

**DEVELOPING A METHOD FOR CONTINUOUSLY MONITORING
DISSOLVED ORGANIC CARBON CONCENTRATION IN BOREAL
FOREST HEADWATER STREAMS**

By

Christian Javier Gaviria Salazar

A Thesis submitted to the Department of Environmental Science in
partial fulfillment of the requirements for the degree of

Master of Science

Department of Environmental Science

Memorial University of Newfoundland and Labrador

May 2021

St. John's

Newfoundland

Abstract

Headwater streams are an important medium through which carbon from the landscape is transported into aquatic ecosystems in the form of dissolved organic carbon (DOC), an ecologically significant and, until the last decade, underestimated pool of mobile carbon. Boreal forests contain a great fraction of the worlds terrestrial carbon and are considered a large carbon sink. However, they are vulnerable to climate change and can easily become sources of atmospheric carbon. To better understand how our landscapes are responding to climate change we can monitor DOC within headwater streams which integrate and quickly respond to changes of the surrounding landscape. However, continuous monitoring for DOC is difficult and must be monitored via proxy-measurements. This thesis demonstrates an approach to develop and monitor the performance of a model developed to estimate DOC from UV-VIS absorbance and other data from *in-situ* probes.

Acknowledgements

Firstly, I would like to thank my supervisor, Dr. Susan Ziegler, for granting me the opportunity to learn and expand my research skills: first as an undergraduate student, and now as a graduate student culminating with this M.Sc. project, you have always provided excellent guidance and support. It was a great fit when Dr. Ziegler introduced me to Dr. Allison Myers-Pigg, at the time looking for an undergraduate student to partake in a project involving spectrophotometry (at which time I had some interest in spectroscopy and minor experience with colorimetry) and dissolved organic matter. That project then evolved into this thesis, and so I also wanted to thank Dr. Myers-Pigg for being present every step of the way and forming part of my M.Sc. committee. You both have provided such invaluable feedback, mentorship and guidance.

I would also like to thank my committee members Dr. Heather Reader and Dr. Alan Roebuck, the latter of who was gradually assimilated as a committee member as we had more and more discussions about the optical properties of DOM stimulated by the various 8-hour long drives across Newfoundland (but not limited to then!). Your inputs throughout my journey here were invaluable and I am very grateful you were a part of my committee.

Many thanks also go to Andrea Skinner, Gordon Butt and Darrell Harris for their skillfull field assistance and field instrumentation maintenance, as well as Jamie Warren and Keri Bowering for their stream probe installations. I would also like to thank Nicole Spehn for her efforts in winter sample collection and processing, Sean Boyd, Jamie Warren and Matthew Norwood for DOC analyses and Maryam Hajheidari, Alexandra Gamble and Melissa Mills for spectrophotometric analyses.

Funding for this work was provided by the Natural Sciences and Engineering Research Council of Canada's Discovery Program and Strategic Partnership Grant (479224), the Canadian Forest Service of Natural Resources Canada, and the Canada Research Chair Programme.

Table of Contents

Abstract.....	ii
Acknowledgements.....	iii
Table of Contents.....	v
List of Tables	ix
List of Figures	x
Appendix	xi
List of Abbreviations and Symbols.....	xii
1 Introduction	1
2 Developing a PLS model to predict dissolved organic carbon concentration from absorbance and water quality parameters	11
2.1 Abstract	11
2.2 Introduction.....	13
2.2.1 Relating absorbance to DOC concentration through regression.....	13
2.2.2 Prediction uncertainty can be estimated and give confidence to predictions.....	15
2.2.3 Water quality and optical parameters can help reduce model uncertainty	16
2.2.4 Chapter Objectives.....	18

2.3	Methods	19
2.3.1	Sample Collection and Analysis	19
2.3.2	Sample Datasets.....	23
2.3.3	Data Processing.....	26
2.3.4	Model Training.....	27
2.3.5	Testing how water quality parameters affect accuracy of model predicted DOC concentration.....	35
2.3.6	Model Evaluation	36
2.4	Results and Discussion	42
2.4.1	Heterogeneity among samples in the training and validation datasets:.....	43
2.4.2	PLS models performed as good as typical MLR models:	51
2.4.3	Additional optical and water quality parameters did not improve model performance:	55
2.4.4	PLS model biased by region but not season:	60
2.5	Conclusion and Implications	63
3	Training and evaluating a PLS model to predict dissolved organic carbon concentration in remote boreal forest stream sites	65
3.1	Abstract	65
3.2	Introduction.....	67

3.2.1	DOC in headwater streams is dynamic.....	68
3.2.2	Using the full CDOM absorbance spectra to predict DOC concentration	70
3.2.3	Using factors that interfere with CDOM absorbance for better predictions	73
3.2.4	Chapter Objectives.....	74
3.3	Methods	77
3.3.1	Sample Collection and Analysis	77
3.3.2	Data Processing.....	81
3.3.3	Model Training.....	83
3.4	Results	96
3.4.1	Different water characteristics between stream sites and storm and baseflow periods	96
3.4.2	Model component selection indicates 8 model components is most robust	97
3.4.3	Hierarchical variable selection.....	102
3.4.4	Training a new model using turbidity corrected absorbance spectra.....	109
3.5	Discussion.....	115
3.5.1	Additional variables in concert with absorbance are useful but may require additional model components in order to characterize them and enhance the prediction of DOC concentration.....	116
3.5.2	PLS model checks are used to determine what <i>in-situ</i> absorbance data we can use and then evaluate which data we should use to predict DOC concentration.....	117

3.5.3	Correction for turbidity-induced absorbance likely requires year-round training data	120
3.6	Conclusion and Implications	122
4	Conclusion.....	125
5	References	134
	Appendix.....	151

List of Tables

Table 2.1: NL-BELT regions climate normals.....	21
Table 2.2: Training and validation datasets.....	24
Table 2.3: Model descriptions.....	29
Table 2.4: Model coefficients of PLS-Plain model.....	30
Table 2.5: Comparing PLS to MLR models	51
Table 2.6: Including other optical parameters as predictors	56
Table 2.7: Testing performance of water quality parameters	60
Table 3.1: Description of 3 sets of model iterations	86
Table 3.2: Sample counts and test-set switch datasets	92
Table 3.3: Test-set switch model names and training datasets	92
Table 3.4: Naming order of hierarchical models.....	94
Table 3.5: Results from 2000 randomized cross-validations.....	101
Table 3.6: Variable importance and SPE contribution in Model 2.5	103
Table 3.7: Comparing validation dataset performance of turbidity-corrected model to scatter-corrected model	112
Table 3.8: Result of 2000 randomized cross-validations for Model 2.4	114

List of Figures

Figure 1.1: The importance of sampling frequency	3
Figure 1.2: CDOM of natural waters.....	5
Figure 2.1: Map of Newfoundland and Labrador.....	19
Figure 2.2: Steps in model development.....	25
Figure 2.3: Bias-Variance trade-off (N. M. Faber & Rajkó, 2007).	34
Figure 2.4: RMSE-CV per component for the PLS-Plain model.....	36
Figure 2.5: Distribution of water quality parameters for the whole dataset	44
Figure 2.6: Bulk DOC and optical parameters related to composition	48
Figure 2.7: Relationship between A_{350} and DOC.....	50
Figure 2.8: Recovery function for PLS-Plain model.....	54
Figure 2.9: Sensitivity analysis testing season and region	62
Figure 3.1: Map of PBEWA catchment	78
Figure 3.2: <i>In-situ</i> instrumentation	81
Figure 3.3: Distribution of water quality and compositional parameters.....	99
Figure 3.4: Result of 2000 randomized cross-validations.....	100
Figure 3.5: Fall period predictions of DOC concentration using 6 PLS-component hierarchical models.....	104
Figure 3.6: Fall period predictions of DOC concentration using 8 PLS-component hierarchical models.....	105
Figure 3.7: Comparing model performance using Hotelling's T^2	107
Figure 3.8: Comparing model performance using SPE	108
Figure 3.9: Development of a correction for turbidity-induced absorbance	110
Figure 3.10: Testing performance of the correction for turbidity-induced absorbance on continuous dataset ..	111
Figure 3.11: Recovery function for 8-component Model 2.4	113

Appendix

Appendix 1: Relationship between stream water DOC and A350.....	152
Appendix 2: Model residuals by season, site and sample type.....	153
Appendix 3: Detailed NL-BELT site information.....	154

List of Abbreviations and Symbols

DOC – Dissolved organic carbon; an operationally defined pool of organic carbon small enough to pass through a 0.7 μm GF/F filter

[DOC] – Concentration of DOC

T-A – Terrestrial-to-aquatic ecosystem

DOM – Dissolved organic matter; an operationally defined pool of organic matter small enough to pass through a 0.7 μm GF/F filter

UV-Vis absorbance - absorbance of light in the ultraviolet and visible light wavelength ranges (~190 to ~700 nm)

CDOM – Chromophoric dissolved organic matter; a fraction of DOM that can absorb light in the UV-vis range (~250 to ~450 nm)

pH – Hydrogen ion activity

WL – Stream water level

SPC – Specific conductance at 25°C

ORP – Oxidation-reduction potential

[Fe] – Concentration of total dissolved iron (Fe^{2+} and Fe^{3+})

λ - Wavelength in nanometers

SUVA – Specific UV absorbance; a measure of absorbance per unit DOC related to aromatic content.

$S_{275-295}$ – Spectral slope of absorbance from 275 to 295 nm

$S_{350-400}$ - Spectral slope of absorbance from 350 to 400 nm

S_R - Spectral slope ratio; $S_{275-295} / S_{350-400}$

NL-BELT – Newfoundland and Labrador Boreal Ecosystem Latitudinal Transect

PBEWA - Pynn's Brook Experimental Watershed Area

UHS – Upper Horseshoe Brook stream site

LHS – Lower Horseshoe Brook stream site

MLR – Multiple linear regression

PLS – Partial least squares

RMSE – Root mean square error; can be adjusted for degrees of freedom

RMSEP – RMSE of predictions; from validation datasets

RMSEE – RMSE of estimation; from training datasets

RMSE-CV – RMSE of cross-validation; calculated from the results of a cross validation of a PLS model for each model component

T^2 – Hotelling's T^2 ; statistic calculated from a PLS models' component scores that measures distance of a projection to the center of the model

SPE – Square prediction errors; also known as Q-statistic or DModX, measures the distance of an observation to the surface of the model.

1 Introduction

Terrestrial-to-aquatic (T-A) fluxes of carbon are an important, yet often overlooked, component of the global carbon cycle, which describes the lateral movement of carbon from terrestrial landscapes to aquatic ecosystems. In recent decades T-A carbon fluxes have been receiving more attention as they are estimated to move a significant amount of carbon, comparable to terrestrial net ecosystem productivity (Battin et al., 2009; Butman et al., 2016; Chi et al., 2020; Cole et al., 2007; Friedlingstein et al., 2020; Tank et al., 2018; Wallin et al., 2013; Webb et al., 2019). Particularly, high-latitude boreal forests are an important carbon sink, storing nearly one-third of global terrestrial carbon (Lal, 2005) and are highly vulnerable to changing temperatures (Lindroth et al., 1998) as a result of climate change (Jennings et al., 2010; Laudon et al., 2012; Worrall et al., 2004; Ziegler et al., 2017). Therefore, it is imperative to quantify and monitor global T-A carbon fluxes to understand and predict landscape feedbacks resulting from climate change.

Headwater streams are intimately connected to their surrounding landscape, and as a result integrate dissolved organic matter (DOM) sourced from different parts of the landscape (Creed, McKnight, et al., 2015). Therefore, when monitored, stream DOM greatly informs about the T-A flux of carbon from the surrounding landscape as DOM represents a broad and ecologically significant proportion of carbon-based natural substances in the aquatic environment (Thurman, 1985). Dissolved organic carbon (DOC) is the typical metric used to quantify DOM and is an important fraction of the total organic carbon (TOC) transported from terrestrial ecosystems through aquatic systems. The flux of TOC from streams that drain different boreal landscapes can

range from $\sim 4000 \text{ kg C km}^{-2} \text{ yr}^{-1}$ from upland catchments to $\sim 12500 \text{ kg C km}^{-2} \text{ yr}^{-1}$ from peatlands (Rantakari et al., 2010), equivalent to $\sim 21\%$ and $\sim 66\%$, respectively, of yearly global carbon emissions from fossil fuels (Friedlingstein et al., 2020). In order to accurately constrain headwater DOC fluxes, high temporal-resolution datasets are needed. This is a result of headwater DOC content being temporally dynamic particularly over storm events when the mobilization of landscape-derived DOC is facilitated and more easily integrated into stream water (Creed, McKnight, et al., 2015; Jollymore et al., 2012; Werner et al., 2019). Using low temporal-resolution datasets to estimate stream DOC flux can lead to great underestimates. For example, high-resolution sampling estimates can capture 1.5-times more carbon exported than with traditional sampling approaches (Figure 1.1; Jollymore et al., 2012). However, DOC concentration cannot be directly measured *in-situ* and thus requires an indirect measurement approach.

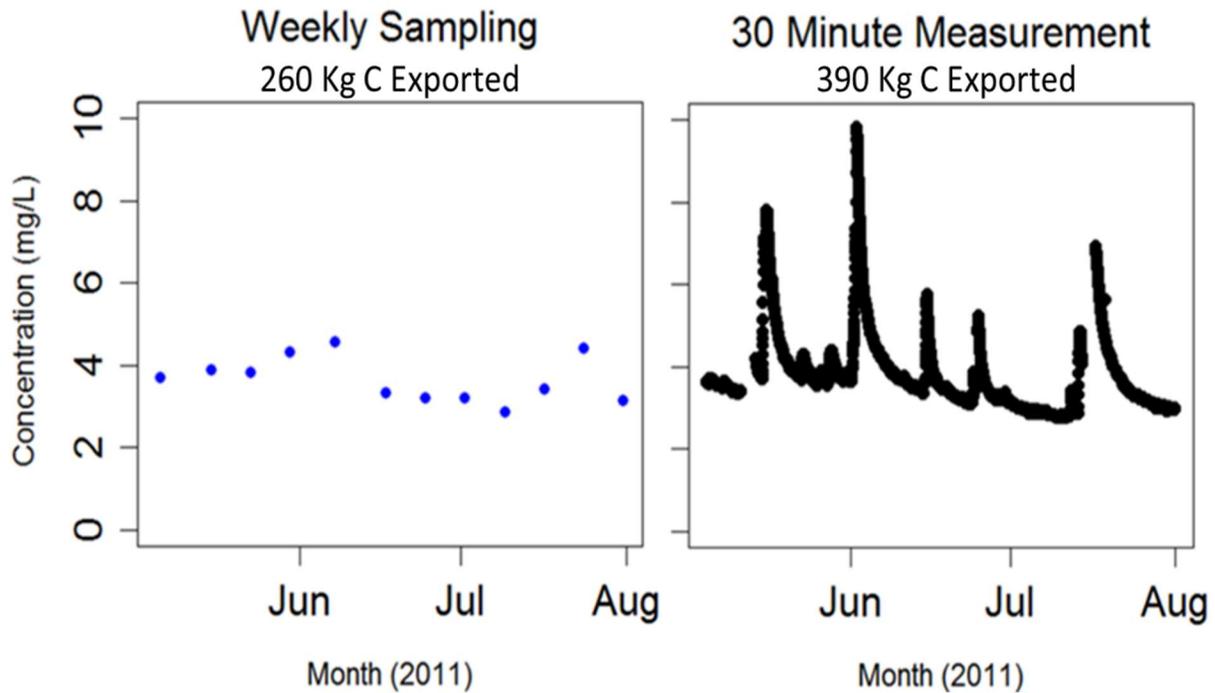


Figure 1.1: The importance of sampling frequency

Adapted from Jollymore et al. (2012), <https://doi.org/10.3390/s120403798>, licensed under Creative Commons Attribution 4.0.

Carbon export estimates calculated from stream dissolved organic carbon sampled at weekly and 30-minute intervals.

Chromophoric dissolved organic matter (CDOM) is the fraction of DOM that absorbs light in the UV-visible wavelength range (~250 to ~450 nm), dependent on the presence of specific functional groups in molecules of DOM (e.g. aromatic rings). In other words, DOC is composed of a light-absorbing component of DOM (CDOM) and non-light-absorbing component of DOM. The Beer-Lambert law relates absorbance of a specific wavelength to the concentration of a solute in a solution as:

$$A_{\lambda} = \epsilon l c \quad (1.1)$$

where A_{λ} is the absorbance of the attenuating species at a specific wavelength, ϵ is the molar attenuation coefficient, an intrinsic property, of the attenuating species, l is the optical path

length and c is the concentration of the attenuating species. Therefore, theoretically, using the linear relationship between A_λ and c , UV-visible absorbance can be used to indirectly measure *in-situ* DOC concentration by estimating it using a trained model and CDOM absorbance measured from an *in-situ* spectrophotometer. However, headwater streams integrate different components of the landscape depending on hydrology that varies with season and storm events (Creed, Hwang, et al., 2015; Creed, McKnight, et al., 2015; Creed & Band, 1998; Vaughan et al., 2019). Consequently, stream DOM is temporally variable, containing different proportions of CDOM and non-CDOM mixtures and different compositions of CDOM (e.g. aromaticity) depending on which components of the landscape are connected to the stream at a given time. This highlights the dynamic relationship between CDOM absorbance and DOC concentration expected in headwater streams. Due to the heterogenous nature of DOM and variability in the proportion that CDOM represents in DOM, ε (Eq. 1.1) of stream water is variable. In this case, while the Beer-Lambert law will work with dilutions of a single sample, it may not work well across different samples due to the heterogeneity in DOM (resulting in the ε from stream water being variable). However, the properties of CDOM, described by specific UV absorbance, spectral slopes and their ratio derived from UV-visible absorbance (Helms et al., 2009), can be used to estimate properties of DOM that can be related back to different components of the landscape and source of DOM (Creed, McKnight, et al., 2015; Franke et al., 2012; Peacock et al., 2014; Vaughan et al., 2019).

The ability to predict bulk stream DOC content (which includes the non-CDOM component of DOC) from CDOM absorbance is affected at two main levels: (1) by the proportion of CDOM to non-CDOM contributing to bulk DOC content in the water, as clearly evident when looking across

sample from low CDOM groundwater contrasted with high CDOM soil water (Figure 1.2), and (2) the composition of CDOM that affects the shape of the absorbance spectra primarily governed by the presence of various functional groups that make up CDOM, evident when looking across high CDOM samples with different origin (stream vs. soil water; Figure 1.2) and further influenced by the presence of metals and other contaminants (Helms et al., 2009; Li & Hur, 2017). Each of the aforementioned factors are dependent on the degree to which the landscape is connected to the aquatic environment as defined primarily by catchment morphology and hydrology (baseflow or stormflow conditions; Kaplan & Cory, 2016).

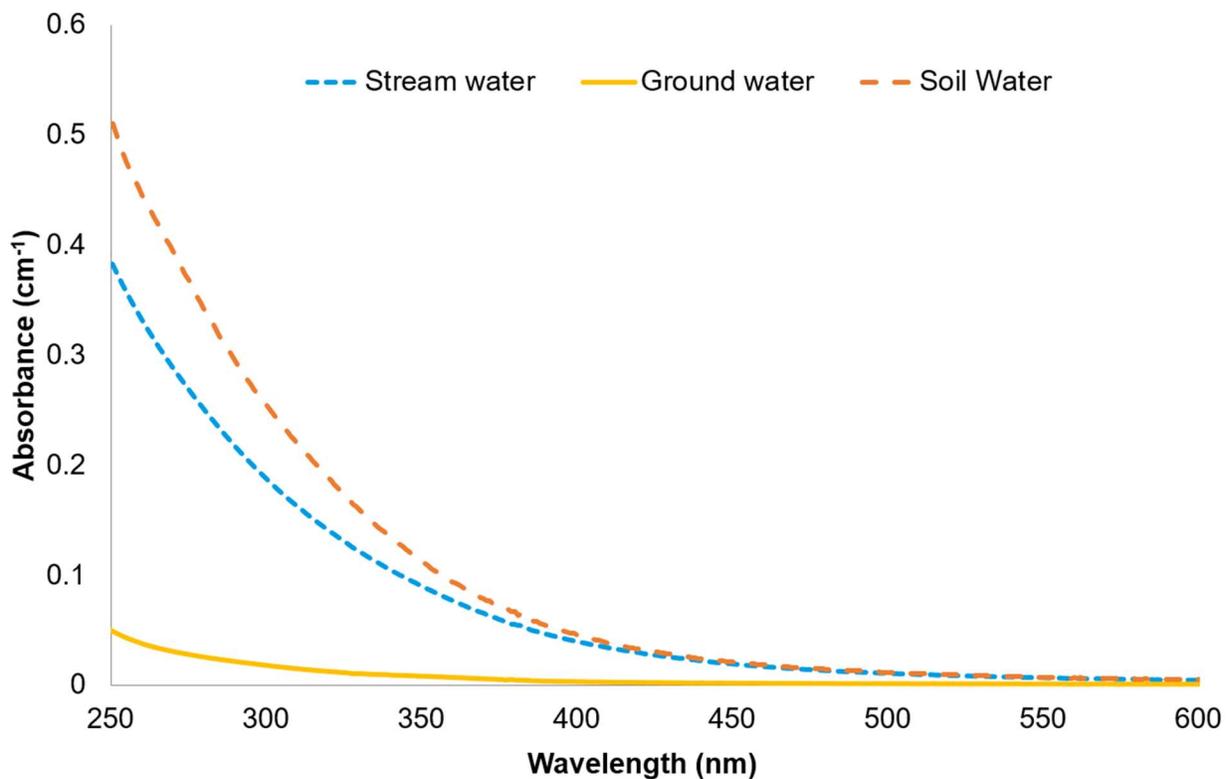


Figure 1.2: CDOM of natural waters

The different relationships between chromophoric dissolved organic matter (CDOM) absorbance and dissolved organic carbon concentration in different natural waters. Each sample had DOC ~8 mg C L⁻¹, but different

compositions of CDOM and different proportions of chromophoric and non-chromophoric DOC contributing to bulk DOC concentration.

Further variability in the relationship between CDOM absorbance and DOC concentration is introduced in streams as turbidity and pH. Turbidity from organic and inorganic particles directly interferes with CDOM absorbance by scattering and absorbing light (Huber & Frost, 1998; Jeong et al., 2012; Langergraber et al., 2003). Turbidity typically correlates with storm events, as increased stream discharge allows for larger particle loads (Downing et al., 2012; Wymore et al., 2019) and flushing particles settled along streambeds. On the other hand, pH affects CDOM absorbance by altering DOM structure (Baalousha et al., 2006; Pace et al., 2012; Spencer et al., 2007), altering conjugated systems in the molecule and its ability to absorb light. Stream water pH is typically found to be inversely correlated to water level during storm events as organic acids, a component of DOC, from the surrounding landscape, are mobilized and integrated into the stream (Ågren et al., 2010; Laudon et al., 2001; Laudon & Buffam, 2008). Thus, part of the change in CDOM absorbance could be attributed to a shift in pH that does not necessarily correspond to a change in DOC concentration.

Models used to predict DOC concentration are typically developed for less dynamic systems that are dominated by a single main landscape component such as a peat bog or mire (e.g. Avagyan et al., 2014), where the relationship between CDOM absorbance and DOC concentration is less variable than from headwater streams draining hillslopes that consequently integrate different landscape components. In such cases, multiple linear regression (MLR), used as an extension of the Beer-Lambert law, has been used to predict DOC concentration from absorbance at discrete wavelengths. In systems where multiple landscape components contribute to stream DOC; DOC concentration and composition are expected to vary widely

across spatial and temporal scales. Training a model intended for a dynamic system, using CDOM absorbance from discrete wavelengths, would not capture all the information available in the full CDOM spectra. Using the full CDOM absorbance spectra would be ideal as the resulting model would use all available information to predict DOC concentration from a dynamic system where a wide variety of relationships between CDOM absorbance and DOC content with varying compositions, and is expected. However, absorbance data has many variables that are also multicollinear. Using too many or multicollinear variables for an MLR can result in overfit- and unstable models that are overly sensitive to minor differences in correlation structure, including random noise, which is expected from *in-situ* data where bubbles, particles and biofouling are a regular occurrence. The resulting models would assume a rigid CDOM absorbance to DOC concentration relationship, possibly resulting in poor DOC predictions when certain compositions of DOC are dominant or present in stream water.

A different model approach is needed for dynamic headwater streams that drain forested hillslopes with variably connected landscape components and, therefore, a temporally variable relationship between CDOM absorbance and DOC concentration. Partial least squares (PLS) is a regression method where a large number of variable and multicollinear data can be used (Dunn, 2019). As such, PLS regression is a chemometric tool used to predict concentrations from absorbance spectra (Wold & Trygg, 2004) and has been used for the purpose of predicting *in-situ* DOC concentration (Avagyan et al., 2014; Codden et al., 2020; Langergraber et al., 2003; Werner et al., 2019), though not to my knowledge in dynamic hillslope-dominated headwater streams with varying contributions from landscape components.

The PLS method has the further benefit of having built-in statistics that are used for process monitoring, e.g. in manufacturing plants, to identify outlier samples or faulty products and diagnose problems in a process (Chen et al., 2004; Dunn, 2019; Mujica et al., 2011; Villegas et al., 2010) which could be helpful in tracking model development and performance for prediction of DOC concentration. One of these statistics, Hotelling's T^2 , identifies high-leverage samples outside the model's scope, from which predictions should be made with caution, and therefore inform the scope of the model and representativeness of the training dataset. Another statistic, squared prediction errors (SPE), identifies samples that are incompatible with the model and cannot be used to make a prediction (Dunn, 2019; Mujica et al., 2011). These two statistics (Hotelling's T^2 and SPE) resulting from different models applied to the same continuous *in-situ* dataset can be compared to identify the model most likely to perform best when predicting DOC concentration in a future setting. Better performing models would have fewer T^2 and SPE outliers in testing datasets.

Model performance is typically reported as the root mean square error (RMSE), which summarizes the variability of prediction residuals in a dataset, and the coefficient of determination (R^2) of each dataset. The RMSE alone does not give information about outlier predictions. Also, R^2 is not helpful as a measure of predictability as it says nothing about prediction error or prediction intervals and tends to be arbitrarily large or small regardless of a model being correct (Shalizi, 2015; 2013). Therefore, other metrics of model fit should be explored when comparing models to identify the best-performing model. To complement RMSE, mean absolute error (MAE) can also be reported and together inform about the magnitude of outlier errors. Prediction intervals give a confidence range around each prediction made by a

model based on the performance with training and validation datasets with more accurate models generating smaller intervals.

Several significant knowledge gaps have been identified regarding the approach and development of DOC-predicting models: (1) The application and performance of models to predict DOC concentration in dynamic headwaters where different landscape components contribute at different times of year have not been described. Therefore, it is unknown whether a PLS model trained using the full CDOM absorbance spectra can improve prediction accuracy of DOC concentration over an MLR model trained with absorbance from discrete wavelengths for these cases. (2) It is also unknown if other properties of CDOM spectra related to composition (i.e. spectral slopes and their ratio) provide information that improves prediction accuracy when used in these PLS models. (3) Other metrics related to stream water quality that respond to seasonality and storm events (e.g. conductivity, pH, water level) should also be tested as predictors of DOC concentration in PLS models because they provide information about the stream water that may improve prediction accuracy of DOC concentration. (4) Other PLS statistics, T^2 and SPE, used to monitor samples and model performance, identify, and diagnose outlier samples, and inform model updates should be evaluated in a stream setting to aid in developing responsive monitoring of stream DOC concentration. (5) The effects of turbidity are known, and there is much effort in measuring *in-situ* turbidity, yet there is not much consensus on the processing of *in-situ* absorbance data to remove the effects of turbidity.

This thesis is divided into two main chapters preceding the conclusion chapter. The first of these describes a study that, by taking advantage of archival samples collected from a series

of upland boreal forest headwater streams and surface soil water from sites across a climate transect, aims to:

1. Use different model performance metrics such as the index of agreement or prediction intervals to evaluate and compare the performance of MLR to PLS models developed to test how: (a) indicators of DOM composition such as spectral slopes and their ratio, and (b) common continuously monitored water quality parameters, such as conductivity, pH, and redox potential, affect model performance.

The second chapter describes a study that, by using high-frequency *in-situ* absorbance spectral data, conductivity and water level data coupled with targeted event sampling from a headwater catchment with an upstream low-relief wetland- and pond-dominated site and a downstream forested hillslope-dominated site, aims to:

2. Train new PLS models and test their performance using T^2 and SPE when applied to the continuous *in-situ* dataset to: (a) test *in-situ* data that relate storm events and seasons to DOC concentration such as conductivity, water level, spectral slopes from 275-295 nm and 350-400 nm, and their ratio to identify best predictors, and (b) develop and test an absorbance-based correction for turbidity to process *in-situ* absorbance data used for predicting DOC concentration.

2 Developing a PLS model to predict dissolved organic carbon concentration from absorbance and water quality parameters

2.1 Abstract

Accurate quantification of dissolved organic carbon (DOC) fluxes from terrestrial to aquatic systems is imperative for understanding landscape and downstream aquatic responses to climate change. This requires high-frequency headwater stream DOC concentration data, which can now be estimated using *in-situ* spectrophotometers by modelling DOC from absorbance of chromophoric dissolved organic matter (CDOM). The challenge is that the relationship between DOC concentration and CDOM is impacted by stream DOC composition that varies significantly with hydrology as DOM composition from several landscape components varies due to different sources and processing. Therefore, a wide breadth of samples spanning sources or different hydrographic stages is still needed in this model development. Furthermore, the use of discrete wavelengths as a bulk DOC proxy can lead to inaccurate or poorly constrained estimates. Therefore, a range of wavelengths, optical-derived indicators of DOC composition and samples representative of catchment sources, e.g. surface and groundwaters, were used to train partial least squares (PLS) models that integrate these qualities for bulk DOC predictions. Stream water quality parameters such as pH, specific conductance (SPC), or redox potential (ORP) were tested to determine whether prediction accuracy of DOC concentration and uncertainty could be

improved. Prediction intervals were used to constrain prediction uncertainty. Using Willmott's refined index of agreement, model performance was assessed on an independent validation dataset for each model. DOC concentration was predicted within ± 3.7 mg C L⁻¹ with 95% confidence in a range of 1-40 mg C L⁻¹ with a PLS-model using only CDOM absorbance. The mean absolute error (MAE) for the validation dataset was 0.40 mg C L⁻¹ over a range of 4-30 mg C L⁻¹ using the same model. PLS model performance was comparable, if not better, to that of typical multiple linear regression (MLR) models that use highly collinear discrete wavelengths to predict DOC concentration. Two tested MLR models had a similar MAE (0.40 and 0.52 mg C L⁻¹) as the PLS model (0.40 mg C L⁻¹) for the same dataset. It was found that adding either spectral slopes, slope ratio, pH, ORP or SPC as an independent variable in the regression did not lead to the development of a more accurate model with reduced uncertainty, most likely due to an over-diverse training dataset (i.e. groundwaters and soil waters). Future work will include testing model performance using high-resolution stream data where the full spectrum of DOC concentration, CDOM absorbance and associated interactions with water quality parameters (e.g. pH and conductivity) can be expected. This will enable further refinement of the PLS models and is expected to improve predictions' accuracy of DOC concentration.

2.2 Introduction

Greater availability of continuous *in-situ* optical measurements provides a benefit to the great need to continuously monitor dissolved organic carbon (DOC) concentrations from streams in support of understanding landscape carbon balance globally (Cole et al., 2007; Jollymore et al., 2012). Headwater streams are intimately connected to their surrounding landscape and integrate features of the land they drain. As such, small headwater streams can be the most significant contributors of terrestrial DOC export per unit area (Ågren et al., 2007). Therefore, high temporal-resolution DOC concentration measurements in small headwater streams are imperative to determining how terrestrial to aquatic transport of carbon responds to climate change. However, frequent sampling to obtain accurate DOC flux is needed as DOC concentration is temporally variable dependent on hydrologic conditions. Thus infrequent sampling may lead to severe underestimations of DOC flux estimates (Jollymore et al., 2012). Therefore, these techniques using *in-situ* optical measurements can improve the accuracy of reported carbon transported from the landscape as DOC.

2.2.1 Relating absorbance to DOC concentration through regression

Chromophoric dissolved organic matter (CDOM) describes the fraction of DOC that can absorb light in the 250 nm to 450 nm wavelength range. *In-situ* spectrophotometers measuring CDOM absorbance in stream water can be used as a proxy for DOC concentration measurements, as CDOM content can be highly correlated to bulk DOC content (Peacock et al., 2014).

Furthermore, the optical properties of DOM described by specific UV absorbance, spectral slopes and their ratio can be used to estimate properties of DOM (Helms et al., 2009) that can be related to different components of the landscape or source of DOM (Creed, McKnight, et al., 2015; Franke et al., 2012; Peacock et al., 2014; Vaughan et al., 2019). Often, absorbance at 254 nm is used as a surrogate for DOC concentration measurements by developing a model from a calibration curve, i.e. linear regression. However, depending on the composition of DOC or CDOM, specific wavelengths can be more suitable than others to use as a proxy for DOC concentration (Peacock et al., 2014), highlighting a dynamic relationship between CDOM absorbance and DOC concentration (Figure 1.2). As headwater streams integrate different landscape components depending on hydrology, which varies with season and storm events, the relationship between CDOM absorbance and DOC concentration in small headwater streams is expected to vary across different flow regimes.

For determining DOC concentration from more complex systems where different sources of water are mixed, the use of CDOM absorbance from multiple wavelengths in multiple linear regression (MLR) models can generate more robust models (Avagyan et al., 2014; Fichot & Benner, 2011). However, multicollinearity among wavelengths in spectral absorbance data violates the assumption MLR makes of no multicollinearity among variables. The violation of this assumption can lead to unstable model coefficients that are very sensitive to the data's covariance structure. However, this is not detrimental to predictions made by the model, only as long as the covariance structure of future samples is near identical to the model's covariance structure. In other words, the model is rigid and could be considered over-fit and can greatly over-or under-predict future samples that do not have a similar covariance structure as the

original data. Due to the dynamics of headwater streams, where hydrologic events that connect different landscape sources to the stream (Vaughan et al., 2017) alter CDOM observed in the water (Creed, McKnight, et al., 2015; Franke et al., 2012; Peacock et al., 2014; Vaughan et al., 2019), changes in the covariance structure of the data is expected over time (Figure 1.2).

Partial least squares (PLS) is a multivariate latent variable regression method that allows for multicollinear variables to be used to develop a model (Dunn, 2019). Multicollinearity is eliminated by reducing variables to determine latent variable scores that maximize covariance between independent- and dependent-variables (Dunn, 2019). PLS regression allows for the use of the entire absorbance spectra to predict DOC concentration as accurately as possible and therefore represents an approach that can take full advantage of *in-situ* instrumentation that provides complete absorbance spectral datasets.

2.2.2 Prediction uncertainty can be estimated and give confidence to predictions

Predictions made by models have an uncertainty associated with them that is typically only partially reported. However, constraining this uncertainty gives more confidence in predictions of DOC concentration. With less uncertainty around predictions, catchment-wide carbon flux estimates derived from these predictions become more accurate. Therefore, prediction uncertainty should be reported to compare and constrain terrestrial to aquatic fluxes represented by DOC export. The model error typically reported as the root mean square error (RMSE) assesses the spread of regression residuals or errors observed, but no confidence around predictions. Alternatively, prediction intervals predict the distribution of future points where one

estimates upper and lower limit boundaries where the real value may be found in with $100(1-\alpha)\%$ confidence (Montgomery & Runger, 2018), and can be further used to provide confidence to the catchment-scale estimates derived from the predictions of DOC concentration made from the model using *in-situ* absorbance data. The prediction interval is derived from model error; thus, more accurate models result in smaller intervals.

2.2.3 Water quality and optical parameters can help reduce model uncertainty

Additional optical or water quality parameters may be helpful as additional predictor variables for DOC concentration to generate more accurate models. For example, model performance can be impacted by the variability in the relationship between DOC concentration and CDOM absorbance in stream water resulting from different flow regimes across seasons or storm events. This is a result of headwater streams integrating different landscape components according to hydrology dependent on season and storms and each component being a different source or having a different signature of DOM (Creed, Hwang, et al., 2015; Creed, McKnight, et al., 2015; Creed & Band, 1998; Vaughan et al., 2019). Parameters derived from CDOM absorbance, such as spectral slopes and their ratio, are related to DOM or CDOM composition (Helms et al., 2009) and are found to change in stream water throughout hydrologic events (Ågren et al., 2010; Vaughan et al., 2017). Therefore, incorporating these parameters into models that predict DOC concentration for streams could enhance the accuracy of predictions of DOC concentration.

Water quantity or quality parameters that are also monitored continuously *in-situ*, such as water level, pH, specific conductance (SPC) and oxidation-reduction potential (ORP), could be beneficial when included in regression models developed for continuously predicting stream DOC concentration from absorbance. Due to the organic acids that contribute to stream DOC, stream water pH is negatively correlated to DOC concentration in many watersheds (Ågren et al., 2010; Laudon et al., 2001; Laudon & Buffam, 2008), yet pH is positively correlated with absorbance as a result of pH affecting DOM molecular structure (Pace et al., 2012; Spencer et al., 2007). Separately, stream water DOC concentration can be positively correlated with stream discharge (Jeong et al., 2012), but can also exhibit hysteresis patterns with discharge indicating temporal variation in near and distal sources of DOC (Creed, McKnight, et al., 2015; Vaughan et al., 2019). Considering these relationships between stream water DOC concentration, pH, absorbance and discharge, the overall relationship between CDOM absorbance and DOC concentration would vary throughout different hydrologic regimes when all these parameters shift in different ways. This suggests pH may be a valid additional parameter to help predict stream DOC concentration from CDOM absorbance. Similarly, stream water SPC changes with seasons and storm events associated with water source and likely DOC source (Inserillo et al., 2017; Kobayashi, 1986). Therefore, SPC can help reduce uncertainty when used in a model as a predictor for stream DOC concentration during such periods. Using ORP as another predictor may also help reduce uncertainty of DOC concentration predictions as together the effects of pH and ORP affect the solubility of metals (Rosecrans et al., 2017) that bind with DOM but also interfere with absorbance (Maloney et al., 2005).

2.2.4 Chapter Objectives

In this chapter, PLS-based models were developed to predict DOC concentration from absorbance spectra. The performance of these PLS models was compared with two MLR models developed to estimate DOC concentration based on parameters from Avagyan et al. (2014) and Fichot & Benner (2011). Uncertainty of model-estimated DOC concentration using 95% prediction intervals was estimated and used to assess the model performance. Furthermore, the effectiveness of using optical-derived parameters indicative of CDOM composition (e.g. spectral slope) and water quality parameters (e.g. pH, SPC, and ORP) were tested as additional regression variables in concert with CDOM absorbance to predict DOC concentration. The usefulness of these parameters yielding a more accurate model to estimate DOC concentration with lower uncertainty was assessed. The influence of spatial and temporal dynamics on absorbance-based PLS models' ability to predict DOC concentration was also investigated and compared to other MLR methods that use absorbance from discrete wavelengths to predict DOC concentration.

2.3 Methods

2.3.1 Sample Collection and Analysis

Study Area:

Water samples were collected from forest and stream sites located within the Newfoundland and Labrador Boreal Ecosystem Latitudinal Transect (NL-BELT; Figure 2.1). The NL-BELT consists of four regions delineated by watersheds of the four major rivers located in each: Grand Codroy (GC), Eagle River (ER), Salmon River (SR) and Humber River (HR). The ER is the northernmost region, GC is the southernmost region, with HR and SR found in between.

Sites within NL-BELT span a 5.2°C gradient in mean annual temperature (MAT) (Table 2.1), where northern sites (ER) have a MAT of 0°C and southern sites (GC) have a MAT of 5.2°C.

Mean annual precipitation (MAP) spanned a range from 1074mm for northern sites and 1505mm for southern sites (Table 2.1, Environment Canada Climate Normals 1981-2010). The NL-BELT

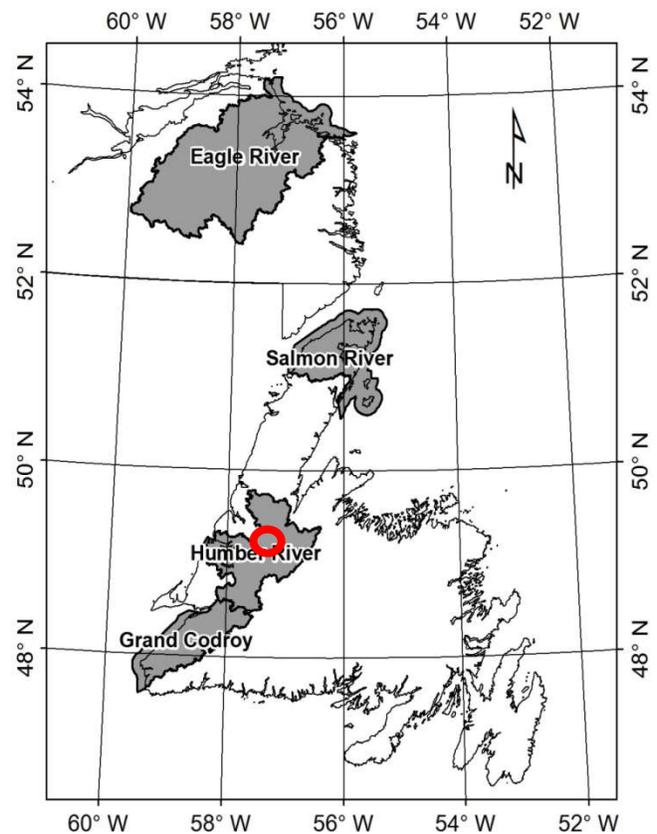


Figure 2.1: Map of Newfoundland and Labrador

The four regions comprising the Newfoundland and Labrador Boreal Ecosystem Latitudinal Transect (NL-BELT). The red circle marks the Pynn's Brook Experimental Watershed Area (PBEWA) located within the Humber River region.

forest sites consist of closed-canopy mature stands dominated by balsam fir (*Abies balsamifera* (L.) Mill.) underlain by humo-ferric podzolic soils (Ziegler et al., 2017). Each region contains three headwater stream sites with catchments that primarily drain forests akin to the forest sites instrumented and studied along the transect (Table 2.1). Additional samples were collected from the Pynn's Brook Experimental Watershed Area (PBEWA), within the HR region, where a greater variety of sample types were available within an instrumented hillslope, including passive pan lysimeters, piezometers and groundwater wells. The forest site in PBEWA differs from the NL-BELT forest sites by the PBEWA site being predominantly dominated by black spruce (*Picea Marina* (Mill.) Britton, Sterns & Poggenburg) rather than balsam fir and is a mixture of age-class ranging from mature stands (~80 years) to a post-14-year stage of regeneration. The stream catchment of the PBEWA contains a mixture of black spruce and balsam fir in addition to fens and bogs (Table 2.1). Of the two sampling sites located along the headwater stream within PBEWA, one upstream site was dominated by wetlands and ponds, while a hillslope forest dominates the downstream site.

Sample Collection:

Samples were collected from NL-BELT stream sites and lysimeters within forest sites from June 2011 to June 2019. Soil water was collected from passive pan lysimeters using an electric pump and polyethylene tubing from the buried 20 L reservoirs (Ziegler et al., 2017). Similarly, shallow groundwater samples were pumped and collected from 1 to 2.5-meter deep piezometer wells. Baseflow stream grab-samples were taken directly from stream sites in each region.

Opportunistic sampling events occurred for two storm events where grab-samples were collected from a stream site within PBEWA and a site within GC in 60 to 120-minute intervals capturing a window of each event's rising and receding limbs. All samples were collected in acid-washed, amber, high-density polyethylene bottles to avoid photochemical degradation and stored in coolers for transportation back to the laboratory for processing. The same protocol was followed for sample collection in the PBEWA for streams and lysimeters, while additionally sampling from passive piezometers, installed to collect water from the top 30 and 60 cm of mineral soil, as well as groundwater wells installed within the till layer below the lowest groundwater level in late summer. For more information on PBEWA infrastructure and site location, see Figure 3.1.

Table 2.1: NL-BELT regions climate normals

Mean annual temperature (MAT), mean annual precipitation (MAP), and catchment composition ranges for each NL-BELT region. Asterisk – catchment cover % data for HR comes from Horseshoe Brook in the Pynn’s Brook Watershed Area.

NL-BELT Region	MAT (°C)	MAP (mm)	Catchment Area (Km ²)	% Forest Cover	% Bog Cover	% Water Cover
ER	0	1073.5	4 - 11.3	64 - 92	5 - 34	0 - 3
SR	2 - 2.4	1211 - 1224	3.3 - 4.2	71 - 88	12 - 29	0
HR*	4.4 - 5.2	1132 - 1286	3.8 - 11.4	63 - 74	25 - 30	0 - 7
GC	4.3 - 5.2	1340 - 1595	0.6 - 2.6	100	0	0

Climate data from Environment Canada Climate Normals (1981-2010):

https://climate.weather.gc.ca/climate_normals/index_e.html

Sample Preparation and Analysis:

Within 72 hours of collection, all samples were vacuum filtered through pre-combusted glass fibre filters (GF/F; nominal pore size 0.7 μm) to remove particulates. After filtering, stream and groundwaters were sub-sampled in triplicate while piezometer and lysimeter samples were sub-sampled in duplicate for absorbance, DOC concentration, and total dissolved iron ([Fe]) analyses. Piezometer and lysimeter samples were sub-sampled in duplicate as they are better spatially replicated across sites. Sub-samples taken for DOC concentration were acidified to pH ~ 2 using 30% phosphoric acid to halt biological activity. Changes in pH can interfere with absorbance spectra; therefore, mercuric chloride was added (1 μL saturated solution per mL of sample), rather than acid, to fix the samples for absorbance and halt biological activity. Samples for total dissolved [Fe] analysis were kept dark and refrigerated until time for analysis.

Absorbance at wavelengths shorter than 250 nm was excluded due to inorganic compounds influencing these measurements (Tipping et al., 1988). However, wavelengths lower than 275 nm were also excluded as the HgCl_2 was found to affect absorbance up to that wavelength. Because of HgCl_2 , using models from this training dataset will not be applicable for directly predicting DOC concentration from *in-situ* data. However, these results provide a basis for both an approach for model development and testing the validity and accuracy of a PLS-based model that can be repeated with future datasets.

The DOC concentrations were determined via the combustion catalytic oxidation method (Shimadzu TOC-V with autosampler ASI-V) with a limit of quantification of 2.00 mg C L^{-1} (limit of detection of 0.60 mg C L^{-1}) and precision of $<3\%$ for concentrations ranging 0.5 to 50 mg C L^{-1} .

Absorbance spectra (A_λ) of filtered samples were measured using a dual-beam UV-Vis benchtop spectrophotometer (Perkin Elmer Lambda 25) from 200-800 nm at 1 nm intervals. Samples with high absorbance (defined here as containing more than 20 mg C L⁻¹) were diluted to <20 mg C L⁻¹ for samples to minimize inner-filter effects so that samples remained in the linear range of the Beer-Lambert law and then corrected with the appropriate dilution factor. The precision of these absorbance measures was <5% coefficient of variation (RSD) among analytical replicates (n = 3) for decadic absorbance at 300 nm (A_{300}) in the range of 10-60 m⁻¹ with a pathlength of 0.01 m. An RSD threshold of 10% was used to initiate a Q-test between analytical replicates to determine potential outliers.

Water sample pH, oxidation-reduction potential (ORP), and specific conductance (SPC) were all measured following each sample's absorbance measurement. Sample from each analytical replicate from a field replicate was pooled to measure pH, ORP and SPC. pH and SPC were measured using a benchtop meter (Thermo Scientific Orion Versa Star Pro), while ORP was measured using a handheld Oakton ORPTestr (WD-35650-10, Vernon Hills, Illinois).

2.3.2 Sample Datasets

For the development and testing of models, a set of 287 analyzed samples was used with a subset of 33 stream samples reserved and used as an independent validation dataset to test and compare every model's performance (Figure 2.2). The 33 validation samples were collected from June 2018 to June 2019, where 31 of the samples were collected from the Horseshoe Brook, found within the PBEWA where this method and model is intended to be applied in some near-

term studies. The remaining 254 samples were collected from November 2011 to August 2018 and were used as the primary training dataset to develop partial least squares (PLS) and multiple linear regression (MLR) models (Table 2.2). Furthermore, the stream samples from this same training dataset (n = 106) were used alone in a further MLR to test whether water quality parameters (WQP) such as pH, ORP or SPC could be used to improve stream prediction accuracy of DOC concentration and reduce prediction uncertainty when using PLS model predictions.

Table 2.2: Training and validation datasets

Sample subset used for model development and testing. Samples collected from NL-BELT and PBEWA include stream (St), lysimeter (Lys), piezometer (Pz) and groundwater well (GWW) samples. Training datasets were used to develop partial least squares (PLS) or multiple linear regression (MLR) models using a different combination of variables at each iteration. Water quality parameters (WQP) such as pH, oxidation-reduction potential and specific conductance were additionally tested in further MLR models trained by stream samples.

Set	Subset From:	Used for:	Comprised of:	Total n =
Full Data Set		Training and Validation datasets	<ul style="list-style-type: none"> • 139 St • 95 Lys • 50 Pz • 3 GWW 	287
Main Training dataset	Full Dataset	PLS and MLR Models	<ul style="list-style-type: none"> • 106 St • 95 Lys • 50 Pz • 3 GWW 	254
Stream Training dataset	Training dataset	WQP Correction (MLR) Models	<ul style="list-style-type: none"> • 106 St 	106
Validation dataset	Full Dataset	Testing Model Performances	<ul style="list-style-type: none"> • 33 St 	33

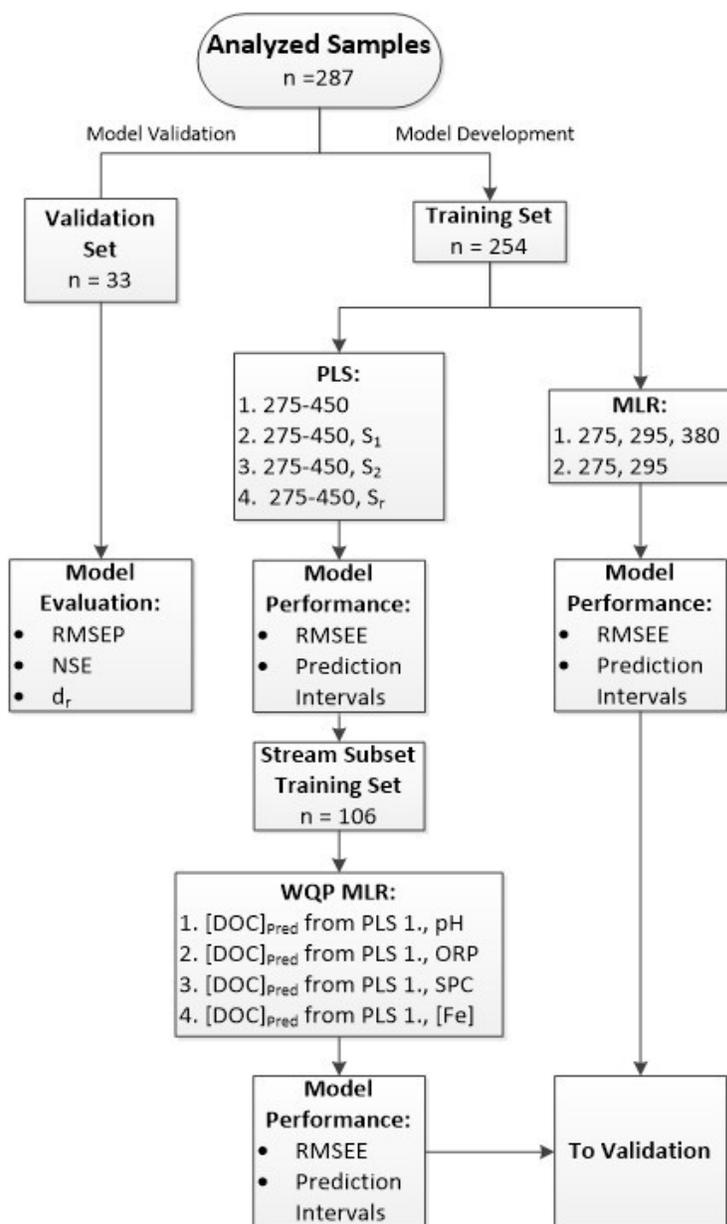


Figure 2.2: Steps in model development

Steps taken for model training and validation. Training datasets of laboratory-measured dissolved organic carbon concentration ([DOC]) and absorbance (analyzed samples) are used to train partial least squares (PLS) and multiple linear regression (MLR) models to estimate [DOC] from absorbance. PLS models regress laboratory [DOC] on absorbance from the 275-450 nm region or additionally spectral slope of 275-295 nm ($S_{275-295}$ or S_1), 350-400 nm ($S_{350-400}$ or S_2) or the slope ratio (S_r). MLR models regress laboratory [DOC] on absorbance from multiple discrete wavelengths 275 nm and 295 nm, or additionally 380 nm. Further MLR models developed to test water quality parameters (WQP) were developed by regressing laboratory-measured [DOC] on PLS-estimated [DOC] ($[DOC]_{pred}$) and a WQP: pH, oxidation-reduction potential (ORP), specific conductance (SPC) or total dissolved iron ([Fe]). Root mean square error (RMSE) was calculated for each model on both training and validation datasets. The 95% prediction interval for each model was calculated from the training dataset. The Nash-Sutcliffe model efficiency coefficient (NSE) and refined index of agreement (d_r) were calculated for each model using the validation dataset.

Absorbance for wavelengths greater than 450 nm was omitted as CDOM absorbance at those wavelengths often approaches the instrument’s detection limit. The laboratory-measured DOC concentration and absorbance for each sample used and reported here is the average of its analytical measurements from the field replicates (field $n = 2-3$; total replicate $n = 4-9$).

2.3.3 Data Processing

Before performing any regressions or developing models, integrative optical metrics related to CDOM composition, such as specific UV absorbance (SUVA) or spectral slope and slope ratio, were calculated. Only absorbance from wavelength intervals of 5 nm was used for the model development as it was the lowest common multiple between *in-situ* stream data collected in 2.5 nm intervals and laboratory-measured absorbance collected in 1 nm intervals.

Specific UV absorbance is the DOC concentration-normalized absorbance at a specific wavelength ($L \text{ mg C}^{-1} \text{ m}^{-1}$), typically 254 or 350 nm. Here $SUVA_{350}$ was calculated for each sample as:

$$SUVA_{350} = \frac{A_{350}}{L \times [DOC]} \quad (2.1)$$

where A_{350} is the absorbance at 350 nm, L is the cuvette path-length (m) and $[DOC]$ is the laboratory-measured DOC concentration (mg C L^{-1}).

Spectral slope (S) refers to the slopes of the line of best fit for the natural log-transformed absorption coefficients (a_λ) across a specified range of wavelengths. Raw decadic absorbance values were converted to Napierian absorbance with the following equation:

$$a_\lambda = \frac{2.303 \times A_\lambda}{L} \quad (2.2)$$

where 2.303 is the constant used to convert values from $\log_{10}(A_\lambda)$ to $\ln(A_\lambda)$, A_λ is decadic (raw) absorbance at wavelength λ (nm) and L is the cuvette path-length (m). Spectral slopes were

calculated using 1 nm wavelength intervals rather than 5 nm intervals by rearranging the following equation:

$$a_{\lambda} = a_{\lambda_{ref}} e^{-S(\lambda - \lambda_{ref})} \quad (2.3)$$

where λ_{ref} is the reference wavelength. The spectral slope ratio (S_r) was calculated by dividing the slope of log-transformed absorption from 275 to 295 nm ($S_{275-295}$) by the slope of log-transformed absorption spectra from 350 to 400 nm ($S_{350-400}$) (Helms et al., 2009).

2.3.4 Model Training

Model calibration, or training, refers to the relating of a set of variables together, as described in the field of chemometrics (Brereton, 2003). In this study, one set of predictors was a series of physiochemical measurements, such as UV-Vis absorbance spectra, pH or other water quality parameters, and another set was DOC concentration. A model is trained when one of these sets is regressed on the other, resulting in a model that relates the variables. To predict DOC concentration using absorbance, the Y-data consisting of DOC concentration data (single variable) is regressed on the X-data containing absorbance measurements (single, multiple discrete or full range). In the case where multiple wavelengths, or additional variables like pH, ORP, SPC, $S_{275-295}$, $S_{350-400}$, S_r or [Fe] comprise the X-data, multiple linear regression (MLR) or partial least squares (PLS) regression is performed. The regression results in DOC concentration-predicting models following a linear structure ($Y = mX + b$), with multiple variables in the X-data simplified to:

$$Y = \sum(m_k \times X_k) + b \quad (2.4)$$

wherein all cases, Y is predicted DOC concentration (or natural log of DOC concentration), m_k is a regression coefficient associated with the predictor variable, X_k , from the X-data used for the regression and b is the regression coefficient constant. All models from this study have an expanded version of equation (2.4) where, for example, m_k and b would be substituted with values from Table 2.4, which are the result of a PLS regression, and X_k would be substituted with absorbance from the corresponding wavelength to predict DOC concentration.

For all models, laboratory-measured DOC concentration (or $\ln[\text{DOC}]$) was used as the Y-data, while a variety of parameters were used as the X-data for MLR or PLS regression (Table 2.3). In all cases, X-data included either multiple discrete A_λ (MLR), a range of A_λ (PLS), or multiple discrete $\ln(a_\lambda)$ (MLR). If a range of A_λ was used, either $S_{275-295}$, $S_{350-400}$ or S_R was used (Table 2.3). To test each water quality parameter, each model's X-data contained PLS predicted DOC concentration and one of pH, ORP, SPC or [Fe] (Table 2.3).

The $\text{lm}()$ function in R (R Core Team, 2019; V. 3.6.6) was used to generate the MLR models, while the $\text{pls}()$ function in the R 'pls' package (Mevik, B.H., Wehrens, R., Hovde, K., Hiemstra, P., 2019; V. 2.7-2) was used to generate the PLS models within R-Studio (V. 1.2.5042).

Table 2.3: Model descriptions

Detailed model parameters were used across model iteration in this study, where Y-data were regressed on X-data. Models were generated through multiple linear regression (MLR) or partial least squares (PLS) regression. PLS models used a segment of the absorbance spectra ($A_{275-450}$) and spectral slopes ($S_{275-295}$, $S_{350-400}$) or slope ratio (S_r) to predict dissolved organic carbon concentrations ([DOC]). Some MLR models used absorbance from multiple discrete wavelengths (A_λ). X-data of the water quality parameter (WQP) correction models used either pH, oxidation-reduction potential (ORP), specific conductance at 25°C (SPC), or concentration of total dissolved iron ([Fe]) with predictions of DOC concentration ($[DOC]_{Pred}$) made by the PLS-plain model to test if WQP would improve [DOC] prediction accuracy.

Model	Y-data	X-data	# Variables in X-data	Regression Method	Training dataset Used
Fichot & Benner (2011)	ln[DOC]	ln(a_{275}), ln(a_{295})	2	MLR	Median Split Main
Avagyan et al. (2014)	[DOC]	A_{275} , A_{295} , A_{380}	3	MLR	Main
PLS-Plain	[DOC]	$A_{275-450}$	36	PLS	Main
PLS- $S_{275-295}$	[DOC]	$A_{275-450}$, $S_{275-295}$	37	PLS	Main
PLS- $S_{350-400}$	[DOC]	$A_{275-450}$, $S_{350-400}$	37	PLS	Main
PLS- S_R	[DOC]	$A_{275-450}$, S_R	37	PLS	Main
pH	[DOC]	$[DOC]_{Pred}$, pH	2	MLR	WQP Correction
ORP	[DOC]	$[DOC]_{Pred}$, ORP	2	MLR	WQP Correction
SPC	[DOC]	$[DOC]_{Pred}$, SPC	2	MLR	WQP Correction
[Fe]	[DOC]	$[DOC]_{Pred}$, [Fe]	2	MLR	WQP Correction

Performance of each model resulting from the regression of DOC concentration on X-data from training dataset samples was measured using different metrics, such as prediction intervals, root mean square error, the Nash-Sutcliffe model efficiency coefficient or Willmott’s refined index of agreement, which relate to accuracy or error of the model and uncertainty around predictions of DOC concentration (Figure 2.2).

Table 2.4: Model coefficients of PLS-Plain model

Model coefficients for variables used in PLS-plain would be substituted into equation (2.4) to predict the concentration of dissolved organic matter ($[DOC]_{Pred}$). In this case, absorbance from each wavelength 275-450 nm ($A_{275-450}$).

Coefficient	Value	Coefficient	Value	Coefficient	Value
m ₂₇₅	-16.85	m ₃₃₅	-6.04	m ₃₉₅	-2.57
m ₂₈₀	-1.14	m ₃₄₀	-6.75	m ₄₀₀	-4.85
m ₂₈₅	6.82	m ₃₄₅	-1.58	m ₄₀₅	-4.57
m ₂₉₀	10.45	m ₃₅₀	-1.33	m ₄₁₀	-0.50
m ₂₉₅	12.18	m ₃₅₅	-0.33	m ₄₁₅	-3.80
m ₃₀₀	10.16	m ₃₆₀	-0.38	m ₄₂₀	7.49
m ₃₀₅	6.63	m ₃₆₅	2.98	m ₄₂₅	-11.94
m ₃₁₀	4.02	m ₃₇₀	-3.76	m ₄₃₀	-6.31
m ₃₁₅	1.37	m ₃₇₅	-3.47	m ₄₃₅	0.15
m ₃₂₀	-1.14	m ₃₈₀	12.15	m ₄₄₀	-8.45
m ₃₂₅	-2.61	m ₃₈₅	7.74	m ₄₄₅	5.57
m ₃₃₀	-8.69	m ₃₉₀	2.80	m ₄₅₀	14.07
				constant	1.33

Multiple Linear Regression:

The Method from Fichot & Benner (2011)

Following the steps from Fichot & Benner (2011) $\ln a_{275}$, $\ln a_{295}$ and $\ln[\text{DOC}]$ were calculated and used. The training dataset was then sorted by a_{275} and subsets are taken to provide two separate training datasets, one containing samples higher than the median a_{275} , and the other containing samples lower than the median a_{275} . The regression was carried out with both subsets resulting in the following equation having a similar linear structure ($Y = mX + b$) to equation (2.4):

$$\ln([\text{DOC}]_{\text{Pred}}) = m_1 \times \ln(a_{275}) + m_2 \times \ln(a_{295}) + b \quad (2.5)$$

where $[\text{DOC}]_{\text{Pred}}$ is the predicted DOC concentration by the model, b is the regression constant and m_n are the regression coefficients associated with the respective variable. The same steps were followed with the validation dataset, and $\ln[\text{DOC}]_{\text{Pred}}$ was calculated using the regression equations.

The Method from Avagyan et al. (2014)

Avagyan et al. (2014) used stepwise regression to determine the wavelengths most suitable to be used to predict DOC concentration according to the dataset (or subset of data). The best model used A_λ from 257.5, 380, 730 and 292.5 nm. Avagyan et al. (2014) used A_{730} as it was found to improve predictions of the *in-situ* measurements of unfiltered water as it likely captures sample turbidity. However, since grab-samples in this study were filtered and CDOM

absorbance in the 275 to 450 nm range at 5 nm intervals was used, A_{730} was not used in the regression while A_{275} was used instead of $A_{257.5}$ and A_{295} instead of $A_{292.5}$. Similar to the linear model structure of equation (4) and (5), the model resulting from these adjustments is as follows:

$$[\text{DOC}]_{\text{Pred}} = m_1 \times A_{275} + m_2 \times A_{295} + m_3 \times A_{380} + b \quad (2.6)$$

Partial Least Squares Regression:

PLS regression was used to develop models to predict DOC concentration using absorbance from a range of wavelengths as PLS allows for multicollinear variables can be used. The regression proceeds by determining latent variables (components) that simultaneously:

- Best explain or summarize the input set of data (X; in this case absorbance and other absorbance-based characteristics),
- Best explain or summarize the output set of data (Y; in this case DOC concentration),
- And maximize the co-variance between X and Y (Dunn, 2019).

When doing so, the variables are projected onto a new space where collinearity among variables is eliminated. As with MLR, the Y-space is DOC concentration. However, the X-space is a range of A_λ and additional absorbance characteristics such as $S_{275-295}$, $S_{350-400}$ or S_R , rather than two or three discrete A_λ as used in MLR (Table 2.3).

The PLS regression was performed on column mean-centred and standard deviation scaled data using a *leave-one-out* (LOO) cross-validation (CV; LOO-CV) method to determine the

root mean square error (RMSE) from the training dataset through cross-validation (RMSE-CV). RMSE-CV was then used to identify the potential number of components to be used. A cross-validation is a form of pseudo-validation, where a subset of the training dataset is kept separate while the rest of the set is used to form a model that is then validated with the subset. The process is then repeated with a different subset of samples from the same training dataset. In the case of LOO-CV, one sample is kept from the training dataset to validate the model with the rest of the dataset (n-1), and this is repeated n times.

The PLS regression determines latent variables iteratively, where each subsequent latent variable explains additional variance between the X- and Y-data. Using too many latent variables can lead to an over-fit model with low RMSE-CV but a high RMSE of predictions (RMSEP) for independent or future samples (Figure 2.3). The number of potential components to use for the model was chosen where the minimum RMSE-CV was observed (Figure 2.4) plus the one prior and the one following. Models using each number of components were tested with the validation dataset, and RMSEP was calculated for each.

The number of components was selected by the lowest RMSEP. Figure 2.4 shows the PLS-plain model case, where the minimum RMSE-CV was observed at six components. Then, models using 5, 6 and 7 components were determined, and RMSEP from the validation dataset was calculated using each of the models. The 5-component model had the lowest RMSEP and was therefore selected as the PLS-plain model. The model resulting from a PLS has a linear structure ($Y = mX + b$) similar to equation (2.4):

$$[\text{DOC}]_{\text{Pred}} = \left[\sum_{i=1}^k (m_i \times x_i) \right] + b \quad (2.7)$$

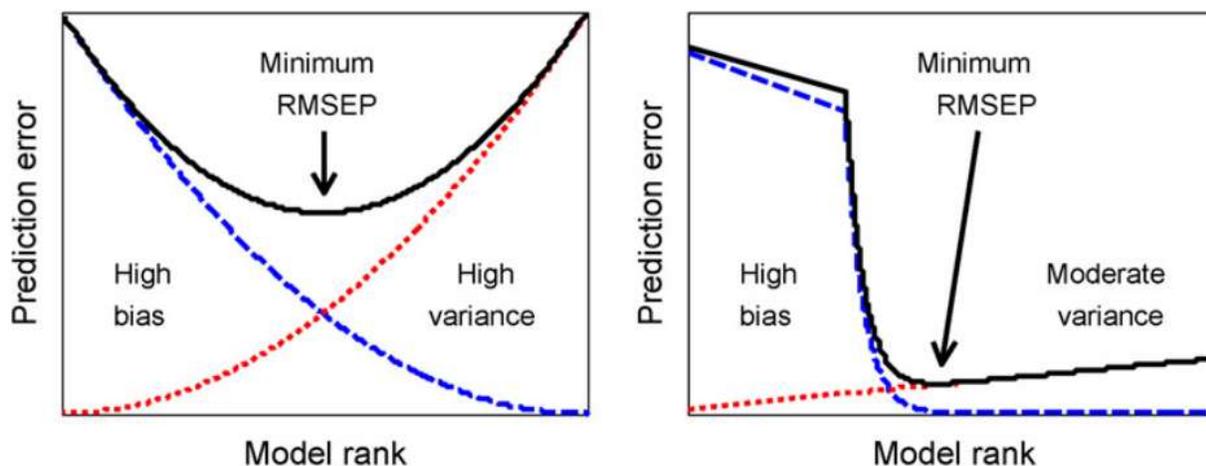


Figure 2.3: Bias-Variance trade-off

Reprinted from *Analytica Chimica Acta*, 595, Faber, N. M., & Rajkó, R., How to avoid over-fitting in multivariate calibration-The conventional validation approach and an alternative, 98–106, Copyright (2007), with permission from Elsevier.

Schematic representations of bias (dashed) and variance (dotted) contributions to the *actual* root mean squared error from future predictions (RMSEP; solid) as a function of model dimensionality (latent variables or components). The minimum RMSEP is ideal for a model. These cases depict RMSEP from hypothetical independent samples, not part of the training dataset. One can think of the bias as prediction error from the training dataset and variance as the prediction error from the validation dataset. The left panel depicts a standard presentation, while the right panel depicts a more realistic presentation (Figure 2.4). In either panel, RMSEP values left of the minimum are suggestive of under-fitting where too few components are used, and predictive features are missed. Alternatively, RMSEP values to the right of the minimum suggest over-fitting where too many components are used, and noise is fit.

where each x_k is associated with each predictor variable used as the X-space and m_k is the regression coefficient associated with that variable. In this case, $A_{275-450}$ was used in every PLS model, while adding one of each $S_{275-295}$, $S_{350-400}$ or S_R were used in separate models to evaluate their potential improvement in model prediction (Table 2.3, Figure 2.2).

2.3.5 Testing how water quality parameters affect accuracy of model predicted DOC concentration

Estimates of DOC concentration from the Plain-PLS model are, first, robustly informed by the full CDOM absorbance spectra of baseflow stream water and stream water source end-members (piezometer and lysimeter samples) that would contribute to stream water during storm events (Table 2.3). Then, using only stream samples from the training dataset, the predictions of DOC concentration from the Plain-PLS model (Table 2.3) were used in concert with one of pH, SPC or ORP in an MLR (Table 2.3) in an attempt to improve prediction accuracy of DOC concentration from stream samples. This step was taken as future *in-situ* data including pH, SPC and ORP are typically collected in stream water and not lysimeters or piezometers. Therefore, the developed models should be informed by the pH, SPC, or ORP signatures observed naturally in a stream. Nevertheless, due to most stream samples in the training dataset being from baseflow conditions, there is minimal variability in the relationship between CDOM absorbance and DOC concentration resulting from different landscape components connecting to the streams during storm events. As such, lysimeter and piezometer samples were used instead to represent end-member sources of DOM that would contribute to stream water during storm events.

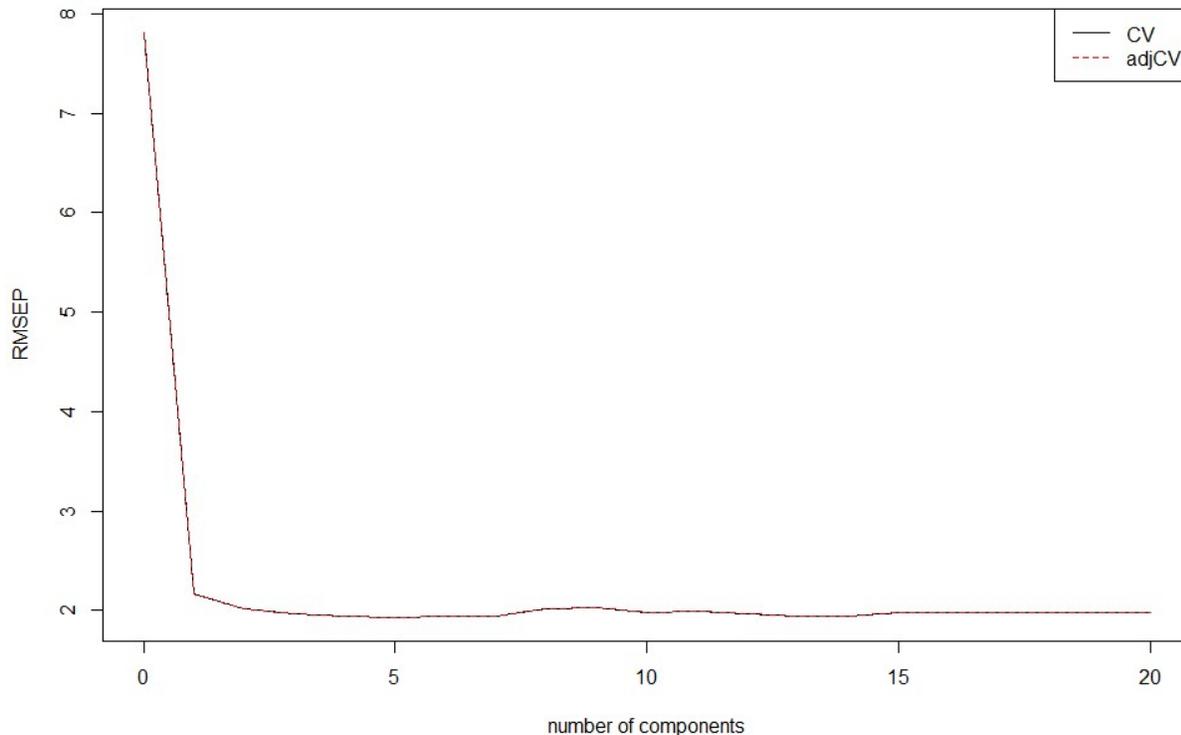


Figure 2.4: RMSE-CV per component for the PLS-Plain model.

Root mean square error from cross-validation (RMSE-CV) curve as a function of model dimensionality for a partial least squares model regressing dissolved organic carbon concentration on A₂₇₅-A₄₅₀. The minimum RMSE-CV was observed at six components. Then, models using 5, 6, and 7 components were tested with the validation dataset, where a 5-component model resulted in lower RMSEP than 6- or 7-component models.

2.3.6 Model Evaluation

Typical model evaluation metrics reported include the root mean square error (RMSE) and coefficient of determination (R^2) of the model. However, while RMSE has the desirable trait of penalizing larger errors, it relates to the spread of errors and does not inform average error. R^2 is not helpful as a measure of predictability as it says nothing about prediction error or prediction intervals while also being arbitrarily large or small regardless of the model being correct (Shalizi, 2015).

Model Residuals

Training datasets are used to generate the predicting models through regression, while the validation dataset made from independent samples is used to test each model's performance by predicting DOC concentration and calculating prediction residuals. Training dataset residuals of fitted DOC concentration from each model were calculated to determine the RMSE and prediction intervals. Validation dataset residuals from predictions of DOC concentration of each model were calculated to determine root mean square error from predictions (RMSEP), the Nash-Sutcliffe model efficiency coefficient (NSE) and Willmott's refined index of agreement (d_r) and compared among models. Prediction residuals were calculated by:

$$[\text{DOC}]_{\text{Res}} = P_i - O_i \quad (2.8)$$

where $[\text{DOC}]_{\text{Res}}$ is the residual DOC concentration (error) from the prediction, P_i is the fitted or predicted DOC concentration by a model, O_i is the corresponding observed or laboratory-measured value. Calculated this way, a positive $[\text{DOC}]_{\text{Res}}$ indicates the model overpredicted DOC concentration for the sample, a negative $[\text{DOC}]_{\text{Res}}$ indicates an underprediction and $[\text{DOC}]_{\text{Res}} = 0$ indicates a perfect prediction. Using prediction residuals, RMSE, also known as the standard error of the model, was calculated for each model by:

$$\text{RMSE} = \sqrt{\frac{\sum([\text{DOC}]_{\text{Res}i})^2}{n-k}} \quad (2.9)$$

where n is the sample size and k is the number of variables used in an MLR model or the number of components for the case of PLS models (Dunn, 2019). RMSE is the standard deviation of residuals adjusted for degrees of freedom used by the model and, therefore, suitable to compare the performance of different models and penalize for using too many or redundant parameters to fit the model. However, RMSEP is not adjusted for degrees of freedom and therefore does not penalize for each parameter used:

$$\text{RMSEP} = \sqrt{\frac{\sum([\text{DOC}]_{\text{Res}_i})^2}{n}} \quad (2.10)$$

where RMSEP was calculated from validation dataset residuals.

Prediction Intervals:

Model residuals from the training dataset were used to determine the 100(1- α)% prediction interval. The prediction interval is an upper and lower limit associated with a predicted value designed to show on a probability basis the range of error associated with the prediction (Montgomery & Runger, 2018). They describe the error in estimation from a regression model and estimate a constraint on uncertainty around that model's predictions. The prediction's real value is estimated to be found within the prediction interval with 100(1- α)% confidence. In the case of this study, an α level of 0.05 was used. Prediction intervals for each sample was calculated as:

$$P_i \pm t_{crit} * \sqrt{\mathcal{V}\{P_i\}} \quad (2.11)$$

where P_i is a model predicted value, t_{crit} are the corresponding critical t-value at a specified $\alpha/2$ with $n - k$ degrees of freedom, and $\mathcal{V}\{P_i\}$ is the variance of predicted values, calculated by:

$$\mathcal{V}\{P_i\} = \text{MSE} * \left(1 + \frac{1}{n} + \frac{(P_i - \bar{P})^2}{SS_P} \right) \quad (2.12)$$

where all the parameters come from the training dataset. n is the sample size, MSE is the mean squared error ($\frac{1}{n} \sum ([\text{DOC}]_{\text{Res}_i})^2$), P_i is a predicted value, \bar{P} is the mean of all predicted values, SS_P is the sum of squared deviations about the mean of predicted values ($\sum (P_i - \bar{P})^2$). In order to confirm the validity of the prediction intervals, at least $100(1 - \alpha)\%$ of the residuals from the training dataset should fall within the prediction interval ($| [\text{DOC}]_{\text{Res}_i} | < t_{crit} * \sqrt{\mathcal{V}\{P_i\}}$).

To calculate prediction intervals for the validation dataset and future datasets, $\mathcal{V}\{P_i\}$ was calculated using the RMSE, SS_P , n and \bar{P} values from the training dataset used to train the model while using the appropriate P_i from the validation dataset or future samples.

The Nash-Sutcliffe Model Efficiency Coefficient:

The Nash-Sutcliffe model efficiency coefficient was developed to assess hydrological models' predictive ability by comparing the variance of prediction errors to the variance of observed data (NSE; Nash & Sutcliffe, 1970). It was calculated from the validation dataset residuals. When used to evaluate models here to predict DOC concentration, it was found to be near-identical to R^2 , with higher values indicating better model performance, calculated as:

$$\text{NSE} = 1 - \frac{\sum([\text{DOC}]_{\text{Res}_i})^2}{\sum(O_i - \bar{O})^2} \quad (2.13)$$

where \bar{O} is the mean of measured values. An NSE value equal to 1 would mean $\sum([\text{DOC}]_{\text{Res}_i})^2 = 0$ and therefore perfect predictions, while $\text{NSE} = 0$ indicates $\sum([\text{DOC}]_{\text{Res}_i})^2 = \sum(O_i - \bar{O})^2$ meaning that using \bar{O} to predict DOC concentration is as accurate as using the model predictions from a particular dataset. Therefore, $\text{NSE} < 0$ indicates that using the mean observed value to predict DOC concentration is more accurate than using model predictions.

The Refined Index of Agreement:

The refined index of agreement (d_r) is another unitless indicator of model performance similar to NSE (Willmott et al., 2012). It is an updated and more robust version of the original index of agreement (d) that, unlike NSE, which uses squared error or deviation values, compares the absolute error of the model predictions to the absolute deviations about the observed mean (Legates & McCabe, 1999; Willmott et al., 2012). Therefore, it is a more robust indicator of model performance that can be used instead of R^2 . The d_r is calculated as:

$$d_r = 1 - \frac{\sum|[\text{DOC}]_{\text{Res}_i}|}{c * \sum|O_i - \bar{O}|} \quad (2.14)$$

where c is a scaling factor, suggested being set equal to 2 (Willmott et al., 2012). Ranging from +1 to -1, a d_r value of 0.5 indicates that the sum of the error-magnitudes ($\sum|[\text{DOC}]_{\text{Res}_i}|$) is one

half the sum of the observed-deviation magnitudes ($\sum |O_i - \bar{O}|$) with the perfect-model-deviation (where $P_i = O_i$, hence using $c = 2$ and the denominator becoming $2 * \sum |O_i - \bar{O}|$). Inasmuch, d_r is interpreted in terms of mean-absolute error (MAE; $\frac{1}{n} \sum |[DOC]_{Res_i}|$) from model predictions and the mean-absolute deviation (MAD; $\frac{1}{n} \sum |O_i - \bar{O}|$). When $c = 2$ and the denominator becomes $2 * \sum |O_i - \bar{O}|$, one MAD accounts for the observed MAD (around O_i) while the second represents the average magnitude of the perfect-model deviations (around P_i , when $P_i = O_i$). The interpretation of $NSE = 1$ is also applicable to $d_r = 1$, as perfect predictions mean $P_i = O_i$ and $\sum |[DOC]_{Res_i}| = 0$. However, as the lower limit of d_r is approached, interpretations should be made cautiously as it could mean model estimated deviations about \bar{O} are poor estimates of observed deviations, or there simply is little observed variability (Willmott et al., 2012). As opposed to R^2 , d is a better indicator of model fit (Legates & McCabe, 1999), and d_r further improves the utility of d as an index of model fit by addressing caveats of the original d (Pereira et al., 2018; Willmott et al., 2012).

2.4 Results and Discussion

Developing a model to predict DOC concentration from absorbance will allow us to monitor stream DOC content using in-situ spectrophotometers continuously. With a dataset of continuous DOC concentration estimates, more accurate carbon fluxes from the landscape can be determined and used to inform how these fluxes may respond to climate change. However, stream water chemistry varies with source and hydrologic events (Ågren et al., 2010; Avagyan et al., 2014; Laudon et al., 2011; Peacock et al., 2014; Tiwari et al., 2014; Vaughan et al., 2017) and thus, the relationship between DOC concentration and CDOM absorbance in stream water varies as different components of the landscape are connected with changes in hydrology. The PLS approach is tested here to demonstrate whether it captures the variability in the relationship between DOC concentration and CDOM absorbance in stream water. PLS allows for a more extensive range of the absorbance spectra to be used to predict DOC concentration, which may reduce the uncertainty around predictions, as more information is used to make the predictions; unlike MLR models where using multicollinear absorbance data to predict DOC concentration risks forming an unstable model with more uncertainty around predictions. Optical metrics related to CDOM composition and water quality parameters that change about discharge are also used with absorbance to improve accuracy of predicted DOC concentration. Uncertainty around predictions was estimated as 95% prediction intervals that take into account the variability in the relationship between CDOM absorbance and DOC concentration among samples and give confidence to predictions of future samples made by the model as well as catchment-wide flux estimates derived from these predictions.

2.4.1 Heterogeneity among samples in the training and validation datasets:

Variation in pH, SPC and [Fe], but not ORP, among soil waters is reflected in stream waters

Water quality parameters among sample types demonstrate that stream samples exhibit the largest variance in [Fe], pH, and SPC, consistent with the role of hydrology and water source changes with discharge and season in controlling stream water chemistry (Figure 2.5). Piezometer samples cover a wide range of [Fe] from 200 to 400 ppb, lysimeter samples range from 75 to 200 ppb (Figure 2.5 B), while stream samples range in [Fe] from 50 to 350 ppb, spanning the ranges across other sample types. Similarly, the SPC of lysimeter and piezometer samples exhibited a narrower range of 25 to 75 $\mu\text{S cm}^{-1}$ compared with the stream samples that ranged from 50 to 100 $\mu\text{S cm}^{-1}$ (Figure 2.5 A). The pH varied across all sample types with lysimeter samples ranging from 4.5 to 5.5 pH, piezometers ranging from 5 to 6 pH, groundwater wells ranging from 6 to 7 pH, and streams ranging from 5.5 to 7 pH, again spanning the range of values observed within the other sources (Figure 2.5 C). At this point, [Fe] was no longer considered a water quality parameter as the intention of the following PLS models was to estimate DOC concentration from *in-situ* data and continuous *in-situ* monitoring of [Fe] is not typically possible.

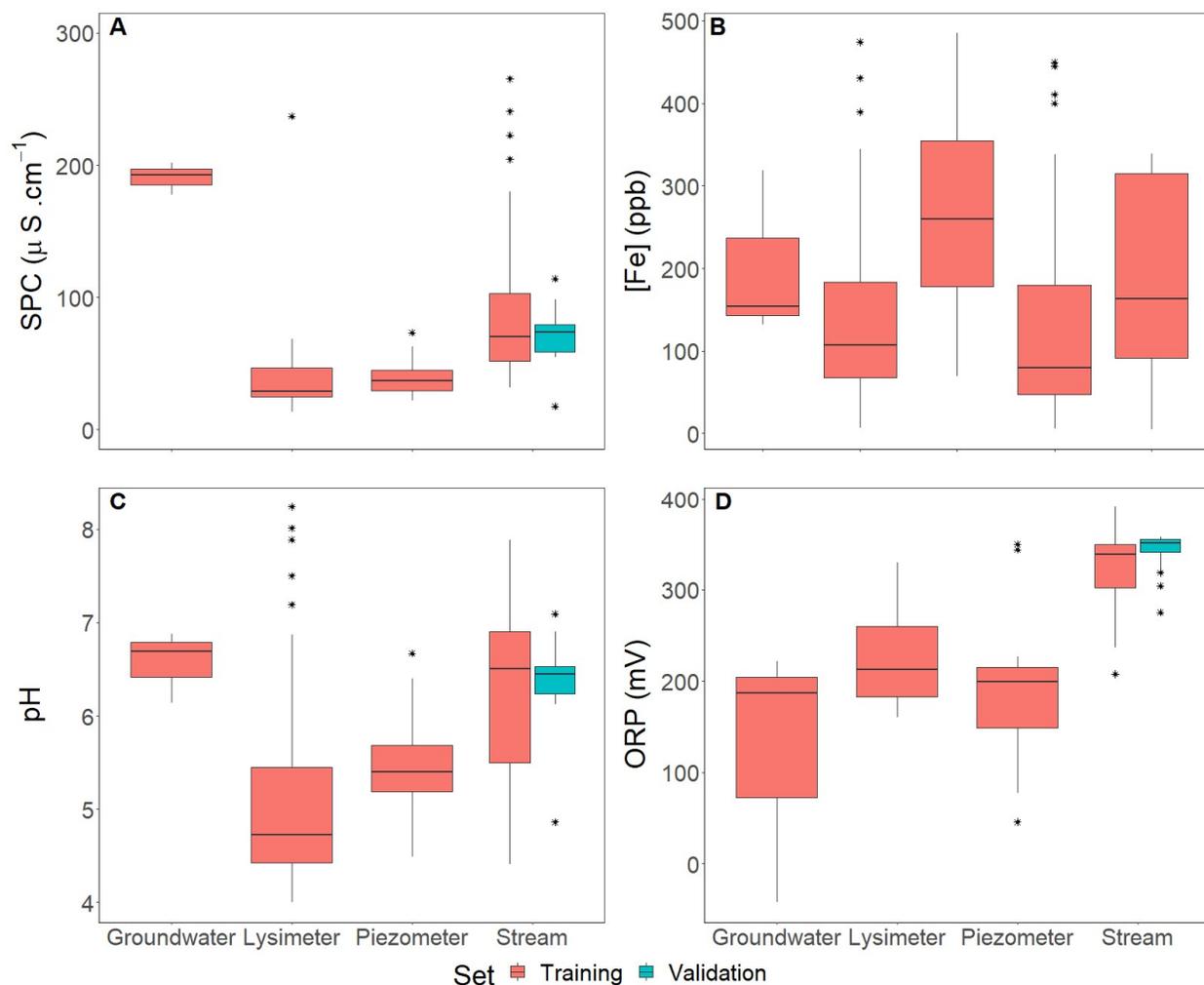


Figure 2.5: Distribution of water quality parameters for the training and validation datasets

Water quality parameters, specific conductance at 25°C (A; SPC), total dissolved iron (B; [Fe]), pH (C), and oxidation-reduction potential (D; ORP), across sample types for the model training and validation datasets. All stream and lysimeter samples from the training dataset were collected from across the four NL-BELT regions and the PBEWA, while groundwater, piezometer and most stream samples from the validation dataset were collected exclusively from the PBEWA. The asterisks refer to extreme outliers outside 1.5 times the interquartile range.

Models derived from this particular training dataset will not be applicable for *in-situ* data due to the addition of HgCl_2 to preserve samples. However, the method is equally applicable when using a new dataset where HgCl_2 was not used. The addition of HgCl_2 was observed to cause a drop of pH by as much as a whole unit across the samples collected in 2019. This was observed when the measurements for this subset of samples (those collected in 2019 from Table

2.2.) were made before and after the addition of HgCl_2 , the latter immediately following absorbance analysis. Considering that continuous *in-situ* stream pH measurements below pH 6 from these stream sites have rarely been observed, the addition of HgCl_2 likely contributed to the values observed below 6 (Figure 2.5) for the rest of the sample set (collected pre-2019). Similarly, the addition of HgCl_2 also increased the SPC by as much as $40 \mu\text{S cm}^{-1}$ across the same 2019 samples, suggesting the pre-2019 samples experienced the same effect. Even though these models are not applicable for *in-situ* data due to the altered pH, SPC and ORP, these methods provide a basis for model development and testing that can be repeated with future datasets.

In contrast to other water quality parameters, the ORP of the stream water samples was the highest and exhibited the least variation (300 to 350 mV) among sample types consistent with the turbulence, and therefore likely reaeration, experienced at each stream site (Figure 2.5 D). The most variable ORP was observed in groundwater well samples (-50 to 200 mV), as expected with changes in oxidizing or reducing conditions dependent on the water table variation over the year. Collectively, these results suggest that ORP measurements may not be helpful in further constraining the modelling of stream DOC from absorbance measurements as the ORP values of the sources are likely overprinted by the reaeration occurring in the stream.

The relationship between CDOM absorbance and DOC concentration is variable among all samples types

The concentration of DOC exhibited a wide range spanning 1 to 40 mg C L⁻¹ across all sample types (Figure 2.6 A). As expected, baseflow stream samples, in either the training or validation dataset, exhibited lower values and a narrower range in DOC concentration (1 to 15 mg C L⁻¹) compared to those collected during storm events (5 to 22 mg C L⁻¹; Figure 2.6 A). Among sample types, lysimeters were found to have the most extensive range in DOC concentration (5 to 40 mg C L⁻¹; Figure 2.6 A). Derived from water percolated through overlying organic soil, this variation in lysimeter DOC concentration is due to both water fluxes and organic soil compositional changes throughout the year (Bowering et al., 2020). Samples taken from deeper mineral soil horizons exhibit lower maximum concentration of DOC, with deeper piezometers ranging from 5 to 25 mg C L⁻¹ and the groundwater wells ranging from 5 to 10 mg C L⁻¹ (Figure 2.6 A). During baseflow, stream DOC concentration was closer to those of the deeper soils (groundwater wells and piezometers), while during hydrologic events, stream DOC concentration was closer to values found for surface organic soils (lysimeters).

Absorbance at 350 nm (A_{350}) was found to have a very similar trend to DOC concentration among all sample types (Figure 2.6 A-C). This was expected, as these two metrics are typically highly correlated (Peacock et al., 2014), with A_{350} sometimes used as a proxy for DOC concentration (Asmala et al., 2012; Muller & Tankéré-Muller, 2012; Peacock et al., 2014). However, that is not the case with S_R (Figure 2.6 D), which seems more independent of DOC concentration as it relates to CDOM composition (Helms et al., 2009; Weishaar et al., 2003).

Stream samples were found to have the largest variance in these optical characteristics compared to other sample types, likely resulting from changing proportions of different sources, including those represented by the sample types measured here. For example, the lower $SUVA_{350}$ values in stream samples typical of those collected during baseflow are consistent with groundwaters, while higher values typically of those collected during storm flow are consistent with the piezometers and lysimeter samples (Figure 2.6 C). On the other hand, higher S_R values (1-1.2) were observed in the groundwater well samples (deepest depth), lower S_R values (0.7-0.9) were observed in the lysimeter samples (shallowest depth) and piezometer samples (intermediate depth) and stream samples (variably integrating soil waters) were observed to range between those extremes (0.8 to 1.2) regardless of base or storm flow conditions (Figure 2.6 D).

One of the caveats with using a single wavelength to train a model to predict DOC concentration is that it may result in a model that is biased to predict one sample type more accurately than another: a result of varying contributions of CDOM, or varying components of CDOM, among sample types (Peacock et al., 2014). Although A_{350} and concentration of DOC are highly correlated and show near-identical distributions across sample types, the overall relationship between DOC concentration and A_{350} among sample types were found to be significantly different from each other, observed as an interaction effect of A_{350} with sample type ($F_{(3, 276)} = 21.4, p < 0.0001$; Figure 2.7) in an ANCOVA regressing DOC concentration against A_{350} , sample type, and their interaction. This suggests that the aromatic content, and therefore overall CDOM composition, among the different sample types, is measurably distinct, and using absorbance from a single wavelength (A_{350}) to predict DOC concentration of these samples would most likely inaccurately predict specific types of sample.

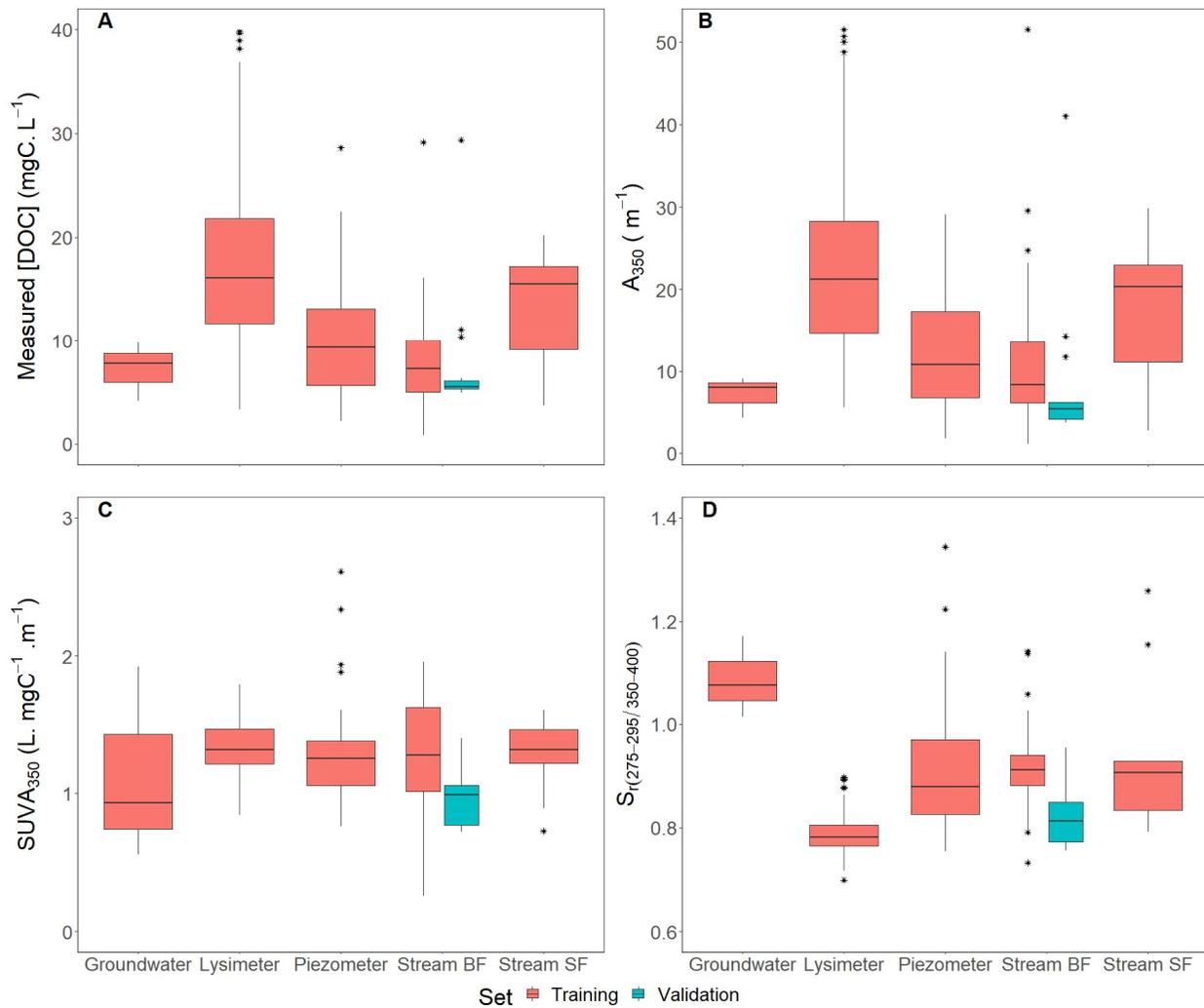


Figure 2.6: Bulk DOC and optical parameters related to composition

Measured dissolved organic carbon concentration (A; [DOC]) and absorbance characteristics including absorbance at 350 nm (B; A_{350}), specific UV absorbance at 350 nm (C; $SUVA_{350}$) and spectral slope ratio (D; S_r) across sample types for the model training and validation datasets. The stream samples collected during baseflow (Stream BF) are provided separately from stream samples collected during storm events (Stream SF). All stream and lysimeter samples were collected from across the four NL-BELT regions and the PBEWA, while groundwater and piezometer samples were collected exclusively from the PBEWA. The asterisks refer to extreme outliers outside 1.5 times the interquartile range.

Also, following a one-way ANOVA where the effect of sample type was marginally significant on $SUVA_{350}$ ($F_{(3,280)} = 2.49$, $p < 0.10$), a post-hoc Tukey's honest significant difference test revealed that the difference in slopes between stream and lysimeter samples were significantly different ($p < 0.05$). These results are observed in Figure 2.7, where the slopes of

each sample type differs, the largest difference in slope is observed between the stream and lysimeter samples. However, these analyses highlight that the relationship between DOC concentration and A_{350} can be variable even within the same sample type. For example, stream samples within a narrow range of DOC concentration (9 to 12 mg C L⁻¹) were associated with a wide range of A_{350} values (5 to 30 m⁻¹) (Figure 2.7, Appendix 1). Similarly, as DOC concentration increases from 20 to 40 mg C L⁻¹, the spread of A_{350} values for lysimeter and piezometer samples becomes wider (20 to 50 m⁻¹) (Figure 2.7). The relationships between DOC concentration and A_{350} observed here (Figure 2.7) are not necessarily what would be observed for a different wavelength in place of 350 nm, due to differences in DOM composition among sample types (Peacock et al., 2014) and even within the same sample type (Appendix 1). This is an example that highlights how the use of single-wavelength models to predict concentration of DOC could result in more uncertainty around predictions than a multi-wavelength model (Avagyan et al., 2014), where Appendix 1 highlights the variable relationship between DOC concentration and A_{350} within stream samples alone, most likely attributed to the region where the sample was collected from.

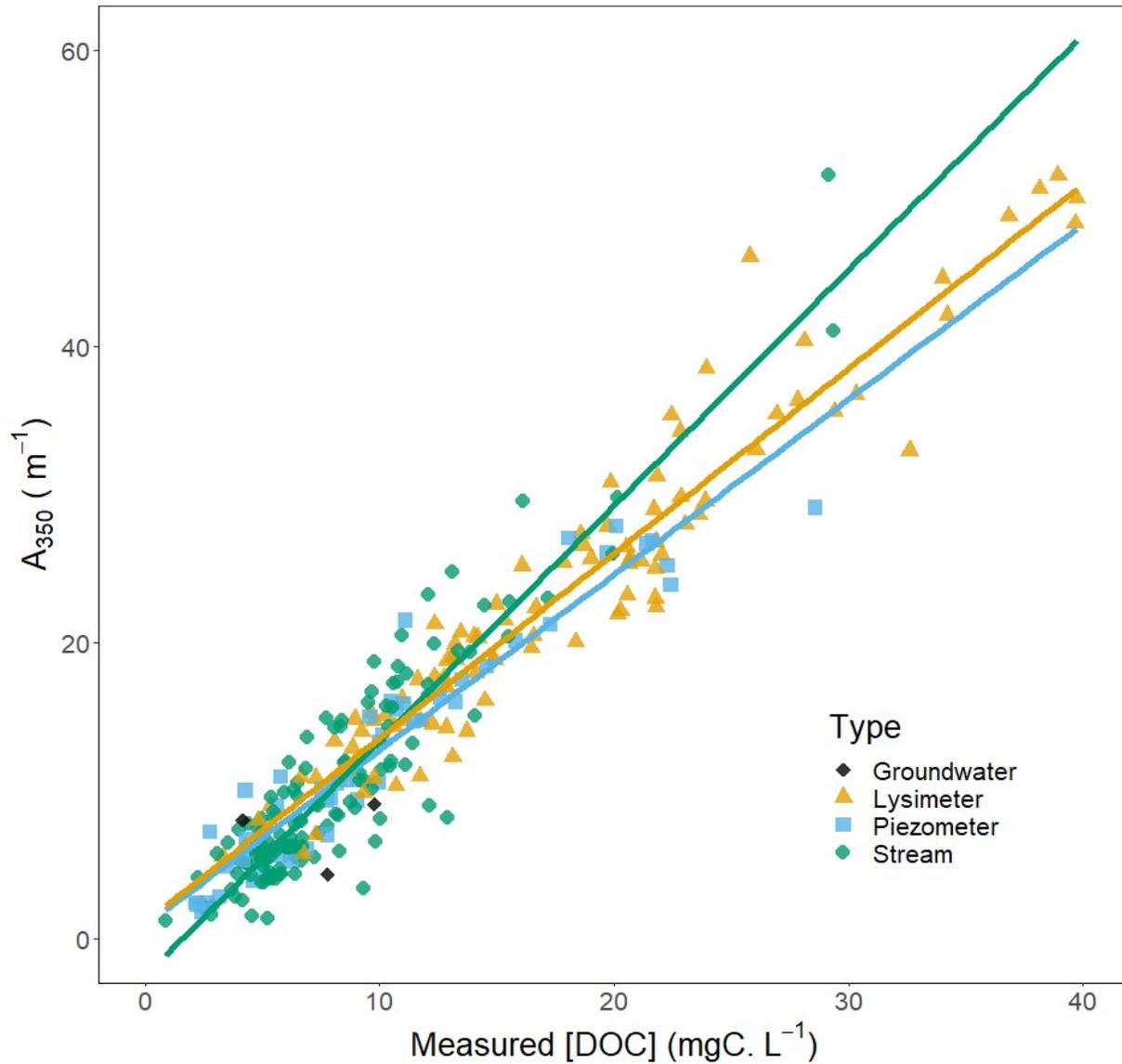


Figure 2.7: Relationship between A₃₅₀ and DOC

The relationship between absorbance at 350 nm (A₃₅₀) and dissolved organic carbon concentration ([DOC]) across sample types. Slopes of best-fit lines from stream and lysimeter samples were found to be significantly different ($p < 0.05$) as assessed through a post hoc Tukey's test on a one-way ANOVA testing SUVA₃₅₀ ([DOC]-normalized A₃₅₀) against sample type. The ANCOVA test of A₃₅₀, [DOC] and sample type exhibited a significant sample type and interaction terms ($p < .0001$).

2.4.2 PLS models performed as good as typical MLR models:

The d_r was used as an indicator of model performance to replace R^2 and prediction intervals were calculated to constrain uncertainty around predictions. These metrics are typically not reported in the literature dealing with the spectrophotometric determination of DOC concentration (Asmala et al., 2012; Avagyan et al., 2014; Etheridge et al., 2014; Fichot & Benner, 2011; Jollymore et al., 2012; Juhls et al., 2019; Lee et al., 2015; Peacock et al., 2014; Ruhala & Zarnetske, 2017; Simonsson et al., 2005; Strohmeier et al., 2013). Prediction intervals from DOC estimates can then be applied to catchment-wide flux estimates. Across all models, NSE values were found to be equivalent to R^2 and therefore d_r was considered the primary metric to compare model performance along with prediction intervals.

Table 2.5: Comparing PLS to MLR models

Comparison of three models generated using the training dataset and tested using a separate validation dataset of this study. The Avagyan et al. (2014) model is a multiple linear regression using absorbance at wavelengths 275, 295 and 325 nm selected through stepwise regression. The Fichot & Benner (2011) model is based upon a multiple linear regression of log-transformed absorption coefficients at 275 and 295 nm, where the median absorption split the dataset at 275 nm. The PLS-plain is the PLS regression developed in this study using absorbance across the spectrum of 275-450 nm at 5 nm intervals. Models are evaluated by comparing the index of agreement (d_r), root mean square error (RMSE) and RMSE of prediction (RMSEP) and the 95% prediction interval (95% PI). Asterisk - the predicted [DOC] from the PLS-plain model was used to develop the models in Table 2.7.

Model	Training dataset				Validation dataset			
	n	df Model	RMSE	95% PI	n	RMSEP	NSE	d_r
Avagyan <i>et al.</i>	254	4	1.92	3.7	33	0.56	0.98	0.89
Fichot & Benner	254	3	2.06	4.2	33	0.66	0.98	0.86
PLS-Plain*	254	5	1.87	3.7	33	0.52	0.99	0.89

Avagyan, A., Runkle, B. R., & Kutzbach, L. (2014). Application of high-resolution spectral absorbance measurements to determine dissolved organic carbon concentration in remote areas. *Journal of hydrology*, 517, 435-446.

Fichot, C. & Benner, R. (2011). A novel method to estimate DOC concentrations from CDOM absorption coefficients in coastal waters. *Geophysical Research Letters*. 38:L03610

Prediction accuracy was similar between the PLS-plain model developed here and each of the MLR models developed using the methodology from Fichot & Benner (2011) and Avagyan *et al.* (2014) (Table 2.5). All models had $d_r \geq 0.86$. This indicates that this PLS approach works as well as previously developed methods using multiple discrete wavelengths. However, the PLS approach has the added benefit of providing more information about the dataset and resulting model that allows for further interpretability and comparability among model iterations and results. This is achieved by utilizing the scores, weights, and loadings provided by the analysis. Additionally, the model coefficients result in being more stable, which means that minor changes in the dataset do not result in significant changes in the model coefficients. Future samples need not have a near-identical covariance structure as the training dataset in order to make accurate predictions of DOC concentration. This contrasts with MLR when using multiple wavelengths that exhibit multicollinearity, resulting in unstable models with highly sensitive covariance structure and inferior performance when predicting samples that do not have the same covariance structure as the training dataset.

The model resulting from the PLS regression of DOC concentration on $A_{275-450}$ was able to predict $\text{DOC} \pm \sim 3.7 \text{ mg C L}^{-1}$ with 95% confidence using only five components (Table 2.5, Figure 2.8 A). The validation dataset testing this PLS model had a d_r value of 0.89, supporting that the model can make very accurate predictions (Willmott *et al.*, 2012). The 3.7 mg C L^{-1} prediction

interval is helpful in estimating a constraint on uncertainty and applicable when utilizing the predictions for catchment-wide DOC flux estimates. However, due to prediction intervals not typically reported in the literature, it is unknown how this estimate of uncertainty fares to other studies with similar goals. The model was able to accurately predict DOC concentration of validation samples ranging from 4-30 mg C L⁻¹ (Figure 2.8 B), with all the residuals from the prediction ($[\text{DOC}]_{\text{Res}}$) of the validation dataset falling well within the ~ 3.7 mg C L⁻¹ prediction interval ($|[\text{DOC}]_{\text{Res}}| < 3.7$ mg C L⁻¹); the largest $|[\text{DOC}]_{\text{Res}}|$ (absolute residual) observed was 1.06 mg C L⁻¹.

Considering that the mean square error (MSE) determines the prediction interval's size (while keeping all other parameters constant; Eq. 12), the estimated ~ 3.7 mg C L⁻¹ prediction interval could have been inflated as squaring errors increase the influence of outliers. Re-fitting the model after examining the outliers or replacing the training dataset to train a new model using stream samples from the PBEWA from storm events to develop a more refined model can result in a more accurate model with lower MSE and, therefore, smaller prediction intervals. Otherwise, an alternative leverage-based prediction interval can be derived and more suitable to be used for constraining prediction error from PLS models (Høy et al., 1998).

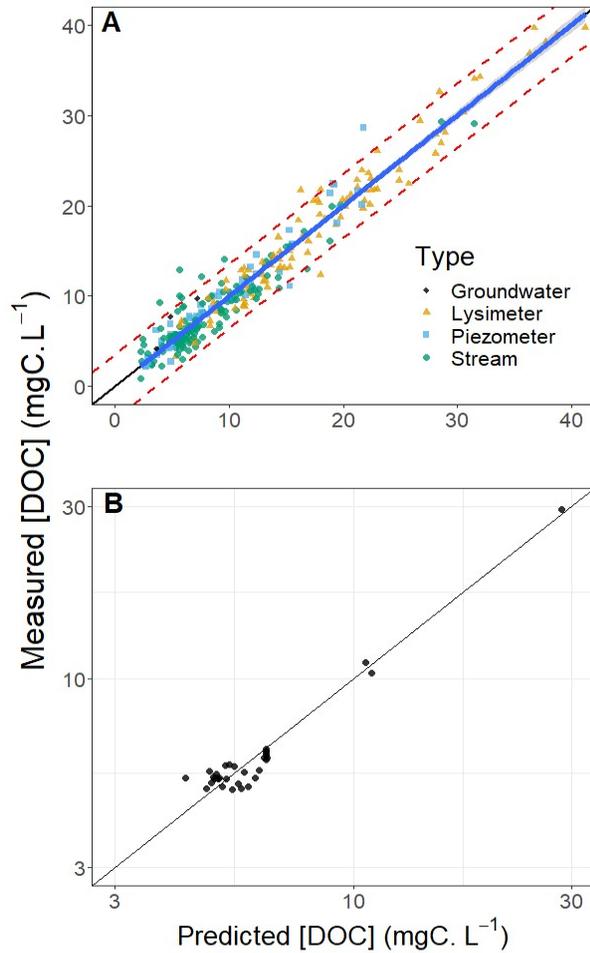


Figure 2.8: Recovery function for PLS-Plain model

Recovery function of the PLS-plain model (using only wavelengths from 275-450 range) across samples from training dataset ($n = 254$; A) and validation dataset ($n = 33$; B) expressed as the model predicted dissolved organic carbon concentration ([DOC]) plotted against lab measured [DOC]. All stream and lysimeter samples from the training dataset were collected from across the four NL-BELT regions and the PBEWA, while groundwater, piezometer and most stream samples from the validation dataset were collected exclusively from the PBEWA. Line of best-fit (A; $y = 1.0007x - 0.0080$) overlaps with 1:1 line representing perfect predictions. The shaded area (A) represents 95% confidence interval of best-fit regression. Dashed lines (A) mark the 95% prediction interval range. Solid 1:1 line (B) indicates perfect predictions. Log-scaled axes (B) highlight model performance in the lower [DOC] range expected during baseflow conditions ($\sim 5 \text{ mg C L}^{-1}$).

2.4.3 Additional optical and water quality parameters did not improve model performance:

The determination of an accurate model that predicts DOC concentration by using CDOM absorbance from continuous *in-situ* stream measurements and constraining its uncertainty around predictions is imperative to accurately estimate DOC export across all spatial and temporal scales within catchments with confidence. Models that use only absorbance to predict DOC concentration are suitable. However, the relationship between DOC and CDOM absorbance is altered during storm events, likely due to changes in DOC source caused by changes in hydrology (Asmala et al., 2012; Vaughan et al., 2017, 2019). Therefore, optical parameters (spectral slopes and their ratio) and water quality parameters (pH, SPC and ORP) were tested to determine if they would improve predictions of DOC concentration and reduce uncertainty. Because future predictions using these parameters will be from *in-situ* stream measurements of pH, SPC and ORP, the training dataset used to test the water quality parameters was composed of a subset of only stream samples to form the main training dataset (Table 2.2 and 2.3).

Using optical parameters in concert with CDOM absorbance in PLS to predict DOC concentration

The inclusion of integrative optical parameters, such as $S_{275-295}$, $S_{350-400}$, or S_R , in additional PLS regressions, did not improve DOC prediction uncertainty. The 95% prediction intervals for the four PLS models (Table 2.5 and 2.6) varied between ~ 3.6 and ~ 3.8 mg C L⁻¹. In fact, the

inclusion of $S_{350-400}$ increased the validation dataset's prediction error, where the RMSEP was double that of the other PLS models, and d_r was 0.1 smaller than the other models. Considering that the PLS- $S_{350-400}$ model had a marginally smaller RMSE but much larger RMSEP while using seven components (Table 2.6) compared to the five components used by the PLS-plain model (Table 2.5), suggests that the PLS- $S_{350-400}$ model is overfit (Figure 2.3).

Table 2.6: Including other optical parameters as predictors

Comparison of models using additional absorbance metrics related to dissolved organic carbon composition to test their impact on the original PLS model developed in this study. All models are generated using the training dataset and tested using a separate validation dataset of this study. These PLS regression models all use $A_{275-450}$ alongside an additional variable: one of either spectral slopes ($S_{275-295}$ or $S_{350-400}$) or the spectral slope ratio (SR). Models are evaluated by comparing the index of agreement (d_r), root mean square error (RMSE) and RMSE of prediction (RMSEP) and the 95% prediction interval (95% PI).

Model	Training dataset				Validation dataset			
	n	df Model	RMSE	95% PI	n	RMSEP	NSE	d_r
$S_{275-295}$ PLS	254	5	1.87	3.7	33	0.53	0.99	0.89
$S_{350-400}$ PLS	254	7	1.8	3.6	33	1.08	0.95	0.78
S_R PLS	254	6	1.85	3.8	33	0.59	0.98	0.88

Even though including these additional optical parameters were found to not improve predictions of DOC concentration and reduce the uncertainty of the models using this specific dataset, they could be potential variables for reducing model uncertainty in future studies. Especially when considering developing a model for *in-situ* stream predictions, as these metrics hold information about the DOM composition (Helms et al., 2009) and therefore, the relationship between these metrics and DOC may capture information relevant to hydrologic events (Vaughan et al., 2017, 2019) that may be useful for improving predictions of DOC concentration and

reducing model uncertainty. However, this training dataset is quite heterogeneous due to samples from different sources of water from different catchment areas with widely different CDOM compositions (Table 2.1, Figure 2.6). This heterogeneity may be aliasing the contribution of the stream sample optical parameters and desensitizing the model stream sample signatures. In other words, the model may be too general and desensitized to catchment-specific signatures that could improve model performance. Thus, it appears as if there was no improvement in prediction accuracy for the validation dataset composed of mostly baseflow samples from specific sites within the PBEWA (validation dataset in Figure 2.6 compared to training dataset). This highlights the possible pitfalls one may encounter when using a global predicting model that is insensitive or overlooks site-specific signatures that may be important for making an accurate prediction within a site.

Further exploration of how these optical parameters contribute to predictions of DOC concentration and uncertainty should be considered when training a model using a training dataset of samples more representative of what the model will be used to predict. In this case, samples collected over baseflow and storm-flow conditions from Horseshoe Brook within the PBEWA.

Using water quality parameters in concert with priorly-predicted DOC concentration in MLR to predict DOC

Using only stream samples from the training dataset, the predictions of DOC concentration from the Plain-PLS model were used in concert with one of pH, SPC or ORP in an MLR (Table 2.3) in an attempt to improve prediction accuracy of DOC concentration from stream samples. This step was taken as future *in-situ* data would be collected in stream water and not lysimeters or piezometers. Therefore, the developed models should be informed by the pH, SPC, or ORP signatures observed typically in a stream. Nevertheless, due to most stream samples in the training dataset being from baseflow conditions, there is minimal variability in the relationship between CDOM absorbance and DOC concentration resulting from different landscape components connecting to the streams during storm events. As such, lysimeter and piezometer samples are used instead to represent end-member sources of DOM that would contribute to stream water during such events. However, there was no further improvement to prediction accuracy of DOC concentration despite a reduction in the 95% prediction interval (from ~ 3.7 to ~ 2.8 - ~ 3.0 mg C L⁻¹, Table 2.5 and 2.7).

The prediction interval for the MLR models is not able to incorporate the uncertainty associated with the original PLS-predicted DOC concentration (~ 3.7 mg C L⁻¹) used in the MLR and, therefore, underestimate the uncertainty.

Despite the lack of clear evidence of model improvement in the comparisons made here, pH and SPC may still be good candidates to improve model accuracy, while ORP remains a weaker candidate for inclusion in further model development for stream DOC prediction. The MLR

models' performance incorporating these water quality parameters has only been validated for baseflow stream samples, where an event would lead to a much more extensive range of observed values of these water quality parameters. The lack of model improvement with the inclusion of these water quality parameters for DOC prediction was likely due to reduced ranges in the SPC and pH values within the training dataset relative to the validation dataset (Figure 2.5). Furthermore, this range was not likely reflective of the range in SPC and pH typically found over hydrologic events. On the other hand, due to the signature of ORP from different sample types being overprinted within the stream because of turbulence and mixing, ORP was the weakest candidate to continue testing in the future (Figure 2.5 D).

Stream water pH and SPC are still considered better candidates to improve model accuracy due to their relationship with discharge and DOC concentration, and by extension absorbance (Ågren et al., 2010; Inerillo et al., 2017; Kobayashi, 1986; Laudon et al., 2001; Laudon & Buffam, 2008; Pace et al., 2012; Spencer et al., 2007). The stream water pH, for example, is negatively correlated to DOC concentration (Ågren et al., 2010; Laudon et al., 2001; Laudon & Buffam, 2008) as organic acids often represent a significant component of DOM delivered to streams, lowering pH while contributing to increasing DOC content. However, CDOM, a measure of DOM's absorptivity, is positively correlated to pH (Pace et al., 2012; Spencer et al., 2007). This causes an interference in the relationship between DOC concentration and CDOM absorbance during storm or snowmelt events where stream water pH has been observed to decrease during the rising limb and increase during the lowering limb of these melt events (Ågren et al., 2010; Laudon et al., 2001; Laudon & Buffam, 2008) or storm events.

Table 2.7: Testing performance of water quality parameters

Comparison of models using a water quality parameter to test their impact on [DOC] prediction accuracy. All models are generated using the same training dataset, a subset of the training dataset used for Table 2.1 and Table 2.2 composed of only stream samples and tested using a separate validation dataset of this study. These multiple linear regression models use predictions of DOC concentration from the PLS-plain model plus one of the factors to predict [DOC]. Models are evaluated by comparing the index of agreement (d_r), root mean square error (RMSE) and RMSE of prediction (RMSEP) and the 95% prediction interval (95% PI). Asterisk – the +5 df represents the 5 PLS-components used to estimate $[DOC]_{Pred}$, which [DOC] is then regressed onto for these water quality parameter models (Figure 2.2, Table 2.3).

Model	Training dataset				Validation dataset			
	n	df Model*	RMSE	95% PI	n	RMSEP	NSE	d_r
No Factor	106	1 (+5)	1.89	3.3	33	0.79	0.97	0.88
pH	81	2 (+5)	1.82	3.0	31	1.22	0.94	0.87
ORP	39	2 (+5)	2.08	3.0	31	1.2	0.94	0.81
SPC	42	2 (+5)	1.75	2.8	31	1.46	0.91	0.79
C	51	2 (+5)	1.67	2.7	31	1.66	0.89	0.83
Fe	93	2 (+5)	1.8	3.3	0		NA	

2.4.4 PLS model biased by region but not season:

Model performance bias was assessed through six one-way ANOVAs testing the prediction residuals of the PLS-plain and both MLR models, using Fichot & Benner’s (2011) and Avagyan et al.’s (2014) methods (Table 2.5), against sample season and region (Figure 2.9). The analyses revealed that all three models might be biased by region while only the model using Fichot & Benner’s (2011) method may be biased by season. The season could not explain the distribution of residuals from the PLS-plain model ($p > 0.3$), suggesting prediction errors are not biased strongly biased by season. The same test performed on region, however, indicated a regional bias for the prediction errors ($p < 0.01$), suggesting there is a regional specificity to the

relationship between CDOM absorbance and DOC concentration that I fail to capture in all three regression methods used (Table 2.5). However, that does not mean the models are not suitable to be used across different regions, as the $\sim 3.7 \text{ mg C L}^{-1}$ prediction interval takes these biases into account. It does, however, suggest that the prediction intervals could be reduced by using a training dataset composed of samples exclusively from the planned site or region, removing prediction error or bias caused by regionality but within similar ecosystem types based upon dominant vegetation and overall climate regime.

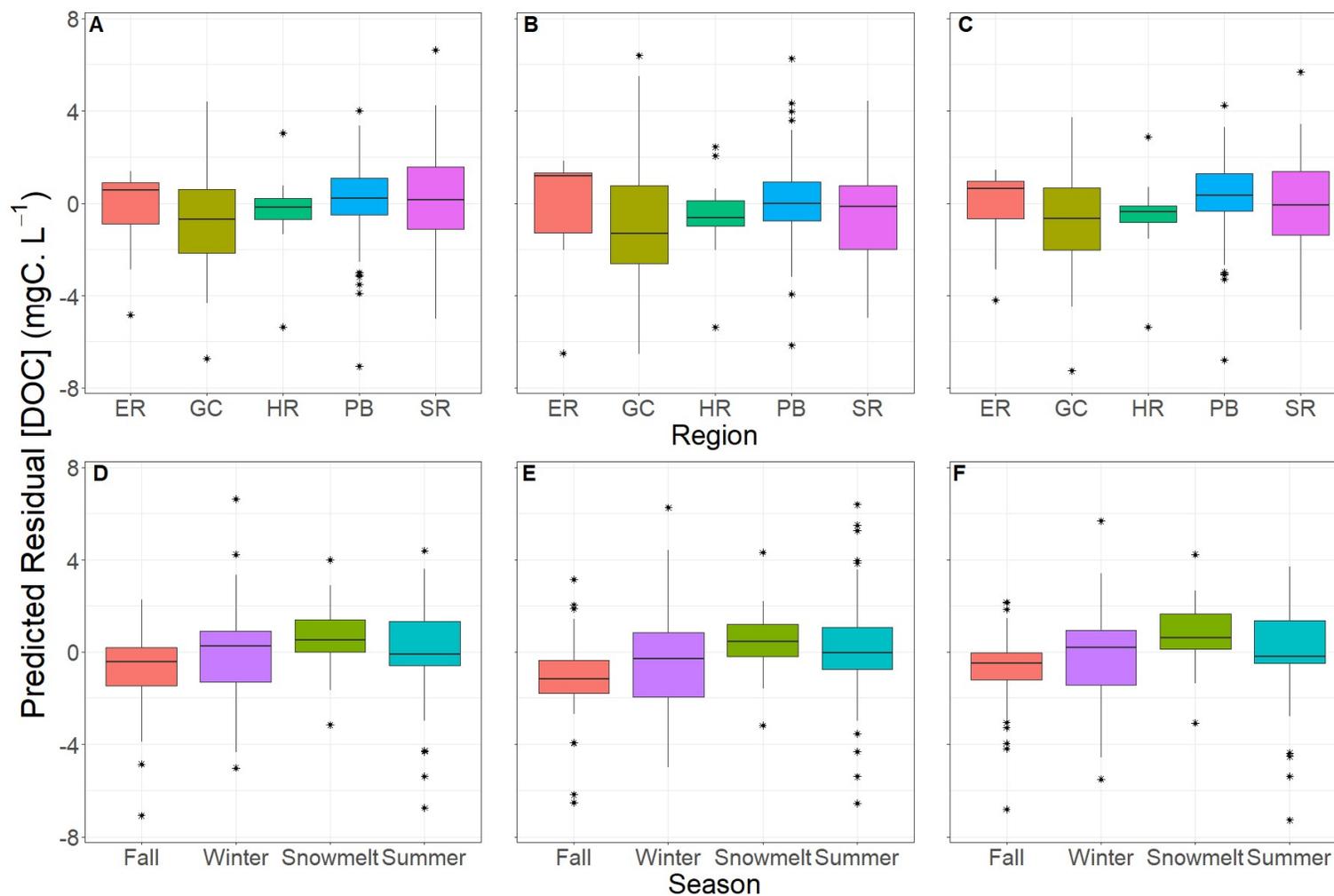


Figure 2.9: Sensitivity analysis testing season and region

Model estimate residual distributions across NL-BELT and the PBEWA regions (top, A, B, C) and seasons (bottom, D, E, F) for the Avagyan (left, A, D), Fichot & Benner (middle, B, E) and PLS-plain (right, C, F) models. One-way ANOVAs testing each of the model residuals to regions or seasons found season to be significant only in the Fichot & Benner methods MLR model residuals ($p < 0.001$) and found region to be significant in all models ($p < 0.001$). The asterisks refer to extreme outliers outside 1.5 times the interquartile range.

2.5 Conclusion and Implications

Accurate determination of DOC concentration using only absorbance was achieved along with an estimate of uncertainty around predictions made by these PLS and MLR models. Prediction accuracy and uncertainty from PLS models were not different from MLR models, indicating they are adequate for building absorbance-based models for predicting *in-situ* DOC. Further testing of optical metrics related to DOC composition ($S_{275-295}$, $S_{350-400}$ and S_R), pH and SPC using a dataset comprised of storm samples from only Horseshoe Brook in PBEWA may improve the model. These parameters ($S_{275-295}$, $S_{350-400}$, S_R , pH and SPC) may be more beneficial when a representative dataset is used to train the model. Such a dataset would capture full storm events, capturing a great proportion of the natural variability in relationships between CDOM absorbance, DOC concentration, SPC, pH, and CDOM composition throughout a storm. Using metrics, such as variable importance in the projection (VIP), squared prediction errors (SPE) and Hotelling's T^2 , made available from PLS model's scores and loadings, could provide more insight regarding whether these parameters can help improve model predictions and reduce uncertainty in predictions of DOC concentration from *in-situ* measurements. Scores used for diagnostic plots such as SPE and T^2 plots can be used to check for dataset or sample consistency (Dunn, 2019) which can be used in further model refinement and help decrease prediction uncertainty. Furthermore, SPE and T^2 plots can identify unusual outlier samples in the training dataset that may be biasing the model or even identify unusual future samples (Dunn, 2019). Scores can be further used to calculate sample

leverages to identify high-leverage samples that would bias the model or also be used to estimate leverage-based prediction intervals (Høy et al., 1998). SPE and T^2 plots can also help determine when the model should be updated or re-fit with more appropriate data if future samples consistently fail to have similar SPE or T^2 results to the training dataset (Dunn, 2019) separately from prediction error. Future steps should involve using these plots when applying the model to a dataset of continuous *in-situ* absorbance measurements before predicting DOC concentration to inform whether the model is suitable to predict a new sample.

As the model is meant to be used to predict stream DOC frequently over long periods, the model was developed further with the inclusion of optical metrics and water quality parameters affected by stream discharge. Optical parameters, like $S_{275-295}$, $S_{350-400}$ and S_R , and water quality parameters, such as pH, ORP and SPC, were found not to improve prediction accuracy of DOC concentration or reduce uncertainty when added to regression models using the dataset of this study. This was most likely due to the high diversity in sample types within the training dataset which reduced representation, or aliased, the range of stream water chemistry indicative of the entire range of hydrologic conditions. Future datasets should use stream samples collected across hydrologic events where the full range of these parameters and DOM compositions can be observed, tested and used to reduce the uncertainty around predictions of stream DOC concentration. These models may not provide the resolution in predicted DOC concentrations needed to inform accurate DOC flux estimates considering the relatively large uncertainty in

prediction intervals. However, these results do indicate the PLS modelling approach is promising, and considering the importance of the accuracy of the DOC flux, next steps in developing this model from *in-situ* stream data capturing storm events is warranted.

3 Training and evaluating a PLS model to predict dissolved organic carbon concentration in remote boreal forest stream sites

3.1 Abstract

Accurate quantification of dissolved organic carbon (DOC) fluxes from terrestrial to aquatic systems is imperative for understanding landscape and downstream aquatic responses to climate change. Accurately quantifying DOC dynamics requires high-temporal resolution data not feasible via discrete *in-situ* sampling. Continuous *in-situ* spectrophotometers are increasingly being used to monitor stream catchments and obtain high-frequency estimates of DOC concentration modeled using chromophoric dissolved organic matter (CDOM) absorbance. The accuracy of predictions of DOC concentration is improved through better representation of changes in the relationship between CDOM absorbance and DOC concentration relationship. This is particularly

important for streams where DOC composition is highly dynamic due to variation in hydrologic connectivity, creating changes in DOC sources and levels of processing (e.g. physiochemical or biological) occurring within the landscape. In this chapter, *in-situ* CDOM absorbance, water level (WL), specific conductance (SPC) and optical-derived indicators of DOC composition were used as predictors to train partial least squares (PLS) models. A discrete dataset capturing baseflow and stormflow conditions from two different sites along a boreal headwater stream was used to train a PLS model that integrates these qualities and data for bulk predictions of DOC concentration. PLS was employed here as it can use multicollinear information from the entire absorbance spectra, which contains information about CDOM composition, to enhance predictions of DOC concentration. Furthermore, consistency checks built into the PLS model, Hotelling's T^2 and square prediction errors (SPE), provide a means to gauge model performance when used with *in-situ* data and identify outlier observations incompatible with the model. A series of test-set switches and randomized cross-validations were performed to identify potential PLS-components to use when there is no validation dataset available. An absorbance-based measurement of light-scatter, average absorbance from 700-730 nm (average $A_{700-730}$), was used to develop a correction for *in-situ* particle-absorbance (turbidity) from field measures and tested on a continuous *in-situ* dataset. The turbidity correction performed poorly as it overcompensated during some periods and likely because of the variability of *in-situ* turbidity-induced absorbance not represented in the training dataset. The scatter correction (average $A_{700-730}$) was found to perform better

across the continuous *in-situ* dataset. Training dataset had a root mean square error of 0.36 mg C L⁻¹ when using 8 PLS components over a range of 4 to 16 mg C L⁻¹. The consistency checks for this model were used to identify data gaps in the model and determine the training dataset's representativeness when the model is applied to new *in-situ* data. This provides an approach for evaluating the model's performance when applied to new data by identifying and diagnosing outlier observations, for example identifying a need for more spring and winter samples to further update and inform the model for predicting DOC in two study sites. Furthermore, these statistics can be used to identify when and how a training dataset should be updated to maintain accurate prediction of DOC concentration when re-training the model for future datasets.

3.2 Introduction

Continuous *in-situ* optical measurements are increasingly being used to monitor some aspects of stream water quality, including at remote sites (Codden et al., 2020; Etheridge et al., 2014; Jollymore et al., 2012; Strohmeier et al., 2013; Vaughan et al., 2017, 2019; Zhu et al., 2020). Coupled with discrete sampling, *in-situ* absorbance is used to develop models that use these continuous optical measurements to predict various stream water components (e.g. dissolved organic carbon or nitrate; Avagyan et al., 2014; Etheridge et al., 2014; Langergraber et al., 2003; Vaughan et al., 2017). The use of the high temporal-resolution data associated with these continuous *in-situ* instruments is beneficial in monitoring headwater streams, intimately connected to their surrounding

landscape and quite responsive to short-term changes in hydrology from the surrounding landscape. At high temporal-resolution, measurements of essential forms of stream carbon or nitrogen can inform their source and pathway of delivery from the landscape, essential for predicting terrestrial to aquatic fluxes (Creed, McKnight, et al., 2015; Creed & Band, 1998).

Dissolved organic carbon (DOC) represents an essential fraction of carbon that is transported from terrestrial landscapes to aquatic ecosystems (Alvarez-Cobelas et al., 2012; Battin et al., 2009; Cole et al., 2007; Neff & Asner, 2001; Olefeldt et al., 2013; Raymond et al., 2016; Raymond & Sayers, 2010) and at potentially increased rates in northern landscapes with ongoing climate change (Jennings et al., 2010; Laudon et al., 2012; Worrall et al., 2004; Ziegler et al., 2017). The concentration of DOC can be accurately predicted from the absorbance of chromophoric dissolved organic matter (CDOM) as demonstrated in both wetland catchments and terrestrially influenced marine environment (Asmala et al., 2012; Avagyan et al., 2014; Fichot & Benner, 2011; Peacock et al., 2014) via a trained model.

3.2.1 DOC in headwater streams is dynamic

Due to the intimate connection between headwaters and their surrounding landscape, compositions of DOC vary as a function of hydrology, both seasonally and by event, due to variation in hydrologic pathways (Creed, Hwang, et al., 2015; Creed,

McKnight, et al., 2015; Creed & Band, 1998; Senar et al., 2018; Werner et al., 2019). As a result, the relationship between DOC concentration and CDOM absorbance can be site-specific and temporally variable within the same site (Helms et al., 2009; Peacock et al., 2014; Vaughan et al., 2019). It has been demonstrated how different sources of water from the same catchment can each have a unique optimum wavelength to predict DOC concentration and a uniquely different relationship with DOC at single wavelengths across the entire CDOM spectra (Peacock et al., 2014). This variability may be significant in streams draining hillslope-dominated landscapes as opposed to more low-relief landscapes dominated by wetlands because of the variety of possible pathways, and thereby sources of DOC linked to hydrology (Ågren et al., 2008; Hinton et al., 1998; Laudon et al., 2011; McGlynn & McDonnell, 2003).

As such, more straightforward model training methods that use absorbance from discrete wavelengths to predict DOC concentration (Asmala et al., 2012; Avagyan et al., 2014; Fichot & Benner, 2011; Peacock et al., 2014) may fail to predict DOC accurately from dynamic headwater streams. Different parts of the CDOM absorbance spectra can be related to DOC's different compositional properties (Helms et al., 2009). Therefore, using the entire CDOM spectra, thereby integrating any information in the CDOM spectra related to DOC composition, can result in more accurate predictions of DOC concentration from the streams where DOC's source and composition are temporally dynamic. However, due to the high number of wavelengths and multicollinear nature of absorbance data, typical multiple linear regression (MLR) methods (Asmala et al., 2012;

Avagyan et al., 2014; Fichot & Benner, 2011; Peacock et al., 2014) have limited application in predicting DOC from the vast information captured in absorbance spectra as they can potentially overfit the model and lead to an unstable model.

3.2.2 Using the full CDOM absorbance spectra to predict DOC concentration

An unstable overfit MLR model would have a limited ability to predict DOC concentration. Such a model would accurately predict only observations with a near-identical covariance structure among measurements as training dataset samples and, therefore, highly sensitive to random noise, which minimally alters the covariance structure and is observed within *in-situ* data. To remedy this, partial least squares (PLS) regression is used to train a model from high-dimensional, multicollinear data. PLS handles these problems by reducing dimensionality to create uncorrelated latent variables, which are used to make the predictions (Dunn, 2019). It is for these reasons the PLS method is commonly used in the field of chemometrics (Wold et al., 2001), and in recent years has been successfully used to train models to monitor *in-situ* DOC concentration from CDOM absorbance (Avagyan et al., 2014; Codden et al., 2020; Etheridge et al., 2014; Langergraber et al., 2003; Vaughan et al., 2017; Zhu et al., 2020).

A further benefit of the PLS method, which is often overlooked in the environmental-sciences community, are two independent consistency checks built into the model, Hotelling's T^2 and square prediction errors (SPE; also known as DModX or the

Q-statistic; Dunn, 2019). These consistency-checking statistics are widely used in the realm of process engineering and manufacturing plants to monitor production, maintain set efficiency standards, detect faults in a process and diagnose them (Chen et al., 2004; Dunn, 2019; Kaistha & Upadhyaya, 2001; Mujica et al., 2011; Villegas et al., 2010). The training dataset defines "normal" and "not-normal" (i.e. faulty) conditions. The statistics identify whether a process is under normal operation (and therefore producing "acceptable" product or observations) or if a process is out of control (and therefore producing a faulty product or observations). An observation found to be faulty can then be diagnosed to identify what is causing the fault, leading to a decision being made whether the model must be updated or a correction (e.g. bias) must be made. When used to make predictions, these consistency checks help us determine whether an observation can be suitably used in the model to make a prediction (Dunn, 2019). The SPE measures how well each sample fits into the PLS model (Mujica et al., 2011) (i.e. how compatible a sample is in a model) and so indicates whether the sample *can* be used to make a prediction. Hotelling's T^2 measures the sample's systematic variations (Villegas et al., 2010) and identifies high leverage outlier observations where a prediction would be extrapolated rather than interpolated. In applying to CDOM spectral data to predict DOC concentration, these consistency checks can be used to evaluate *in-situ* data outside of grab-sampling campaigns to assess the validity of a model's application. This could be helpful in understanding if site-, season- or event-specific models are required to accurately model stream DOC content from *in-situ* absorbance data. Furthermore,

periods with consistently outlying observations can be diagnosed to identify what subset of the observation's measurements are causing the problem and further identify periods where the model fails to work or what new training data would have to be prioritized for collection to update the model.

Because PLS models can use the entire CDOM spectra, they can consider the varying relationship between DOC concentration and CDOM absorbance and predict DOC more accurately than MLR models that use discrete wavelengths. However, training and validation datasets of representative samples that *fully* capture this variability are challenging to obtain from remote sites, if not impossible, especially during inaccessible seasons (e.g. winter and spring freshet in boreal and arctic regions). A training dataset that *mostly* captures the full spectrum of DOC-CDOM absorbance relationships observed in dynamic streams can still perform well, but not without considering gaps in the data where or when the model may not perform adequately. However, these may not be identified and quantified in many instances (i.e. remote locations or inaccessible times of the year) due to a lack of grab-samples from which to check.

Alternatively, T^2 and SPE can be used to potentially identify performance gaps in the model, identified as periods of faulty observations from a continuous dataset to which the model was applied. These can be used and applied to the continuous *in-situ* observations from the stream to identify different kinds of outliers (T^2 or SPE or both) and determine whether the model can suitably make a prediction using such observations or

if the observations are incompatible with the model and should therefore not be used to make predictions.

3.2.3 Using factors that interfere with CDOM absorbance for better predictions

The relationship between DOC concentration and CDOM absorbance can also be impacted by turbidity from particles found naturally in streams and responsive to hydrologic changes (Jollymore et al., 2012; Langergraber et al., 2003; Vaughan et al., 2017, 2019). Turbidity, typically estimated from light scattered at different angles, also registers as an additional absorbance source during *in-situ* spectrophotometric measurements. The high correlation between turbidity and mean absorbance from 710-730 nm from a continuous *in-situ* dataset suggested that the mean absorbance at those higher wavelengths could be used to correct for turbidity (Grayson & Holden, 2016). That is a similar approach typically used in laboratory settings to offset light-scattering effects from particles small enough to pass through filters (Green & Blough, 1994; Helms et al., 2009). Therefore, it can effectively correct particles' light-scattering effect from *in-situ* absorbance and reduce the variability in the relationship between DOC and CDOM absorbance caused by these (e.g. due to turbidity Jeong et al., 2012 over-predicted DOC concentration). This correction, however, assumes the effect to be constant across all

wavelengths. The particles contributing to turbidity at a given stream site may render scattering and impacts on absorbance differently across the spectrum of wavelengths used to detect DOC (Jeong et al., 2012; Langergraber et al., 2003; Vaughan et al., 2017, 2019). Such an influence on absorbance would need to be assessed with *in-situ* turbidity to determine if it impacts the relationship between DOC concentration and CDOM absorbance used to predict DOC.

3.2.4 Chapter Objectives

Additional variables that highly correlate to streamDOC content while also relating to seasonality and shifting sources during storm events, such as conductivity (Inserillo et al., 2017; Kobayashi, 1986) and water level (Jollymore et al., 2012; Vaughan et al., 2017) are good candidates to be considered as predictors for DOC concentration. Also, optical properties related to DOC composition, such as spectral slopes and their ratio (Helms et al., 2009), relate to hydrologic events (Ågren et al., 2010; Vaughan et al., 2017, 2019) and are therefore good candidates to consider as predictors of *in-situ* DOC concentration. Though derived from spectral slopes and therefore multicollinear among them, spectral slope ratio relates to different compositional information of DOC than the spectral slopes from which it is calculated (Helms et al., 2009) and from which the PLS algorithm can still extract information. Including these physical and optical water quality parameters as

predictors in a PLS model means that any flagged outliers (i.e. those not fully represented in the training dataset) are likely to highlight periods related to hydrologic events or seasons not represented by the training dataset. This suggests these faulty periods should not be used to make predictions, or any predictions should be made with caution as they are being made by data outside the model's scope or an extrapolation.

To continuously monitor DOC concentration from *in-situ* data, this chapter aimed to train PLS models using laboratory-measured DOC coupled with *in-situ* CDOM absorbance, specific conductance, water level, spectral slopes and slope ratio collected from two boreal forest headwater stream sites. To understand the representativeness of the training dataset composed of seasonal grab samples and targeted storm samples from the two different sites within the same stream, a series of test-set switches were performed to identify potential redundancies in the training dataset that negatively influence the final model. Furthermore, the test-set switches can help identify the optimal number of PLS-components to be used in a model trained using the combined subsets of data. Considering CDOM absorbance as the primary set of predictors of DOC concentration, the importance of extra predictors (water level, conductivity, spectral slopes, and slope ratio) was measured using the variable importance in the projection (VIP) statistic. Simultaneously, predictions of DOC concentration from continuous *in-situ* data from different model iterations that use different combinations of the extra variables were compared to understand how each variable affected predictions of DOC concentration. Consistency checks built into PLS models (Hotelling's T^2 and SPE) were also

explored and used to assess model performance. When applied here to the continuous *in-situ* dataset, these consistency checks identify periods where a model may fail to perform and diagnose why a model may be failing to perform at a specific site or in a certain period. This can then be related to the training dataset's representativeness, further informing what samples should be prioritized for collection when deciding to re-train the model. Finally, an absorbance-based correction for turbidity was developed and tested via application to a validation dataset, and its suitability was further assessed when applied to the high-resolution dataset of *in-situ* measurements.

3.3 Methods

3.3.1 Sample Collection and Analysis

Study Area:

Water samples were collected from two instrumented sites along the Horseshoe Brook, a boreal headwater stream located within the Pynn's Brook Experimental Watershed Area (PBEWA; Figure 3.1). The Horseshoe Brook catchment is divided into two sub-catchments: a low relief upstream catchment (Upper) of 3.8 km² dominated by ponds and wetlands, and a downstream catchment (Lower) 7.5 km² dominated by forested hillslopes, as defined by the respective instrumented sites (UHS and LHS; Figure 3.1). For details regarding the PBEWA climate normals and catchment land cover composition, please refer to Chapter 2 (Table 2.1).

Sample and In-situ Data Collection:

Discrete stream samples and continuous *in-situ* data were collected throughout February-November 2019. All discrete samples were collected or stored in acid-washed, amber, high-density polyethylene bottles to avoid photochemical degradation and stored in coolers for transportation back to the laboratory for processing. Grab-samples (1-3 per day from each site) were taken directly from the stream during periodic field campaigns

(every 1-2 months) from February 2019 through October 2019. For high temporal resolution event sampling (every 1-2 hours), a peristaltic autosampler (Hach AS950) with the intake tube anchored to deep-driven iron rebar beside *in-situ* data loggers was used to collect up to 24 samples throughout the event, with the collection and transfer to amber bottles occurring within 24 hours of the last sample collection, with processing within 48 hours of the final collection.

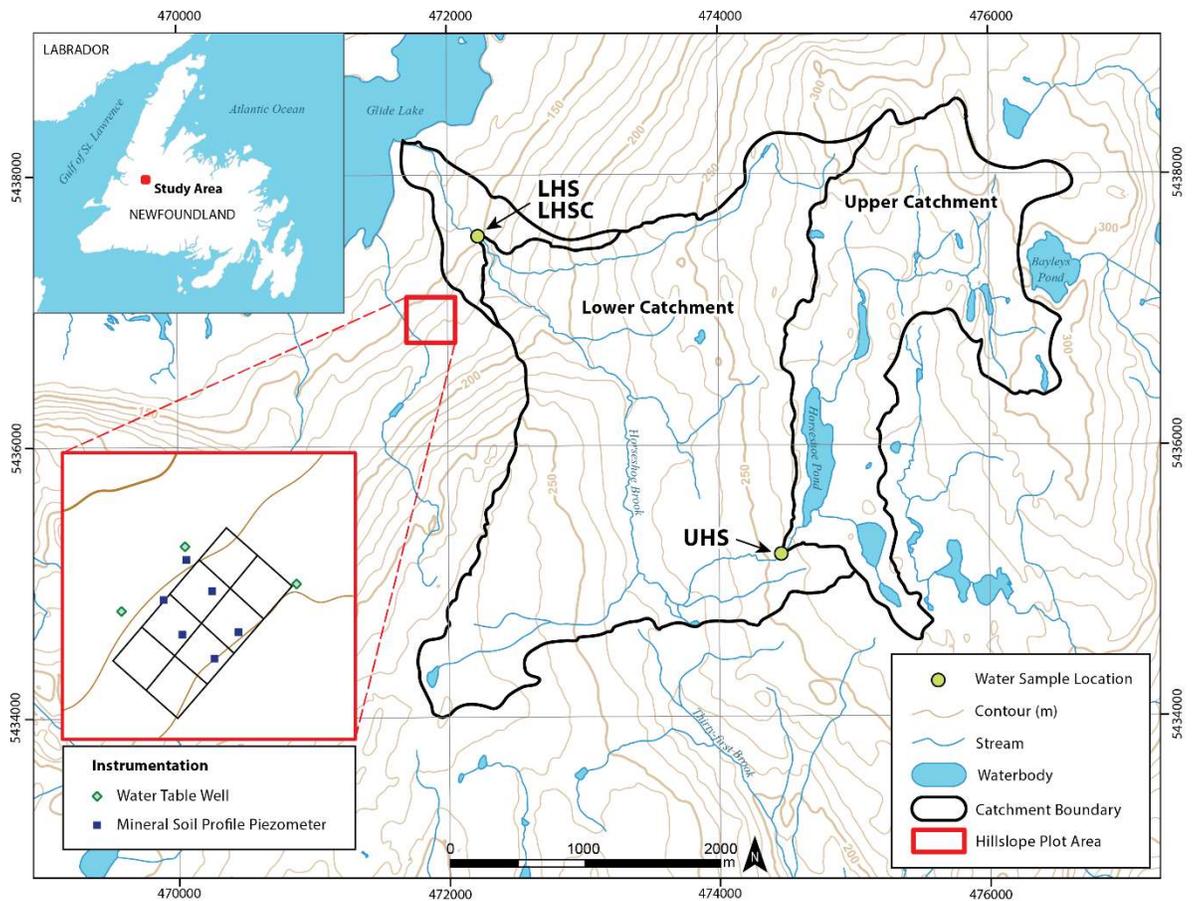


Figure 3.1: Map of PBEWA catchment

Horseshoe Stream catchment in Newfoundland and Labrador, Canada. The catchment is divided into a low relief pond and wetland dominated Upper catchment and a forested hillslope Lower catchment, defined by

the instrumented sampling sites (UHS and LHS). The LHSC site is located 100 m downstream from the main LHS site and is instrumented during the snow and melt periods (November-May)

In-situ data loggers used to monitor absorbance (S::can Spectro::lyser, Vienna), pH and redox potential (ORP) (Seametrics TempHion), conductivity (Onset HOBO Logger U24-001) and water level via a pressure transducer (Onset HOBO Logger U20L-01) were installed at each site (Figure 3.2). Probes were housed in PVC piping and anchored to deep-driven iron rebar. An additional pressure transducer (Onset HOBO Logger U20L-01) was housed on a streamside tree at each site to obtain atmospheric pressure for the water level calculations. Water level, conductivity, and pH/ORP probes were set to log at 30-minute intervals throughout Spectro::lyser deployments. The Spectro::lyser at LHS was set to measure every 30 minutes from February-July 2019 and every 60 minutes since August 2019. The Spectro::lyser at UHS was installed in August 2019 and measured in 60-minute intervals. From June to October, the probes at LHS were housed similar to Figure 3.2. However, during the winter months (November to May), when snow and ice cover are significant, the probes were moved to a more accessible and protective installation on the downstream side of a culvert, ~100 m downstream the original LHS site (LHSC, Figure 3.1).

Sample Preparation and Analysis:

Before filtration, event (autosampler) samples were measured for turbidity using a turbidimeter (Hach 2100P) and pH and conductivity using a benchtop meter (Thermo

Scientific, Orion Versa Star Pro). The precision of the turbidity measures was <10% coefficient of variation (RSD) among analytical replicates (n = 3).

All samples were filtered and prepared for DOC analysis, as outlined in Chapter 2. Mercuric chloride was no longer added to sub-samples meant for UV-Vis absorbance measurements as it interfered with CDOM absorbance in the 200-275 nm range and instead was stored at 4°C until analysis within seven days. Peacock et al. (2014) reported absorbance of filtered pore-water samples after cold and dark storage over 12 weeks had no significant difference than those measured on the first day. We had similar results with filtered stream water tested over two weeks.



Figure 3.2: *In-situ* instrumentation

In-situ instrument housing within the UHS site. Seen in their respective housings are the spectrophotometer (Spectro::lyser), pH (TempHion), conductivity (HOBO), and water level pressure transducer (HOBO) probes during the low flow late summer period in the smaller Upper Horseshoe Brook study site.

As outlined in Chapter 2, DOC concentration was determined via combustion catalytic oxidation method (Shimadzu TOC-V with ASI-V) with a limit of quantification of 1.72 mg C L⁻¹ (limit of detection of 0.52 mg C L⁻¹) and precision of <3% coefficient of variation (RSD) for concentrations ranging 0.5 to 50 mg C L⁻¹.

Absorbance spectra (A_{λ}) of filtered samples were measured using a dual-beam UV-Vis benchtop spectrophotometer (Perkin Elmer Lambda 365) from 200-800 nm at 1 nm intervals. The precision of these absorbance measures using a 1 cm path length was <5% RSD among analytical replicates (n = 3) for A_{300} in the range of 10-60 m⁻¹ for samples with a limit of detection of 0.33 m⁻¹.

Following the absorbance measurement of each sample, sample pH and conductivity (Thermo Scientific, Orion Versa Star Pro, City) and ORP (Oakton ORPTestr, WD-35650-10, Vernon Hills, Illinois) were measured again to provide a direct match with absorbance data collected.

3.3.2 Data Processing

In-situ absorbance was corrected for particle light-scattering effects by subtracting the average absorbance measured across 700-730 nm from the rest of the spectra (Grayson & Holden, 2016).

Integrative optical metrics related to CDOM and DOC composition (specific UV absorbance at 255 nm (S_{255}), spectral slopes from 275-295 nm ($S_{275-295}$) and 350-400 nm ($S_{350-400}$), and slope ratio (S_R); Helms et al., 2009) were calculated as outlined in Chapter 2 from laboratory-measured absorbance and *in-situ* absorbance which was scatter-corrected beforehand. These were also calculated for *in-situ* absorbance spectra corrected for turbidity using the correction method outlined in section 3.3.3.

Absorbance due to turbidity (turbidity-induced absorbance) was measured as the difference between *in-situ* (raw) absorbance and the laboratory-measured absorbance of the corresponding filtered sample in the dataset. To account for differences between instruments, before calculating turbidity-induced absorbance, laboratory-measured absorbance spectra were converted to *in-situ* absorbance using a linear model that relates the two instrument measurements to one another.

Continuous *in-situ* ORP was corrected for time-dependent drift while conductivity data were converted to specific conductance (SPC) using water temperature monitored simultaneously by the sensor to eliminate temperature-related effects from conductivity. Water level (WL), pH/ORP and SPC data, measured every 30 minutes, were linearly

interpolated for instances where measurement times did not match with *in-situ* absorbance measurements.

3.3.3 Model Training

A series of partial least squares (PLS) regression models were trained using laboratory-measured DOC concentration as the response variable and the corresponding *in-situ* data as the predictor variables. Please refer to Chapter 2 for more details regarding the PLS regression model. Specifically, the scatter-corrected *in-situ* absorbance from wavelengths 250-450 nm ($A_{250-450}$) was always included as predictor variables. Other predictor variables used in concert with $A_{250-450}$ to train the PLS models were spectral slopes from the 275- to 295 nm and 350 to 450 nm wavelength ranges, the ratio of these two slopes, specific conductance and water level ($S_{275-295}$, $S_{350-400}$, S_R , SPC, and WL, respectively).

Lastly, a correction for turbidity-induced absorbance was developed and used to train a new PLS model from this alternative turbidity-corrected absorbance data. The correction and PLS model were then applied to a validation dataset and the continuous *in-situ* dataset to test performance and compare to prior iterations where this alternative correction for turbidity was not applied.

Component Selection

These new PLS models underwent 2000 iterations of randomized k-fold cross-validation for up to the maximum number, or no more than 20, of PLS-components during the test-set switch and variable testing steps. The models were developed using the `pls()` function and cross-validation performed using the `crossval()` function from the R 'pls' package (Mevik, B.H., Wehrens, R., Hovde, K., Hiemstra, P., 2019; V. 2.7-2) within R-Studio (V. 1.2.5042). Root mean square error (RMSE) was calculated as outlined in the previous chapter without adjusting for degrees of freedom (Equation 2.9-2.10). The predicted residual error sum of squares (PRESS) calculated from the prediction residuals of each component from each model iteration's cross-validation was recorded and converted to an RMSE from cross-validation (RMSE-CV) by:

$$\text{RMSE-CV} = \sqrt{\frac{\text{PRESS}}{n}} \quad (3.1)$$

where n is the size of the training dataset. The k-fold cross-validation was randomized 2000 times, and using the resulting 2000 PRESS's, an average and standard deviation of RMSE-CV of the 2000 iterations for each PLS-component for each model iteration were calculated and compared. This procedure is similarly outlined in Cruciani et al. (1992), where the standard deviation of 2000 RMSE-CV would be equivalent to the standard deviation of error of calculations (SDEC) from Cruciani et al. (1992). The PLS-components that had among the lowest average RMS-CV and SDEC of the 2000 randomized cross-

validation iterations were considered to generate each resulting model (Table 3.1). For cases where a validation dataset was available, the validation dataset RMSE, MAE and d_r were computed for each component, and components that exhibited minimum RMSE, MAE or highest d_r , or a combination, were considered potential candidates to be used.

The selected component candidates from either method were then compared. Similar outcomes from both tests suggest that the multiple iterations of randomized k-fold cross-validation to compute an average RMSE are a robust way to identify the optimum number of components for cases when there is no validation dataset available.

Model consistency checks, conducted using Hotelling's T^2 and square prediction errors (SPE), from different model iterations were compared to each other to assess the performance of models further and help identify the most robust model to use with continuous data and predict DOC in the future. These metrics measure unusual variation, though, are independent of each other. T^2 looks at error within the model hyperplane, while SPE looks at error outside the model (Dunn, 2019; Mujica et al., 2011). For example, these metrics are typically used in processing plants to identify systematic or random problems in a process, diagnose them and identify when a model needs to be adjusted or updated (Dunn, 2019). Therefore, when used for prediction, these checks can elucidate whether the model is appropriate for a dataset, for example, when predicting DOC concentration in a new stream site or season not represented in the training dataset used to develop the model.

Table 3.1: Description of 3 sets of model iterations

The model naming scheme for each series of model tests. All test-set switch models used the same predictor variables. These were the scatter corrected *in-situ* absorbance from 250-450 nm ($A_{250-450}$), spectral slopes for A in the 275-295 and 350-400 nm range ($S_{275-295}$ and $S_{350-400}$), the ratio of $S_{275-295}$ to $S_{350-400}$ (S_R), specific conductance (SPC) and water level (WL). Three sets of tests were performed: a test set switch (1), variable test (2) and a turbidity correction test (3). Each test has a set of steps indicated by decimals (0.0-0.8) specific to the test such that: for the test set switch, they indicate the subset of data used to train a model validated by the leftover subset and with the entire dataset being used to train Model 1.8, for the variable tests they indicate an additional variable ($S_{275-295}$, $S_{350-400}$, S_R , WL or SPC) being used in concert with $A_{250-450}$ keeping the variable from the prior decimal ending with Model 2.5 being equal to model 1.8, and the turbidity test having only one model (3.0) trained from alternative turbidity corrected subset of data comparable to model 1.5.

Step		Model ID								
		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
Test-Set Switch (Training dataset used)	1	N/A	Baseflow	Upper Event 2	Lower Event 1	Lower Event 2	All Event	All Upper	All Lower	Full Set
Variable Testing (Variables used, Full Set)	2	$A_{250-450}$	2.0 + $S_{275-295}$	2.1 + SPC	2.2 + WL	2.3 + S_R	2.4 + $S_{350-400}$ (Equivalent to 1.8)	N/A		
Turbidity Correction	3	Remake model 1.5 using turbidity corrected Abs.	N/A							

Hotelling's T^2 is a positive scalar number that summarizes observations score values from A -number of components and measures the distance from the center of the model hyperplane to the observation's projection onto the hyperplane (Dunn, 2019). Therefore, an observation that projects to the model's center ($T^2 = 0$) would have every input value ($x^i_{i,k}$) at the mean. T^2 can also be interpreted as measuring the systematic variations of a process (Villegas et al., 2010). The T^2 of i^{th} observation was calculated as:

$$T^2 = \sum_{a=1}^A \left(\frac{t_{i,a}}{s_a} \right)^2 \quad (3.2)$$

where $t_{i,a}$ is the observation's score for component a , s_a^2 is the variance of component a from training dataset observations, and A is the number of components used in the model.

SPE measures the distance of an observation from the model hyperplane (the length of the projection) and summarizes the prediction error of an observation from the E matrix ($n \times K$) when using A -components. An observation with low SPE means that the correlation structure between measurements (K -variables) is observed in the training dataset. An observation with $SPE = 0$ indicates it lies precisely on the model hyperplane and follows the model structure exactly.

$$SPE_i = \sqrt{e'_{i,A} e_{i,A}} \quad (3.3)$$

where $e'_{i,A}$ is the residual row ($1 \times K$ vector) from the i^{th} observation using A components, obtained from:

$$e'_{i,A} = x'_i - \hat{x}'_{i,A} \quad (3.4)$$

and;

$$\hat{x}'_{i,A} = x'_i - t'_i P' \quad (3.5)$$

where P is the PLS model's loadings matrix with dimensions $K \times A$, t'_i are the scores of the i^{th} observation from A -components ($1 \times A$) and x'_i is the i^{th} observation (row) from matrix X . Matrix X is the original data input into the PLS model, with $n \times K$ dimensions, with each column having been mean-centred and variance scaled in the case of training datasets. Validation datasets and the continuous *in-situ* dataset had their respective X matrices centred and scaled using the corresponding models training dataset's means and variances of each variable. The value $\hat{x}'_{i,A}$ are the model predicted (mean-centred and scaled) x'_i for the i^{th} observation using A components.

For validation datasets and the continuous dataset, scores (T) were calculated from:

$$T = X R \quad (3.6)$$

where R is the model's projection matrix:

$$R = W(P'W)^{-1} \quad (3.7)$$

and where W is the PLS model's weights matrix when using A -components.

To help identify outlier observations, 95% and 99% upper threshold limits were calculated for T^2 and SPE from the model's corresponding training dataset. T^2 for future observations (validation datasets and continuous dataset) were calculated using Equations 2.2 and 2.6, with T^2 being distributed closely to the F-distribution. The upper threshold limits for T^2 ($T_{100(1-\alpha)\%}^2$) were calculated from:

$$T_{100(1-\alpha)\%}^2 = \frac{A(n^2-1)}{n(n-A)} \times F_{\alpha,A,n-A} \quad (3.8)$$

where n is the size of the training dataset, A is the number of PLS components, and F is the critical F-value for a specified α level with A and $n - A$ degrees of freedom (Villegas et al., 2010). The threshold value can be interpreted as measuring the systematic or normal variation of a process. An observation found in violation of the threshold would indicate that the systemic variations of that observation are out of control (Villegas et al., 2010), not being observed or captured by the training dataset.

Similarly, SPE for future observations (validation datasets and continuous dataset) were calculated using Equations 2.3-2.6 and is approximately normally distributed

(Jackson, 1991; Jackson & Mudholkar, 1979). The upper threshold limits for SPE ($SPE_{100(1-\alpha)\%}$) were calculated as:

$$SPE_{100(1-\alpha)\%} = \Theta_1 \left[\frac{Z_\alpha \sqrt{2\Theta_2 h_0^2}}{\Theta_1} + 1 + \frac{\Theta_2 h_0 (h_0 - 1)}{\Theta_1^2} \right]^{1/h_0} \quad (3.9)$$

where Θ_i is the sum of the i^{th} power ($i = 1, 2, 3$) of the singular values of $\frac{E}{\sqrt{n-1}}$ from the training dataset (Jackson, 1991; Kaistha & Upadhyaya, 2001; Villegas et al., 2010), Z_α is the critical z-value for a specified α -level, positive when $h_0 > 0$, or negative $h_0 < 0$, and where:

$$h_0 = 1 - \frac{2\Theta_1\Theta_3}{3\Theta_2^2} \quad (3.10)$$

An observation found in violation of this threshold would indicate random noise has significantly changed or an unusual event has occurred and produced a change in the covariance structure of the observation's input vector (Jackson, 1991; Villegas et al., 2010). The violation would signify that the observation cannot be characterized by the model's components (or a subset of components). This could be either due to too few components being used to produce a good model or the observation truly being an outlier for the model (Jackson, 1991) and therefore not compatible with the model. Observations found to violate this threshold are considered inconsistent with the model and are not suitable for making predictions (Dunn, 2019).

Observations from the continuous *in-situ* dataset that exceeded T^2 or SPE thresholds were further used to interpret and diagnose which variables and components contributed to the elevated T^2 and SPE. This was used for identifying potential gaps in the training dataset where representative samples may be lacking.

Test-set Switch Procedure

Samples were grouped based on site (UHS or LHS), flow condition (baseflow or storm event), and storm event date to form different training datasets (Table 3.2, Table 3.3), each being validated by the leftover ($98 - n$) samples. All models from this procedure use scatter-corrected *in-situ* $A_{250-450}$, $S_{275-295}$, $S_{350-400}$, S_R , WL and SPC to predict DOC concentration. The optimal number of PLS-components for Models 1.1-1.7 were identified as previously outlined using their corresponding validation datasets. The number of PLS-components for Model 1.8 was selected based on the best performing component when looking at the averaged RMSE resulting from the 2000 randomized k-fold cross-validations.

Testing additional in-situ variables

By using the last model from the previous section (Model 1.8; Table 3.1) as the final step in a new series of model iterations, five new iterations were performed, each

including one additional variable (SPC, WL, spectral slopes, slope ratio) with each iteration, starting with none (using only $A_{250-450}$ as the predictors, Model 2.0; Table 3.4). The order in which each variable was added to the regression was determined by their importance and overall contribution to SPE.

Table 3.2: Sample counts and test-set switch datasets

Sample count breakdown. Baseflow samples included samples collected throughout the early Fall period (ET Shutdown) and late Fall periods using an autosampler with periodic grab-samples throughout the year. Event 1 corresponds to a storm event in late September (early Fall) and Event 2 corresponds to a storm event at the end of October-start of November (late Fall). Shaded boxes indicate samples used as training data for test-set switches using the leftover ($98 - N$) samples as a validation dataset.

Site	Sample N				
	Baseflow	Event 1	Event 2	Event Total	Total
Upper	7	0	16	16	23
Lower	42	16	17	33	75
Total	49	16	33	49	98*

*Asterisk – models developed after the test-set switch used all available samples to be trained.

Table 3.3: Test-set switch model names and training datasets

Model identifiers for test-set switch training datasets following the same scheme as Table 3.2.

Site	Model ID				
	Baseflow	Event 1	Event 2	Using All Event	Using All
Upper	N/A	N/A	1.2	N/A	1.6
Lower	N/A	1.3	1.4	N/A	1.7
Using All	1.1	N/A	N/A	1.5	1.8

Model 1.8 was used to determine the variable importance in the projection (VIP), a variable selection method (Mehmood et al., 2012) that suggests potential candidate variables that could be omitted or retained when re-training a model. The VIP for each variable was calculated by:

$$VIP_k = \sqrt{\frac{K \sum_{a=1}^A \left[SS_a \left(\frac{w_{a,k}}{\|w_a\|^2} \right) \right]}{\sum_{a=1}^A (SS_a)}} \quad (3.11)$$

where K is the number of variables used as predictors, A is the number of components, SS_a is the sum of squares explained by the a^{th} component, $w_{a,k}$ is the weight of the k^{th} variable in the a^{th} component (Mehmood et al., 2012). The VIP value for each variable (VIP_k) is bounded at 0, and $\sum VIP_k = K$. A variable with $VIP_k < 1$ contributes less to the model projection than a variable with $VIP_k > 1$.

The total error contribution for each variable was determined from each column of the squared E matrix of the training dataset and reported as a % of the total error. Variables with higher VIP and contributing to lower SPE were considered to be added to the model before variables with lower VIP and higher contribution to SPE, resulting in the order outlined in Table 3.4.

Table 3.4: Naming order of hierarchical models

Model identifiers for models used in exploring the importance and influence of each variable on predicting dissolved organic carbon concentration ([DOC]).

Variables Used in PLS X block (to predict [DOC])	Model ID
A ₂₅₀₋₄₅₀	2.0
A ₂₅₀₋₄₅₀ and S ₂₇₅₋₂₉₅	2.1
A ₂₅₀₋₄₅₀ , S ₂₇₅₋₂₉₅ , and SPC	2.2
A ₂₅₀₋₄₅₀ , S ₂₇₅₋₂₉₅ , SPC, and WL	2.3
A ₂₅₀₋₄₅₀ , S ₂₇₅₋₂₉₅ , SPC, WL, and S _R	2.4
A ₂₅₀₋₄₅₀ , S ₂₇₅₋₂₉₅ , SPC, WL and S _R , and S ₃₅₀₋₄₀₀	2.5

Developing and testing a correction for turbidity-induced absorbance

As previously stated, the average absorbance from wavelengths of the 700-730 nm range from *in-situ* spectra was used to capture the light-scattering effect (average A₇₀₀₋₇₃₀) and subtracted from A₂₅₀₋₄₅₀ to remove light-scattering effects assumed to be constant across the absorbance spectra. An additional correction was developed to be applied to *in-situ* absorbance and reduce interference from the light-absorbing effect of particles encountered during *in-situ* monitoring assumed to increase with decreasing wavelengths. The absorbance effect from particles (turbidity-induced absorbance) for each wavelength was determined by, first, subtracting the light-light-scattering effect from the raw (unfiltered) *in-situ* absorbance spectra, and then subtracting the absorbance spectra of the laboratory-filtered grab-sample counterpart from the resulting spectra to obtain the turbidity-induced absorbance of the sample. The resulting turbidity-induced spectra had a similar shape to CDOM absorbance spectra, where absorbance increased as wavelength

decreased. Then, by linearly regressing turbidity-induced absorbance from a specific wavelength onto average $A_{700-730}$, the resulting set of coefficients relate the light-scattering effect to turbidity-induced absorbance for each specific wavelength generating a set of coefficients unique to each wavelength. The whole procedure was done using the same training dataset used for Model 1.5 (Table 3.1). The coefficients were then used to calculate the predicted turbidity-induced absorbance for each wavelength based upon the scatter values from the average $A_{700-730}$. This predicted turbidity-induced absorbance was then subtracted from scatter-corrected *in-situ* absorbance spectra from the training and validation datasets from Model 1.5 and the continuous *in-situ* dataset resulting in fully corrected *in-situ* absorbance spectra. Spectral slopes and slope ratio were then calculated from the fully corrected absorbance spectra.

Using the fully corrected *in-situ* absorbance spectra from the training dataset, a new PLS model was trained (Model 3.0) in the same manner as Model 1.5. However, using the fully corrected absorbance spectra, spectral slopes, and slope ratio. Model 3.0 was then applied to the validation and continuous datasets, with the resulting metrics and predictions being directly compared to those that resulted from Model 1.5.

3.4 Results

To best predict DOC concentration from continuous *in-situ* absorbance data, different model iterations were tested. First, the number of PLS-components used to predict DOC best were determined and held constant through further steps. Then, to identify variables that could be interfering with model predictions, all variables besides $A_{250-450}$ were removed from the model and then re-introduced hierarchically in order of importance, based on VIP and SPE contribution. This enabled the identification of variables that could be redundant or may interfere with the model's predictive ability. Finally, the application of a turbidity correction that subtracts absorbance caused by turbidity across the entire spectra was tested and compared to the alternatively used scatter-correction to identify if the former yielded improved model performance. Following these steps, the best performing model, Model 2.4, had a training dataset RMSE of 0.32 mg C L^{-1} and used 8 PLS-components to characterize scatter-corrected $A_{250-450}$, $S_{275-450}$, S_R , water level and specific conductance using only the scatter correction for absorbance spectra.

3.4.1 Different water characteristics between stream sites and storm and baseflow periods

Overall, DOC concentration within the UHS was found to be much less variable than in the LHS during either base or storm flow periods (Figure 3.3.A). The LHS site was found to have a much more extensive range in DOC concentration across storm events

than those observed from the UHS site (Figure 3.3.A). As expected, *in-situ* SPC from the training dataset during storm events was lower than baseflow, while the greatest baseflow variability in SPC was observed at LHS ranging from 80 to 100 $\mu\text{S cm}^{-1}$ during the baseflow early autumn period and 60 to 100 $\mu\text{S cm}^{-1}$ during the baseflow summer period. In both sites, SUVA_{255} was greatest during storm events and lower in baseflow periods. However, samples collected in baseflow conditions during the LHS winter period contributed to the large range observed in baseflow LHS SUVA_{255} (1.5 to 4.9 $\text{L mg}^{-1} \text{C m}^{-1}$; Figure 3.3.C). The event samples exhibited lower S_R values than baseflow samples, and S_R of UHS samples was always greater than those from LHS. The *in-situ* pH from training dataset samples varied greatly during storms in UHS (5.5 to 7), while pH from LHS samples ranged from 7 to 7.6 (Figure 3.3.E).

3.4.2 Model component selection indicates 8 model components is most robust

Model 1.5 exhibited the lowest validation dataset RMSE and MAE (0.565 and 0.433 mg C L^{-1} , respectively) using 11-components with similar results when using 8-components (0.582 and 0.444 mg C L^{-1} , respectively; Figure 3.4, Table 3.5). However, the lowest training dataset mean RMSE from cross-validation (RMSE-CV) for Model 1.5 was 0.27 mg C L^{-1} observed at 5-components (Figure 3.4, Table 3.5). When using the whole dataset to train Model 1.8 (Table 3.1) there was not much difference in mean RMSE-CV

when using 3- or 6-fold cross-validation. Model 1.8 was observed to have consistently lower mean RMSE-CV and SDEC from 2000 randomizations when using 8-components.

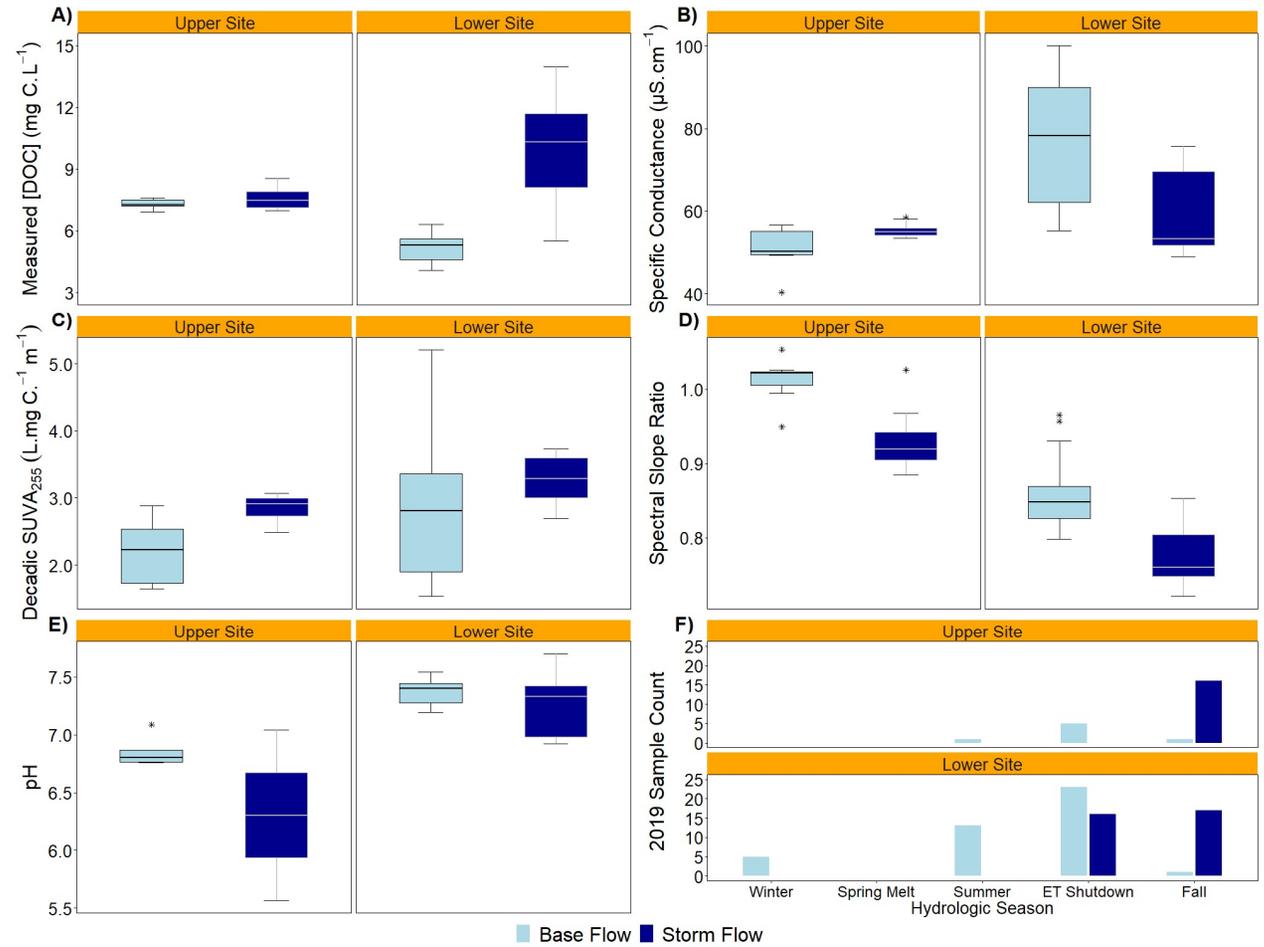


Figure 3.3: Distribution of water quality and compositional parameters

Distribution of A) Dissolved organic carbon concentration ([DOC]), C) specific UV absorbance at 255 nm (SUVA_{255}) and other measured parameters from samples collected in 2019 from Horseshoe stream sites located within the Pynn's Brook Experimental Watershed Area. Box and whiskers mark interquartile range (IQR) and range (excluding outliers), respectively, asterisks mark outlier samples outside 1.5IQR.

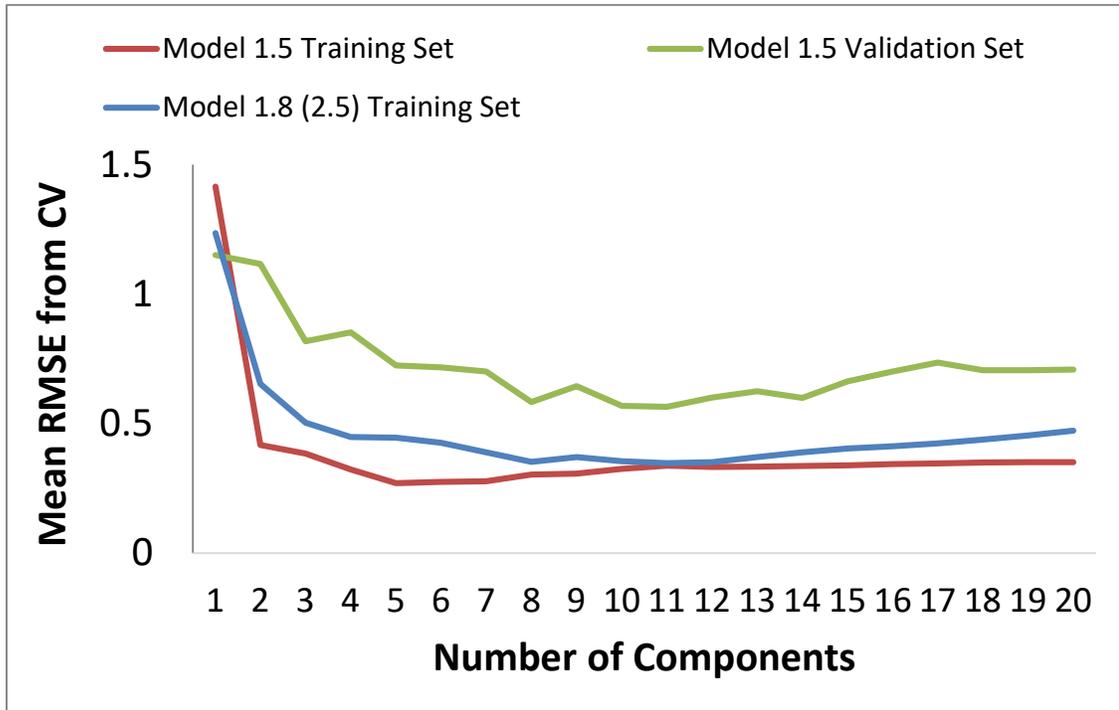


Figure 3.4: Result of 2000 randomized cross-validations

Average root mean squared error (RMSE) from 2000 randomized 3-fold cross-validations for each component from Model 1.5 training and validation datasets and Model 1.8 training dataset, detailed in Table 3.5.

Table 3.5: Results from 2000 randomized cross-validations

Results for component selection – 2000 randomized K-fold cross-validations were performed and root mean square error from cross-validation (RMSE-CV) were calculated from each iteration from the predicted residual error sum of squares for each component up to 20 components. The mean and standard deviation (SDEC) for the 2000 iterations are reported for training datasets. For Model 1.5 where a validation dataset was available, the RMSE and mean absolute error (MAE) from predictions are reported. Three segments were used for CV in Model 1 as there were three distinct sampling periods within the training dataset. With a doubling in sample size for Model 2.5, six segments were used in addition to 3 segments. Before this study, 6-components were used for predicting dissolved organic carbon using absorbance. However, it was determined that 8-components were more appropriate. Models 1.5 and 2.5 were developed using scatter corrected absorbance spectra from 250-450 nm in 2.5 nm intervals, spectral slope from the 275-295 nm range ($S_{275-295}$), spectral slope from the 350-400 nm range ($S_{350-400}$), spectral slope ratio (S_R), specific conductance (SPC) and water level (WL).

N comp.	Split Dataset – Model 1.5				Full Dataset – Model 1.8 (and 2.5)			
	Storm Flow Training dataset n = 49		Baseflow Validation dataset n = 54		Training dataset n = 98			
	RMSE-CV		RMSE	MAE	RMSE-CV			
	3 CV Segments				3 CV Segments		6 CV Segments	
	Variables used for models: Scatter-Corrected Abs. 250-450 nm, $S_{275-295}$, $S_{350-400}$, S_R , SPC, WL							
	Mean	SDEC	RMSE	MAE	Mean	SDEC	Mean	SDEC
1	1.415	0.052	1.151	0.913	1.236	0.034	1.231	0.017
2	0.417	0.019	1.116	0.959	0.653	0.016	0.651	0.009
3	0.384	0.033	0.817	0.64	0.503	0.015	0.5	0.008
4	0.322	0.027	0.852	0.719	0.447	0.017	0.443	0.01
5	0.27	0.023	0.724	0.537	0.445	0.025	0.439	0.015
6*	0.274	0.026	0.716	0.499	0.425	0.029	0.421	0.016
7	0.276	0.034	0.7	0.506	0.389	0.031	0.38	0.018
8*	0.302	0.044	0.582	0.444	0.352	0.02	0.345	0.011
9	0.306	0.05	0.644	0.506	0.37	0.028	0.359	0.018
10	0.325	0.059	0.568	0.433	0.354	0.028	0.336	0.017
11	0.337	0.063	0.565	0.433	0.347	0.026	0.332	0.015
12	0.332	0.052	0.6	0.461	0.35	0.026	0.331	0.015
13	0.333	0.05	0.625	0.474	0.37	0.028	0.353	0.017
14	0.335	0.049	0.599	0.46	0.388	0.029	0.368	0.019
15	0.338	0.049	0.663	0.506	0.403	0.031	0.383	0.021
16	0.343	0.049	0.701	0.537	0.412	0.031	0.388	0.022
17	0.346	0.049	0.735	0.578	0.423	0.032	0.396	0.023
18	0.349	0.049	0.705	0.561	0.437	0.032	0.409	0.024
19	0.35	0.049	0.705	0.562	0.454	0.033	0.424	0.025
20	0.35	0.048	0.708	0.568	0.472	0.034	0.443	0.027

*Asterisks represent potential number of components to use for the model.

3.4.3 Hierarchical variable selection

To understand how each variable contributed to the prediction of DOC concentration and to identify potentially redundant variables, a new set of model iterations was performed (Table 3.4), starting with a model that used only CDOM absorbance as DOC predictors and re-introducing all other variables (SPC, WL, $S_{275-295}$ and S_R) one at a time. Model 2.0 used all 98 samples after determining the entire dataset was suitable from the test-set switch. The model used only absorbance from 81 wavelengths (250 to 450 nm in 2.5 nm intervals) to predict DOC concentration as it was the starting point for subsequent iterations that use additional variables. Models 2.1-2.5 then represent models where we added variables, one at a time from $S_{275-295}$, $S_{350-400}$, S_R , SPC and WL, to the prior one in a specific order, finishing with Model 2.5, which included all five additional variables. Both VIP and % contribution to SPE of each variable from the training dataset were used to determine the order to be added. When using 8-components, $S_{275-295}$ had the highest VIP and lowest % contribution to SPE (Table 3.6), so it was added in Model 2.1, followed by SPC added in Model 2. 2. The S_R was added in Model 2.3 and WL in Model 2.4. Due to the consistently higher % contribution to SPE up until component 7 from $S_{350-400}$ (30.04%), it was considered last, already included in Model 2.5 (which was first known as Model 1.8). Models 2.0-2.5 all performed similarly when predicting DOC from continuous *in-situ* data using 6- and 8-components (Figures 3.5 and

3.6). The addition of SPC into Model 2.2 resulted in a diurnal pattern observed in predictions of DOC concentration which diminish with the addition of WL with Model 2.4 when using 6-components (Figure 3.5), however, remain persistent even with the addition of WL when using 8-components (Figure 3.6). Due to minimal differences in predictions of DOC concentration between Models 2.4 and 2.5, and the consistently higher contribution to SPE from $S_{350-400}$, Model 2.4 was chosen and used moving forward.

Table 3.6: Variable importance and SPE contribution in Model 2.5

Variable importance in the projection (VIP) and % contribution to square prediction errors (SPE) for each variable from the training dataset used to train the 4-, 6-8-component partial least squares Model 2.5. The variables include spectral slopes from the 275-295 nm range ($S_{275-295}$) and the 350-400 nm range ($S_{350-400}$), the spectral slope ratio (S_R), specific conductance (SPC) and water level (WL). The reported VIP and SPE contributions for absorbance are the total of 81 wavelengths (250-450 nm in 2.5 nm intervals).

Model 2.5					
N comp.	8	4	6	7	8
Variable	VIP	% SPE Contribution			
Abs. 250-450 nm (81)	79.89	14.47%	34.83%	46.77%	93.97%
$S_{275-295}$	1.33	27.76%	17.98%	1.10%	0.16%
$S_{350-400}$	1.14	8.12%	33.38%	30.04%	3.05%
S_R	1.11	11.18%	4.41%	8.17%	1.32%
SPC	1.25	4.81%	0.18%	0.19%	0.25%
WL	0.99	33.67%	9.23%	13.74%	1.25%
Total	86	18.89	3.97	2.14	0.64

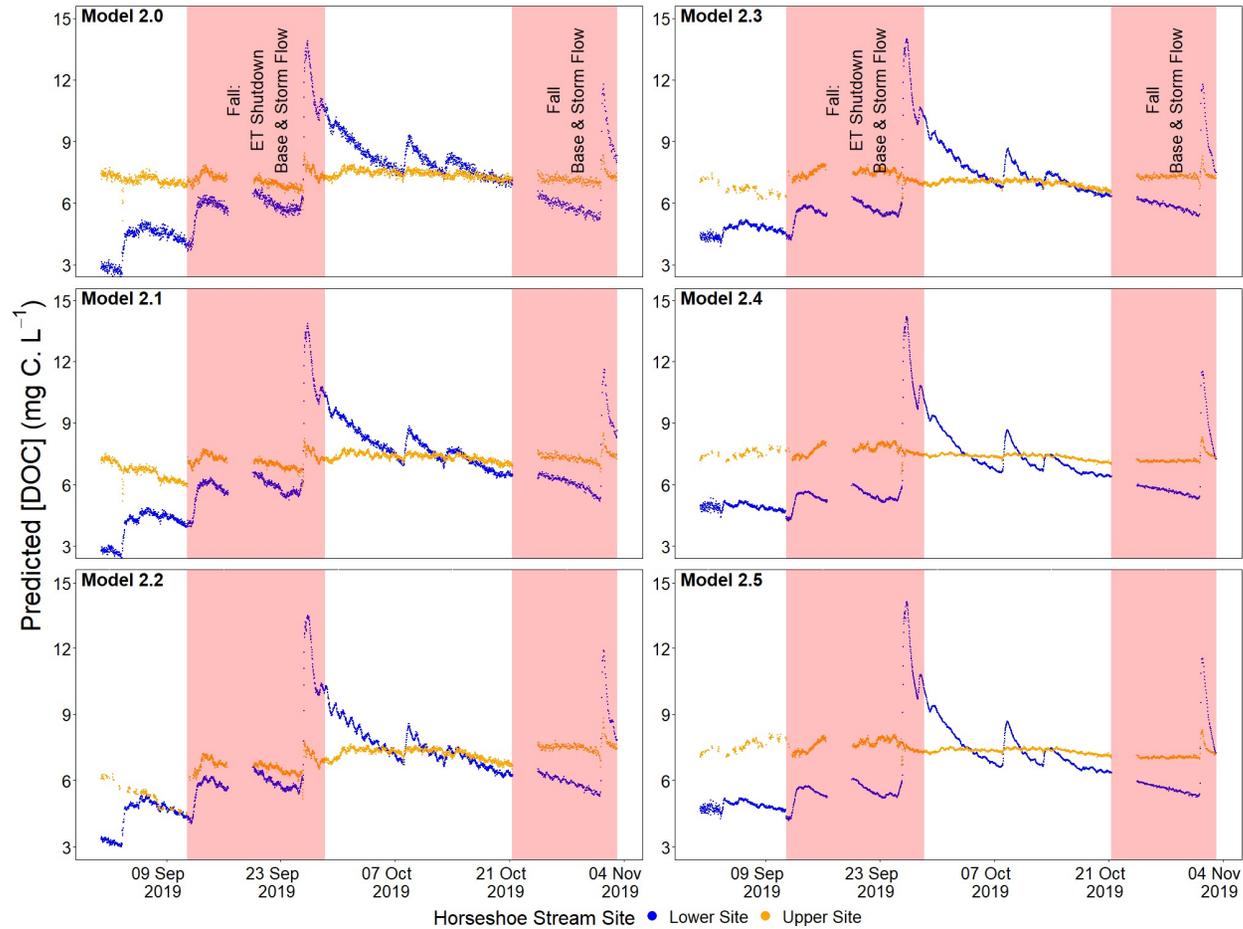


Figure 3.5: Fall period predictions of DOC concentration using 6 PLS-component hierarchical models

Continuously predicted dissolved organic carbon concentrations ([DOC]) for the Upper Horseshoe (orange) and Lower Horseshoe (blue) Brook sites for the Fall period of 2019 using 6-component partial least squares (PLS) regression models. Shaded boxes represent periods for which samples were collected and form part of the training dataset include base flow, and storm events (spikes in [DOC]). Evapotranspiration shutdown (ET Shutdown) represents the time of year when plant evapotranspiration ceases, resulting in a shift in baseflow water levels as observed in both stream sites and groundwater wells. Variables were added sequentially into the PLS regression in order of importance starting with Model 2.0 which uses only absorbance for from 250 to 450 nm to predict [DOC]. Further, Model 2.1 adds spectral slope from the 275-

295 nm range onto Model 2.0, Model 2.2 adds specific conductance onto Model 2.1, Model 2.3 adds water level onto Model 2.2, Model 2.4 adds spectral slope ratio onto Model 2.3, and Model 2.5 adds spectral slope from the 350-400 nm range to onto Model 2.4.

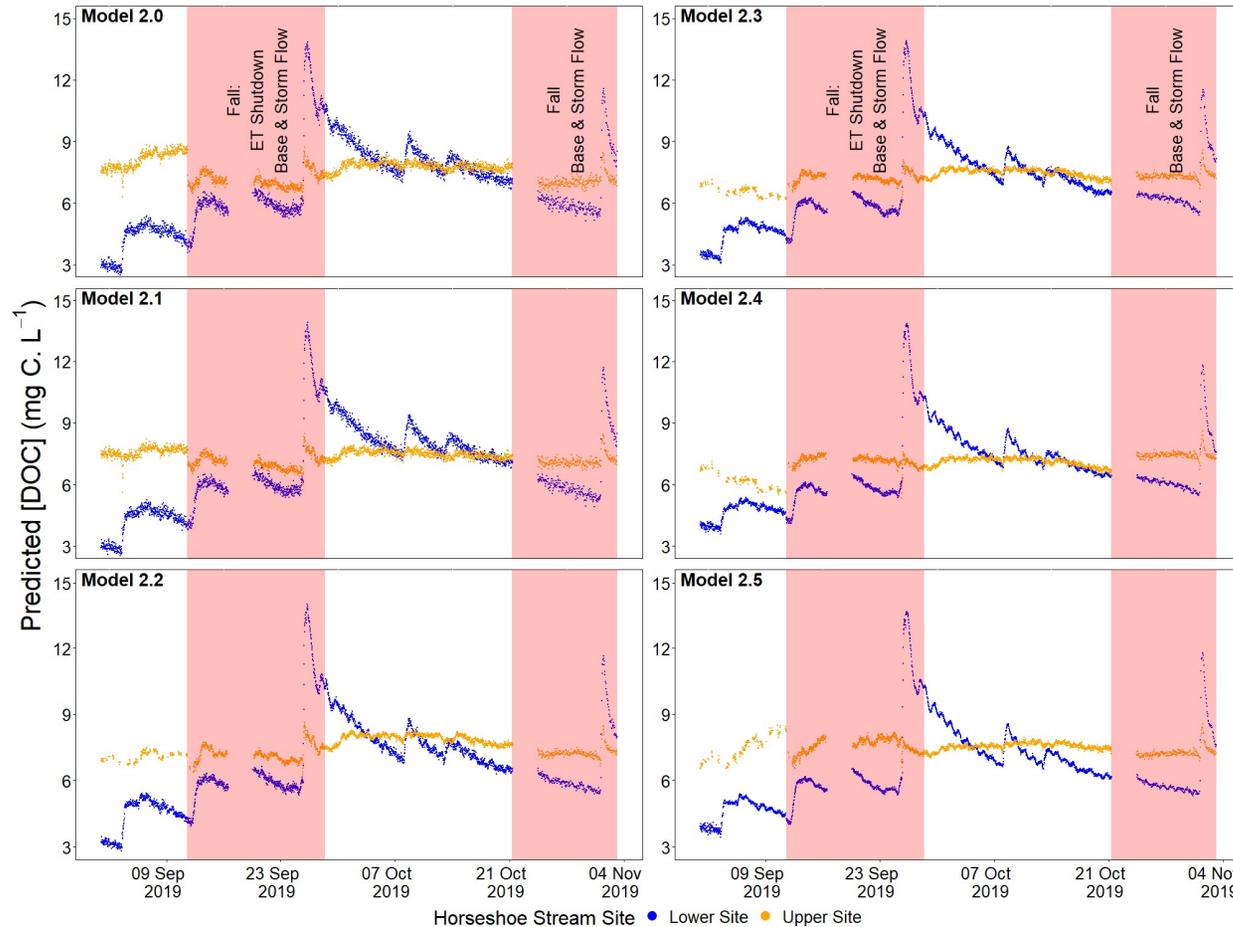


Figure 3.6: Fall period predictions of DOC concentration using 8 PLS-component hierarchical models

Continuously predicted dissolved organic carbon concentrations ([DOC]) for the Fall period of 2019 using 8-component partial least squares (PLS) regression models. Shaded boxes represent periods for which samples were collected and form part of the training dataset include base flow, and storm events (spikes in [DOC]). Evapotranspiration shutdown (ET Shutdown) represents the time of year when plant evapotranspiration ceases, resulting in a shift in baseflow water levels as observed in both stream sites and groundwater wells. Variables were added sequentially into the PLS regression in order of importance starting with Model 2.0 which uses only absorbance for from 250 to 450 nm to predict [DOC]. Further, Model 2.1 adds spectral slope from the 275-295 nm range onto Model 2.0, Model 2.2 adds specific conductance onto Model 2.1, Model 2.3 adds water level onto Model 2.2, Model 2.4 adds spectral slope ratio onto Model 2.3, and Model 2.5 adds spectral

conductance onto Model 2.1, Model 2.3 adds water level onto Model 2.2, Model 2.4 adds spectral slope ratio onto Model 2.3, and Model 2.5 adds spectral slope from the 350-400 nm range to onto Model 2.4.

The training dataset for Model 2.4 had no SPE or Hotelling's T^2 outliers, indicating there were no overly influential samples with high leverages on the model. However, T^2 outlier values were observed across the continuous *in-situ* dataset, particularly during periods and at the stream site where samples used to train the models were lacking, e.g. early Fall period (September) at UHS and winter at either site (Figure 3.7 and 2.8). The addition of $S_{275-295}$, S_R , SPC and WL to Model 2.0 (Model 2.4) helped reduce T^2 values when using 6-components, though not enough to reduce them below the calculated 95% and 99% limits (Figure 3.7), while the use of 8-components further increased T^2 values. The elevated SPE in the 6-component Model 2.4 (Figure 3.8) results from the combined contribution of $S_{275-295}$, S_R and WL (Table 3.6). Though had $S_{350-400}$ also been included, then SPE would have been higher. Using 8-components for Model 2.4 significantly reduced the number of SPE outliers observed in the continuous *in-situ* dataset, allowing the previously outlier observations to compatible with the model and suitable for predicting DOC concentration (Figure 3.8). The SPE values observed in the continuous *in-situ* dataset from Model 2.4 using 8-components is comparable to Model 2.0 using 6-components. However, Model 2.4 has the added benefit of using the additional information from the extra variables (SPC, WL, $S_{275-295}$, S_R). The $S_{350-400}$ was discarded as it was found to be the most significant contributor to SPE across training and continuous datasets even when using 8-components.

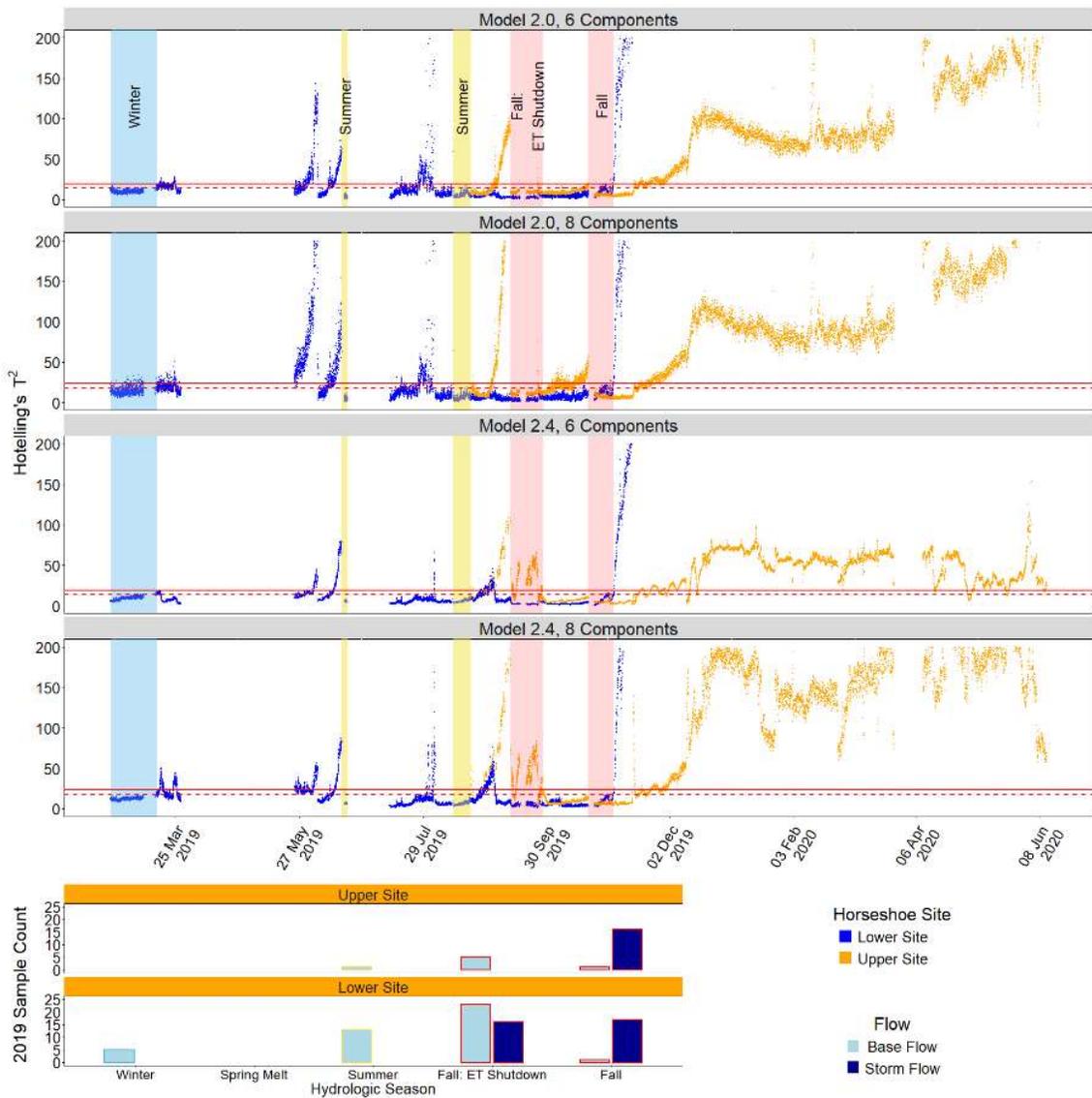


Figure 3.7: Comparing model performance using Hotelling's T^2

Hotelling's T^2 from continuous predictions using 6- and 8-component partial least squares (PLS) regression models. Shaded boxes represent periods of time for which samples were collected and form part of the training dataset. Only base flow samples were collected during Winter and Summer periods while both baseflow and storm event samples were collected during the early Fall (ET Shutdown) and late Fall periods. Variables were added sequentially into the PLS regression in order of importance. Model 2.0 uses only scatter-corrected absorbance from 250-450 nm. Model 2.4 uses the same and additionally spectral slope from the 275-295 nm range, conductivity, water level and spectral slope ratio. Dashed and solid lines represent 95% and 99% SPE limits, respectively.

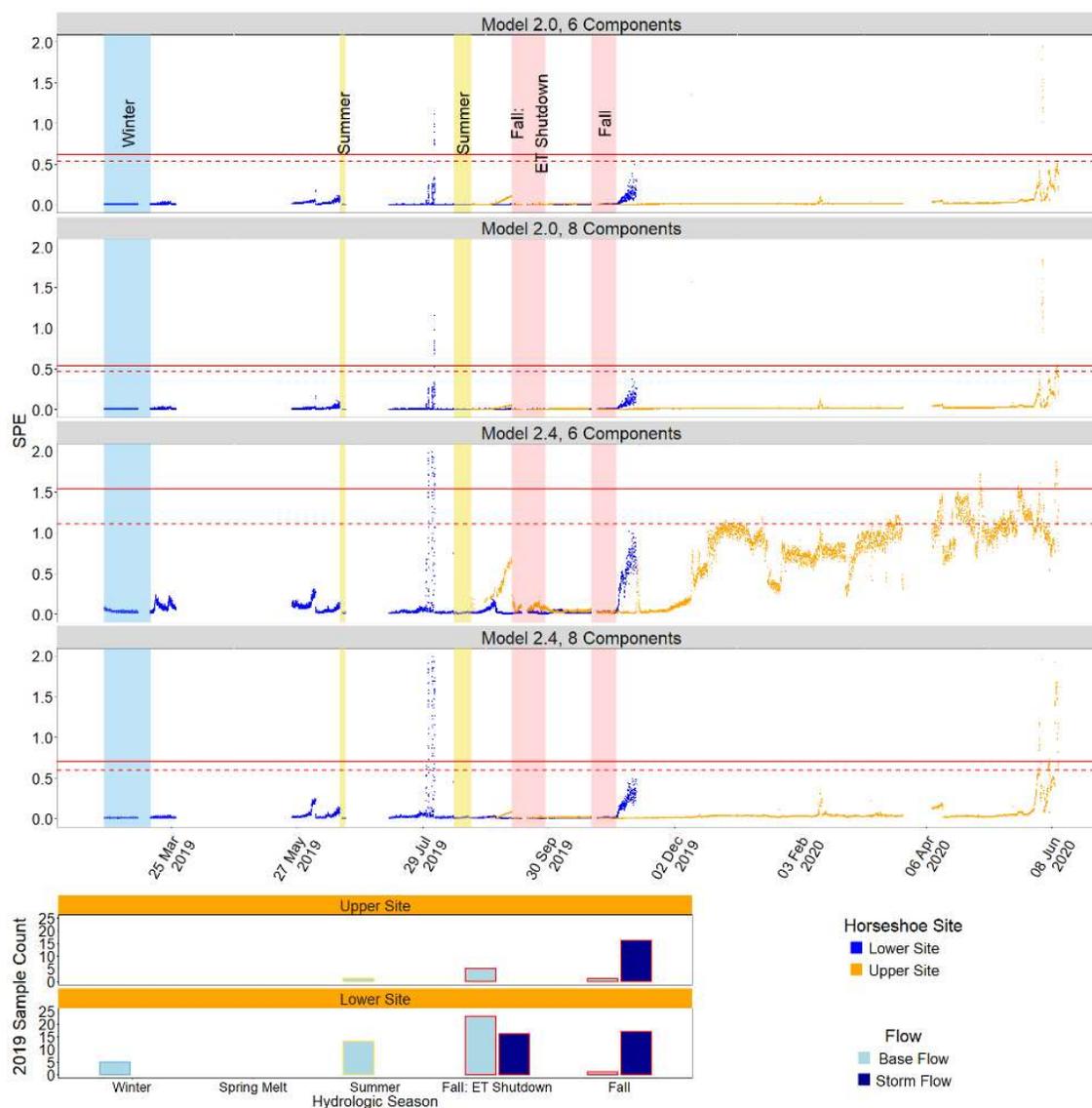


Figure 3.8: Comparing model performance using SPE

Square prediction errors (SPE) from continuous predictions using 6- and 8-component partial least squares (PLS) regression models. Shaded boxes represent periods of time for which samples were collected and form part of the training dataset. Only base flow samples were collected during Winter and Summer periods while both baseflow and storm event samples were collected during the early Fall (ET Shutdown) and late Fall periods. Variables were added sequentially into the PLS regression in order of importance. Model 2.0 uses only scatter-corrected absorbance from 250-450 nm. Model 2.4 uses the same and additionally spectral slope from the 275-295 nm range, conductivity, water level and spectral slope ratio. Dashed and solid lines represent 95% and 99% SPE limits, respectively.

3.4.4 Training a new model using turbidity corrected absorbance spectra

Turbidity due to particles can impact the absorbance spectrum of water in two ways: through light-scattering effects and absorbance from those particles (turbidity-induced absorbance). Scatter deflects light across the entire pathlength, registered as a constant absorbance by the detector at every wavelength. Turbidity-induced absorbance, however, exhibits an increasing absorbance with decreasing wavelength.

The average $A_{700-730}$ was highly correlated to turbidity-induced absorbance (Figure 3.9). The relationship between light-scattering and turbidity-induced absorbance was determined using the storm event samples (N = 49), the same samples used to train Model 1.5. The correction was applied to *in-situ* data, resulting in an absorbance spectrum comparable to a filtered sample (Figure 3.9.C). Corrected spectra from the storm samples were then used to train a new PLS model using the same variables and components as Model 1.5 (Model 3; Table 3.7). The leftover 49 baseflow samples were used as a conservative validation dataset to test the corrections' performance (Table 3.7). The correction was also applied to the continuous dataset to predict DOC concentration, T^2 and SPE, and compare them to predictions made by Model 1.5, which used the same training dataset and variables but only scatter-corrected absorbance rather than absorbance corrected for turbidity-induced absorbance. Because Model 3 was trained using the same samples as Model 1.5 and their parameters were kept constant, the only difference being the correction applied to absorbance data, they can be directly

compared. Overall, Model 3 performed worse than Model 1.5, resulting in higher RMSE and MAE and lower d_r than Model 1.5 (Table 3.7). Model 3.0 did not predict DOC adequately as predictions during specific periods were negative and large proportion of observations from the *in-situ* continuous dataset had large T^2 and SPE, significantly exceeding the suggested limits (Figure 3.10).

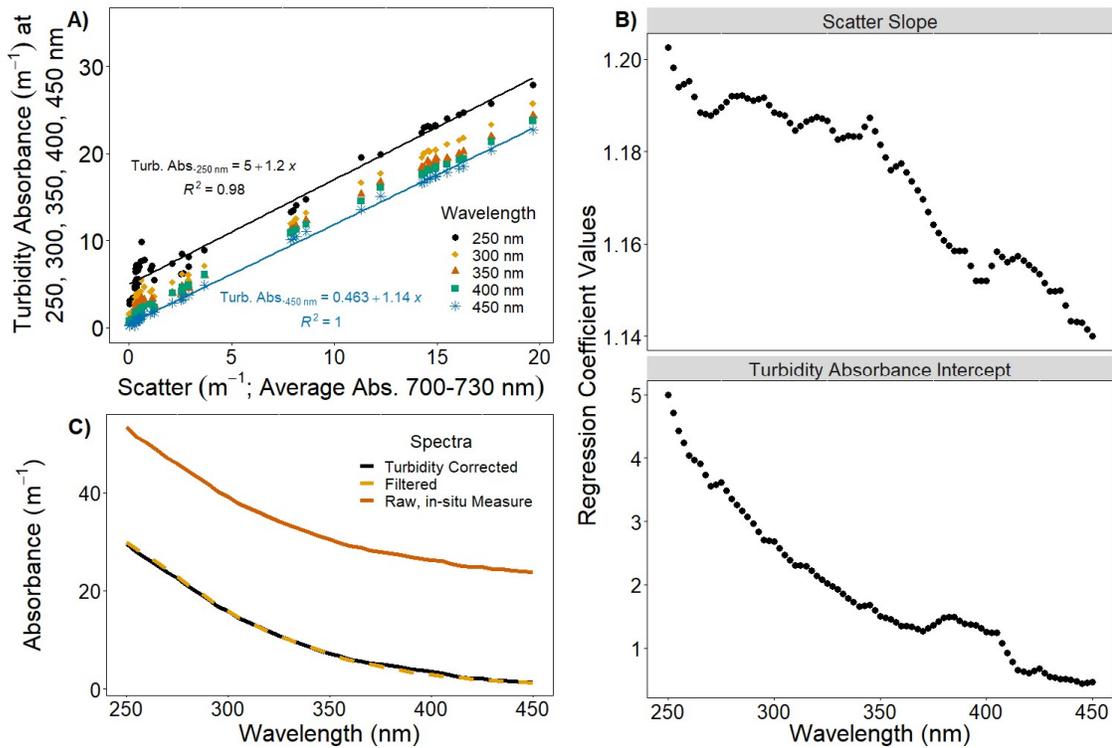


Figure 3.9: Development of a correction for turbidity-induced absorbance

A) Example derivation of scatter-to-turbidity-induced absorbance correction factor (line of best-fit) via least squares regression where the turbidity-induced absorbance from each wavelength from 250-450 nm in 5 nm intervals was regressed onto scatter derived from average *in-situ* absorbance from 700-730 nm. Turbidity-induced absorbance was calculated by subtracting filtered absorbance spectra from *in-situ* (unfiltered) absorbance spectra. B) Regression slopes and intercepts for every wavelength at each 5 nm interval from 250-450 nm as derived from turbidity-induced absorbance vs. scatter relationship in (A). B) Correction factors for 2.5 and 7.5 nm intervals were interpolated from two adjacent factors. C) Absorbance spectra for a training dataset sample where peak scatter was observed. Turbidity-induced absorbance correction was applied as derived from A). Turbidity corrected absorbance spectra used for Model 3 was then calculated by subtracting the turbidity-induced absorbance, predicted using scatter from average

absorbance at 700-730 nm as in A), from raw *in-situ* absorbance spectra for every wavelength using coefficients reported in B).

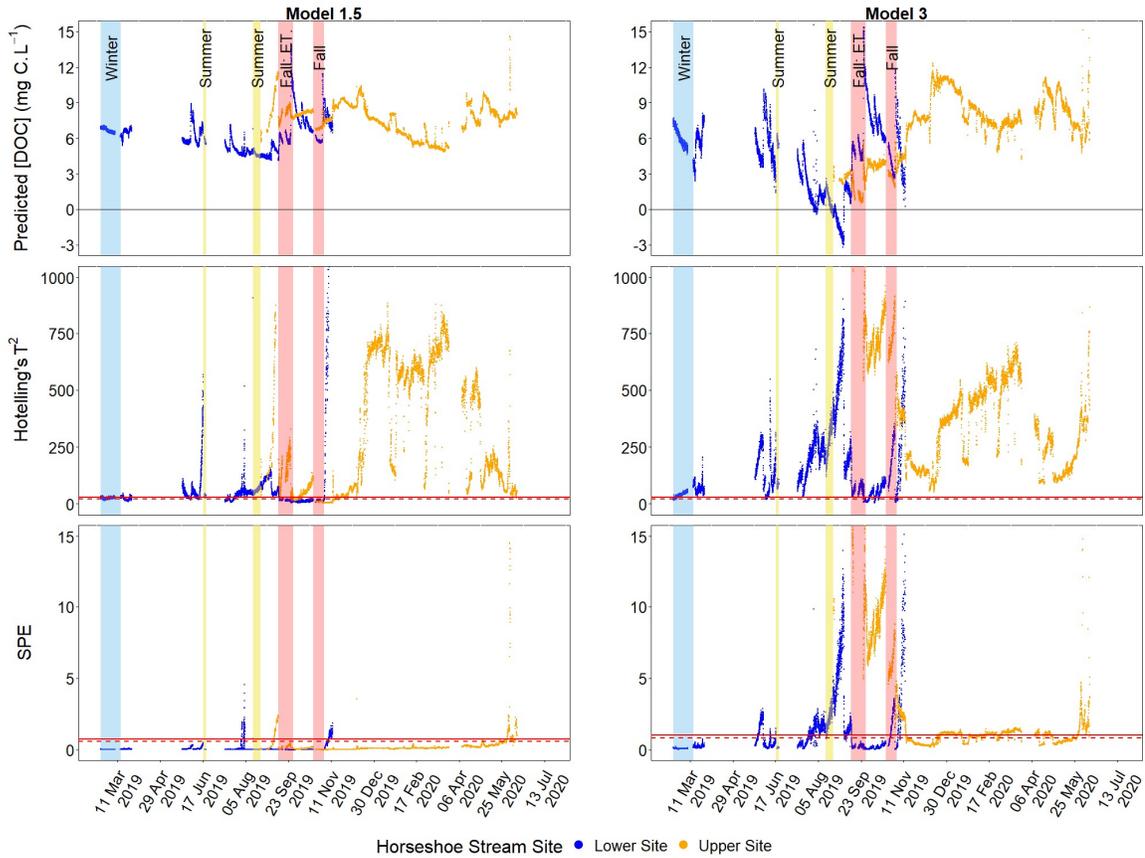


Figure 3.10: Testing performance of the correction for turbidity-induced absorbance on continuous dataset

Continuous dataset prediction results when using specific conductance, water level, spectral slope and slope ratio derived from the respective absorbance correction: scatter-corrected absorbance spectra (Model 1.5) and scatter-derived turbidity-induced absorbance-corrected absorbance spectra method (Model 3). The latter correction caused model predicted dissolved organic carbon (DOC) concentrations to become negative during some periods of the year and found to greatly increase T^2 and squared prediction errors compared to the model developed from only scatter corrected absorbance.

Table 3.7: Comparing validation dataset performance of turbidity-corrected model to scatter-corrected model

Model results for 8-component partial least squares regression models using specific conductance, water level and, either: scatter-corrected absorbance from 250-450 nm, spectral slope from 275-295 nm and slope ratio derived from the same scatter-corrected absorbance spectra (Model 1), or scatter derived turbidity-induced absorbance-corrected absorbance spectra, to predict dissolved organic carbon concentrations. Root mean square errors (RMSE) and mean absolute errors (MAE) are reported for both training and validation datasets, while Willmott's index of agreement (d_r) is reported for the validation dataset.

Models	Training dataset		Validation dataset		Training dataset		Validation dataset		
	Training dataset	N	Validation dataset	N	RMSE	MAE	RMSE	MAE	d_r
Model 1.5	All events	49	Baseflow	49	0.21	0.15	0.72	0.50	0.89
Model 3	All events	49	Baseflow	49	0.20	0.15	0.94	0.73	0.80

Model 2.4, using scatter corrected absorbance from 250-450 nm, $S_{275-295}$, S_R , water level and SPC and using 8-components from a training dataset of 49 baseflow and 49 event samples from different sites across different periods of the year to predict DOC, is considered the best performing model derived from the existing dataset in this study.

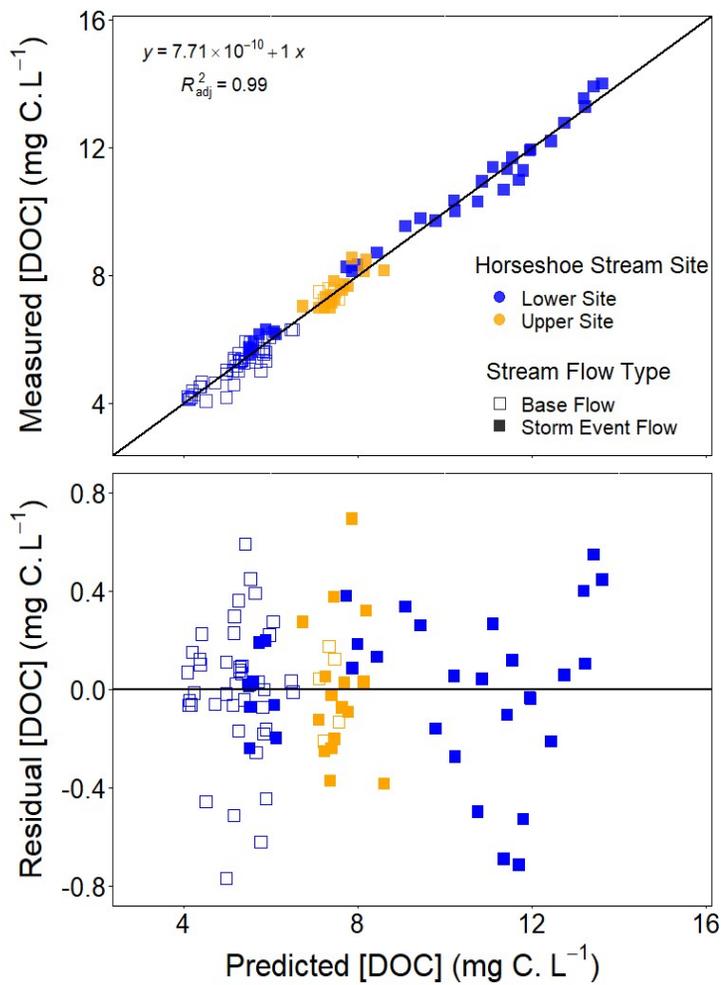


Figure 3.11: Recovery function for 8-component Model 2.4

Performance of Model 2.4 using 8-components to predict dissolved organic carbon concentration ([DOC]) and compare to laboratory-measured [DOC].

Table 3.8: Result of 2000 randomized cross-validations for Model 2.4

Results for component selection – 2000 randomized K-fold cross validations were performed and root mean square error from cross validation (RMSE-CV) were calculated from each iteration from the predicted residual error sum of squares for each component up to 20 components. The mean and standard deviation (SDEC) for the 2000 iterations are reported for training datasets. Model 2.4 was developed using scatter corrected absorbance spectra from 250-450 nm in 2.5 nm intervals, spectral slope from the 275-295 nm range ($S_{275-295}$), spectral slope ratio (S_R), specific conductance (SPC) and water level (WL). Asterisk represents number of components selected to use.

Model 2.4				
Ncomp	Full Dataset Training dataset			
	n = 96			
	RMSE-CV			
	3 CV Segments		6 CV Segments	
	Mean	SD	Mean RMSE-CV	SDEC
1	1.239	0.034	1.233	0.019
2	0.601	0.014	0.598	0.008
3	0.498	0.016	0.495	0.008
4	0.446	0.017	0.442	0.010
5	0.443	0.025	0.436	0.015
6	0.416	0.023	0.415	0.013
7	0.369	0.020	0.361	0.011
8*	0.360	0.022	0.352	0.012
9	0.360	0.031	0.348	0.018
10	0.341	0.025	0.329	0.014
11	0.344	0.025	0.329	0.015
12	0.367	0.027	0.352	0.017
13	0.385	0.028	0.368	0.019
14	0.399	0.029	0.382	0.021
15	0.407	0.030	0.387	0.022
16	0.418	0.030	0.394	0.023
17	0.430	0.030	0.406	0.024
18	0.445	0.031	0.417	0.025
19	0.462	0.032	0.436	0.026
20	0.480	0.032	0.456	0.028

3.5 Discussion

The model that had the best consistency (i.e. best performance) with the continuous dataset was the 8-component Model 2.4, which used $A_{250-450}$, $S_{275-295}$, S_R , SPC, and WL to predict DOC concentration over a range of 4 to 16 mg C L⁻¹. The optimal number of PLS-components for a model depends on how many and how different the relationships among predictor variables and DOC are, with more complex relationships or weaker correlations among variables demanding more PLS-components be characterized. Through a series of test-set switches, 8 PLS-components were identified as a potentially ideal number of components to be used based on the model's performance with a validation dataset. Model iterations following test-set switches lack a validation dataset to confirm optimal PLS-components. A series of 2000 randomized iterations of 3-fold cross-validation was then used to identify the potential optimal number of components, resulting in similar components as the test-set switch. Consistency checks like SPE were a helpful tool for identifying observations incompatible with the model, or alternately, models that were incompatible with a set of observations, and therefore not suitable for predictions. In another vein, T^2 can further evaluate compatible observations to identify potential outliers that are not represented in the training dataset, simultaneously identifying sites or periods where the model can fail to predict DOC concentration accurately and where sampling should be focused to update the model. Finally, the turbidity-induced absorbance correction derived from *in-situ* absorbance-based light-scatter (average $A_{700-730}$) from storm event samples was not suitable to be used year-

round and is likely unnecessary given the performance of the model using only scatter-corrected absorbance.

3.5.1 Additional variables in concert with absorbance are useful but may require additional model components in order to characterize them and enhance the prediction of DOC concentration

The model development indicates that as more and different variables are used in the model to predict DOC concentration (e.g. absorbance data vs. absorbance, WL and SPC), additional model components are required to capture the different relationships among the variables used. Model 2.4 incorporated $A_{250-450}$, $S_{275-295}$, S_R , WL, and SPC variables, with $S_{275-295}$ and SPC being the most important predictors of DOC concentration while also contributing to lowering inter-hourly variability among continuous predictions. Of all the additional variables, SPC had the most significant effect on predictions of DOC concentration as continuous *in-situ* predictions were observed to exhibit a diurnal oscillating pattern, reflecting a naturally occurring pattern observed in the continuous SPC dataset. Additional variables reduced the amplitude of the diurnal oscillations.

When using only 6-components, there was a significant SPE contribution from $S_{275-295}$ and WL, suggesting a relationship among $S_{275-295}$, WL and DOC was not being characterized among the first 6-components (Jackson, 1991). Conversely, it is suggestive that the first

6-components largely characterize the relationships among $A_{250-450}$ and SPC with DOC. This further suggests the relationships among $S_{275-295}$, WL, and DOC are different from $A_{250-450}$ and SPC with DOC and explain why they are characterized among the following components 7 and 8.

When applied to the continuous *in-situ* dataset, the 6-component version of Model 2.4 had poor T^2 and SPE performance, a high proportion of observations exceeding the respective statistics limits, specifically for the winter and spring period of 2020 of the UHS site. Signified as outliers based upon SPE values, data from that site and period is considered inconsistent with the model, and predictions should not be made. However, when using 8-components, the SPE contribution from $S_{275-295}$ and WL were considerably reduced and comparable to other periods of LHS data where the data are consistent and compatible with the model, and therefore predictions can be made from these data.

3.5.2 PLS model checks are used to determine what *in-situ* absorbance data we can use and then evaluate which data we should use to predict DOC concentration

First, we can evaluate modelled observations to confirm that they are compatible with the model and therefore be used to predict DOC concentration. Then, if the observations are within the SPE limit, further evaluation can be made where those found

exceeding the T^2 limit can be investigated and diagnosed for the underlying attributes that cause observations to be outside the limit set by the model's training dataset.

Evaluating modelled observations helps us determine whether they (1) are compatible with the model and therefore can be used to predict DOC concentration via SPE, and (2) to what degree predictions of DOC concentration should be scrutinized via T^2 . For example, in the case of (1), the continuous *in-situ* dataset appears mostly consistent year-round in both sites with the 8-component Model 2.4: a significant proportion of observations are well below the SPE limits and can be used to predict concentration of DOC (Figure 3.8). There is, however, a specific period where SPE suddenly increases during an intense storm in early June 2020 (Figure 3.8) following wet antecedent conditions. A storm event following a wet period where hydrologic connectivity is great could result in more runoff than normal, connecting and allowing the delivery of material from different landscape components not normally connected to the stream (Bracken & Croke, 2007; Creed & Band, 1998; Kaplan & Cory, 2016). Such an event would not be captured in the training dataset used here. Furthermore, the significant increase in discharge due to the storm allows for greater particle load and diversity of particulate matter to interfere with the CDOM spectra (via light-scattering and absorbance), enough to alter the relationships among variables to the degree that the components which initially characterized them cannot recognize. As a result, the observations made during this period at UHS are incompatible with the model and cannot be used to predict DOC concentration.

Observations that are compatible with the model can then be interpreted via T^2 , where outliers defined by exceeding the T^2 limit would have predictions of DOC concentration more scrutinized by first diagnosing what caused the elevated T^2 . For example, throughout September 2019, UHS observations exceeded the T^2 limit (Figure 3.7) due to the observations having a larger S_R than what was observed in the training dataset. This period partly corresponds to the autumnal transition: a period following the cessation of plant evapotranspiration followed by changes in landscape hydrology (Creed, McKnight, et al., 2015) as more components of the landscape become hydrologically connected. The high S_R observed at UHS during this period can be explained by the pond- and wetland-dominated landscape flushing algal-derived DOC (Franke et al., 2012; Helms et al., 2009; Villacorte et al., 2015) into stream water. Overall, this event could shift the relationship between DOC concentration and CDOM absorbance observed at UHS to one that was not captured in the training samples. The outlier T^2 observations from UHS during this period are, therefore, a result of the training dataset not being representative of UHS during this period. Similarly, outlier observations during the 2019-2020 winter from UHS were caused by unusually low SPC values that were also not represented in the training dataset. Predictions could be made from the T^2 outlier observations in these cases where they are not SPE outliers as not being an SPE outlier means the model is valid for these samples and they are similar to those used to train the model. However, predictions would be made keeping in mind that the observations are most likely outside the scope of the model and are extrapolated predictions rather than interpolated predictions (i.e.

the case of “normal”-behaving observations). To better understand the impact of using predictions from observations that are outlier for T^2 but not for SPE, further assessments are needed where test samples of these observations are obtained and compared to model predictions. Known prediction errors that remain consistent with outlying T^2 values (e.g. consistently under-predicting concentration of DOC during a specific period of the year where outlier T^2 values are consistent) can be resolved with a simple bias correction, and if enough samples are collected the model could even be re-fit to incorporate the information that represents the previously outlier period.

3.5.3 Correction for turbidity-induced absorbance likely requires year-round training data

The effects of particles impact absorbance spectra by scattering light, assumed constant absorbance across all wavelengths (using average $A_{700-730}$ as an absorbance-based measurement), and turbidity-induced absorbance where particles absorb light increasingly with decreasing wavelengths. The turbidity-induced absorbance correction derived from absorbance-based light-scatter, though initially showing promising results with the validation samples, performed poorly when applied to the year-round continuous *in-situ* data. The correction was found to overcompensate for specific periods of the year, resulting in negative predictions of DOC concentration. More noticeably, the correction appeared to interfere with the underlying relationships among DOC, $A_{250-450}$,

$S_{275-295}$, SPC, WL and S_R , and $S_{350-400}$ of *in-situ* observations that were already captured in the previous model iterations, specifically Model 2.4. The resulting model from using fully corrected *in-situ* absorbance was more incompatible with the continuous *in-situ* dataset, with a greater proportion of observations being unsuitable to be used for predicting DOC concentration.

These results highlight the temporal variability in natural turbidity-induced absorbance is likely caused by changes in the composition and quantity of particles entrained in streamflow during events. Therefore, a variety of representative samples that capture different qualities of particulate organic matter and inorganic particles are needed to develop a correction for different periods of the year. Although the demonstrated model results using scatter correction alone suggest such correction may not be needed for these two stream sites, turbidity-induced absorbance correction may be necessary in other sites that experience turbidity values above those observed in the study sites (0.5 to 6.0 NTU).

3.6 Conclusion and Implications

A PLS-based model that predicts DOC concentration from continuous *in-situ* absorbance, conductivity, water level and other optical parameters was successfully trained from targeted storm event samples and seasonal grab-samples. Through a series of test-set switches, 8 PLS-components were identified as a potentially ideal number of components to be used based on the model's performance with a validation dataset. The potential 8-components were again identified with a series of 2000 iterations of randomized 3-fold cross-validation when using the entire dataset, resulting in one of the lowest RMSE-CV being observed at 8-components. This cross-validation method, using multiple randomizations, appears to be more robust than traditional cross-validation used to determine how many model components to use when there is no validation dataset available.

Model consistency checks that are built into PLS models, Hotelling's T^2 and SPE, were found to be very useful for understanding the representativeness of the training dataset when the model was applied to a continuous *in-situ* dataset. These consistency checks further allow for the quick and easy identification of outlier observations while also providing an easy method of diagnosing and understanding why the observation is an outlier. It is through this that these statistics can show the representativeness of the training dataset and identify data gaps from the training dataset, identifying where or when or the model may fail to perform. The identified gaps further inform when and

where sampling efforts should be focused for adding to the training dataset and updating the PLS model or when bias corrections could be used. For example, here it was demonstrated how T^2 outlier observations from UHS during the early Fall period (September) were outliers due to elevated S_R , which algal-derived DOC inputs from a nearby pond can explain, flushing DOC, built up throughout the growing season, as water level and discharge increase throughout September.

Other variables tested in concert with $A_{250-450}$ identified SPC and $S_{275-295}$ as very important variables when developing the model and predicting DOC concentration. Water level and S_R were slightly less critical than the former, while $S_{350-400}$ was found to contribute significantly to SPE when trying to use less than eight components. Of all the additional variables, SPC appeared to have the most significant impact on DOC predictions from the continuous dataset as a diurnal pattern in DOC concentration was observed in the *in-situ* dataset. However, the magnitude of the diurnal pattern decreased with each addition of a new variable. Another effect the additional variables had on stream DOC predictions was a significant reduction in inter-hour variance; the additional variables had a stabilizing effect on predictions of DOC concentration and appeared to reduce random noise between measurement intervals.

Though successful with the conservative validation dataset of baseflow samples, the turbidity correction developed here was not appropriate for year-round continuous *in-situ* data as it was found to overcompensate the correction mostly during seasons outside of fall, resulting in negative predictions of DOC concentration from *in-situ* observations.

The training dataset used to develop the correction was composed of Fall storm event samples, suggesting the quality and quantity of particulate matter observed in these sites are temporally variable, most likely according to season.

When applied to continuous *in-situ* data, Model 2.4 exhibited predictions with consistently low SPE values, suggesting the observations follow the model structure very closely and predictions of DOC concentration can be made from the data. Only during a period in early June 2020, corresponding to an extreme storm event, did SPE increase beyond the threshold, most likely due to a “unique” CDOM signature from the integration of surrounding ponds and wetlands being more hydrologically connected to UHS during this powerful event. Similarly, T² identified UHS as a key sampling location as well as winter and spring melt in either site as critical sampling periods to increase training dataset representativeness and more confidently predict DOC concentration. Together, these metrics evaluated the performance of the model when applied to continuous *in-situ* data to confidently make predictions of DOC concentration in LHS during summer and fall periods and UHS in the late fall, while less confident predictions can be made during winter in either site and from UHS during the early fall.

4 Conclusion

This thesis contributes to the present gaps in the literature on developing partial least squares (PLS) models to continuously monitor the concentration of *in-situ* dissolved organic carbon (DOC) using the absorbance of chromophoric dissolved organic matter (CDOM) in headwater streams. Specifically, using the full CDOM spectra and other stream water quality parameters to inform model predictions of the dynamic relationship between CDOM absorbance and DOC concentration: A phenomenon observed in headwater streams that variably integrate different components of the landscape depending on hydrology dictated by season and weather. Furthermore, it contributes to identifying and testing robust model performance metrics that relate to a model's prediction accuracy (refined index of agreement; d_r) and can quantify uncertainty (prediction intervals) to more rigorously evaluate a model's performance or support interpretability of the root mean square error (mean absolute error; MAE) to compare model residuals. This thesis also demonstrates the use of consistency checks built into PLS models (Hotelling's T^2 and SPE) for predictions in a natural stream setting. For example, they identify where and when caution should be taken when using the model to make predictions. Furthermore, they inform how the model can be updated based on the training dataset's representativeness by highlighting periods and sites from continuous *in-situ* datasets where sampling efforts should be focused. These metrics can help describe model suitability, performance, and prediction accuracy when it is impossible to

access remote sites and test samples cannot be collected to assess model predictions directly. These key findings and what they suggest are valuable future avenues of study are discussed in the following section.

This thesis addressed four notions:

1. How comparable are PLS methods to typical multiple linear regression (MLR) methods in predicting DOC from CDOM absorbance?

Though PLS methods are considered more robust for chemometrics, there was not much difference in the validation dataset RMSE of predictions (PLS: 0.52 mg C L⁻¹, MLR: 0.56-0.66 mg C L⁻¹; Chapter 2). Prediction intervals and index of agreement from the best performing MLR model were equal to the PLS model. Due to the limited sample diversity of the validation dataset (i.e. all baseflow condition samples), the similarity in performance between these methods may not be equal when tested on more diverse sample sets, such as those derived from high-frequency storm samples.

PLS models' further advantage is the built-in consistency checks (T^2 and SPE) that identify and diagnose outlier samples that may be incompatible with the model and should not be used to predict DOC. Using a PLS model to predict continuous *in-situ* DOC, I used these metrics to identify periods when the model cannot be used or should be used with caution, further informing the representativeness of the training dataset. I further diagnosed outlier observations from the continuous dataset to understand which variables contribute to elevating SPE, or which component scores contribute to elevating

T^2 , and to determine which variables contribute to elevating the component-score magnitudes that contribute to elevating T^2 . The ability to diagnose outliers allowed for a better understanding of the overall studied system as it quantified why a particular sample or subset of data are outliers to the model and how they are different from non-outlier samples. The diagnoses of outliers also informed variable selection as a redundant variable that consistently contribute to SPE outliers was discarded.

2. New directions of error quantification and model evaluation for biogeochemistry: what metrics are commonly used in other fields that were newly applied here?

An alternative to R^2 had to be identified as it can be arbitrarily large or small and not necessarily related to prediction accuracy while also redundant for model selection as the models tested in each chapter had very similar R^2 . Similarly, a metric to quantify uncertainty had to be tested as it is imperative to have accurate predictions of DOC concentration with uncertainty quantified, to later extrapolate and used for catchment-wide DOC flux estimates. Model performance metrics explored were the Nash-Sutcliffe model efficiency coefficient (NSE) and Wilmott's refined index of agreement (d_r), as well as using 95% prediction intervals to quantify prediction uncertainty (Chapter 2). It was found that NSE was identical to R^2 for these models, and therefore not very informative as these predicting models result in high NSE (and R^2) with little comparability between them. The d_r was a more robust metric to compare models than NSE as it was more conservative and penalized prediction errors more heavily. Mean absolute error (MAE)

was a component for calculating d_r and was informative when reported combined with RMSE to evaluate models (Chapter 3). This was because RMSE measures the spread of residuals while MAE measures the residuals' central tendency.

The prediction intervals all approximated two times the training dataset RMSE and were therefore not as informative of model performance and prediction uncertainty. However, this may be due to the method of calculation in this particular case. Leverage-based methods for estimating prediction uncertainty (De Vries & Ter Braak, 1995; K. Faber & Kowalski, 1997; Høy et al., 1998) can be more robust for models that use a large number of predictors, such as PLS models (De Vries & Ter Braak, 1995; K. Faber & Kowalski, 1997). Therefore, it is recommended for future efforts to determine prediction uncertainty to be focused in testing these leverage-based methods as they could be more appropriate for these datasets that use CDOM absorbance from a spectrum of wavelengths and other variables (specific conductance, water level, spectral slope and slope ratio).

For specific periods of the year, remote sites may be impossible to access, and therefore grab-samples for the intention of testing and confirming model predictions may not be collected. With only continuous *in-situ* data available, consistency checks that are built into the PLS model were calculated from the same input data (Chapter 3). Hotelling's T^2 and SPE allowed me to verify the model's applicability with *in-situ* data and determine a level of confidence in the predictions. At the same time, it highlighted sites and periods for which sampling effort should be focused to improve the model's applicability by improving the training dataset's representativeness. I demonstrated that T^2 and SPE can

be used to monitor predictions of DOC concentration by identifying and diagnosing outlier samples that either: (1) were incompatible with the model and therefore should not be used to predict, or (2) were outside the scope of the model, and therefore predictions should be made with caution as they were extrapolated.

By identifying incompatible samples (SPE outliers) or those outside the model's scope (T^2 outliers), I assessed the representativeness of the training dataset and identified data gaps where and when the model might fail to predict DOC concentration accurately. This was done by diagnosing the outlier data points by determining which variable contributes to elevated SPE or which raw data contribute to the elevated T^2 . These diagnoses identified specifically which variables, and during which periods, they contribute to the sample being an outlier and therefore which periods sampling efforts should be focused for updating the model. In this case, outlier observations from a particular site (Upper Horseshoe Brook) during a specific period (early autumn; September) were caused by high spectral slope ratio values. The sparse number of UHS samples collected during this period, all of which were baseflow, were not enough to represent the likely unique CDOM signatures observed then.

Because these metrics are built into the model and calculated from input data, they can be of vital importance for predicting models used in remote sites where sampling may be impossible and validating measurements cannot be made. They provided information about observations made during inaccessible periods or at inaccessible sites where grab-samples to compare with model predictions directly may not be readily available. Using

these metrics can improve confidence in predictions or warn users that predictions may be inaccurate or unusable. Site- and period-specific trends of these metrics can be further studied to label the causes of anomalous periods of data and give contextual knowledge about predictions made during that time and improve the confidence of predictions.

3. What is the optimal combination of *in-situ* stream water parameters (e.g. conductivity, water level, spectral slopes and slope ratio) to be used as predictor variables for predicting DOC from *in-situ* absorbance?

Headwater streams are intimately connected to the surrounding landscape and have temporally variable DOC sources as different landscape components are connected depending on season and storm events. As a result, the relationship between DOC concentration and CDOM absorbance is variable, while using the full CDOM spectra to predict DOC helps mitigate these problems. However, other interferences with absorbance continue to affect the relationship between CDOM and DOC, even when using the whole CDOM spectra to make predictions. Other metrics that were related to DOC (water level, conductivity, spectral slope) or DOC composition (spectral slopes and their ratios) helped mitigate further disconnects in the relationship between CDOM absorbance and DOC concentration, supporting CDOM absorbance to make more accurate predictions (Chapter 3). The most robust model used specific conductance, water level, the spectral slope of the 275 to 295 nm range and the ratio of the spectral slopes in the 275 to 295 and 350 to 400 nm range. This model resulted in much less inter-prediction variance than a model that did not use them, suggesting the additional

variables help reduce random noise among the *in-situ* absorbance from the 81 wavelengths used, stabilizing the variability between predictions. The 350 to 400 nm range's spectral slope was discarded as it contributed significantly to SPE (Chapter 3) and more than doubled the validation dataset RMSE (Chapter 2) when included as a predictor in a PLS model. Due to site differences in *in-situ* pH behaviour, it was a poor predictor and not used in this model to predict stream DOC across sites (Chapter 3). However, the strong correlation between DOC concentration and pH for the lower Horseshoe Brook stream site suggests it would be a good predictor and should be considered when developing site-specific models in the future.

4. Can we determine an absorbance-based turbidity correction and apply it to *in-situ* absorbance data to predict DOC more accurately?

Due to the low range of turbidity observed in these sites (0.5 to 6.0 NTU) and the success of a simple scatter correction derived from the average absorbance from 700 to 730 nm, a full turbidity correction may not be currently necessary for these sites. However, in other systems or stream sites where a greater range of turbidity is observed, pre-processing of absorbance data via a turbidity correction is necessary before being useable. We were able to determine an absorbance-based correction for turbidity using a training dataset composed of samples of the autumn storm samples and validated using baseflow stream samples. However, after model re-training and application to a continuous dataset, the performance was inadequate as it over-estimated turbidity in specific periods, resulting in negative predictions of DOC concentration. The performance

during the autumn was appropriate and expected considering the training dataset were autumn samples. The correction only working during autumn suggested it is seasonally, and therefore temporally, variable. Therefore, future tests would need targeted event samples throughout different periods of the year to acquire a representative dataset that encompasses a more extensive range of turbidity quantities and qualities observed throughout the year.

This study's findings have demonstrated an approach for developing PLS models to continuously predict DOC concentration from *in-situ* data that could also be used to continuously monitor similar metrics, such as nitrate or dissolved organic nitrogen. It further highlighted how this approach can be used for identifying where training datasets can be improved to use models in other systems or periods of the year. Further efforts to predict stream DOC fluxes from these sites to quantify carbon export from the landscape will be made using these model-predicted DOC concentrations with stream discharge data. Further steps will be taken to apply the model in other sites to determine how applicable the current model is to the new sites by using T^2 and SPE and determine how to improve the model to work in different sites and periods of the year. Over time, as DOC concentration, predicted from these models, is closely monitored in these sites, we can begin to observe and quantify landscape responses to climate change. The findings of this thesis encourage the use PLS models to continuously monitor *in-situ* DOC concentration and the consistency checks built into PLS models, T^2 and SPE, to monitor model

performance, diagnose outliers and identify when and how to update these models to improve their performance.

5 References

- Ågren, A., Buffam, I., Berggren, M., Bishop, K., Jansson, M., & Laudon, H. (2008). Dissolved organic carbon characteristics in boreal streams in a forest-wetland gradient during the transition between winter and summer. *Journal of Geophysical Research: Biogeosciences*, *113*(3), 1–11. <https://doi.org/10.1029/2007JG000674>
- Ågren, A., Buffam, I., Jansson, M., & Laudon, H. (2007). Importance of seasonality and small streams for the landscape regulation of dissolved organic carbon export. *Journal of Geophysical Research: Biogeosciences*, *112*(3), 1–11. <https://doi.org/10.1029/2006JG000381>
- Ågren, A., Haei, M., Köhler, S. J., Bishop, K., & Laudon, H. (2010). Regulation of stream water dissolved organic carbon (DOC) concentrations during snowmelt; The role of discharge, winter climate and memory effects. *Biogeosciences*, *7*(9), 2901–2913. <https://doi.org/10.5194/bg-7-2901-2010>
- Alvarez-Cobelas, M., Angeler, D. G., Sánchez-Carrillo, S., & Almendros, G. (2012). A worldwide view of organic carbon export from catchments. *Biogeochemistry*, *107*(1–3), 275–293. <https://doi.org/10.1007/s10533-010-9553-z>
- Asmala, E., Stedmon, C. A., & Thomas, D. N. (2012). Linking CDOM spectral absorption to dissolved organic carbon concentrations and loadings in boreal estuaries. *Estuarine, Coastal and Shelf Science*, *111*, 107–117. <https://doi.org/10.1016/j.ecss.2012.06.015>

- Avagyan, A., Runkle, B. R. K., & Kutzbach, L. (2014). Application of high-resolution spectral absorbance measurements to determine dissolved organic carbon concentration in remote areas. *Journal of Hydrology*, *517*, 435–446. <https://doi.org/10.1016/j.jhydrol.2014.05.060>
- Baalousha, M., Motelica-Heino, M., & Coustumer, P. Le. (2006). Conformation and size of humic substances: Effects of major cation concentration and type, pH, salinity, and residence time. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, *272*(1–2), 48–55. <https://doi.org/10.1016/j.colsurfa.2005.07.010>
- Battin, T. J., Luysaert, S., Kaplan, L. A., Aufdenkampe, A. K., Richter, A., & Tranvik, L. J. (2009). The boundless carbon cycle. *Nature Geoscience*, *2*(9), 598–600. <https://doi.org/10.1038/ngeo618>
- Bowering, K. L., Edwards, K. A., Prestegard, K., Zhu, X., & Ziegler, S. E. (2020). Dissolved organic carbon mobilized from organic horizons of mature and harvested black spruce plots in a mesic boreal region. *Biogeosciences*, *17*(3), 581–595. <https://doi.org/10.5194/bg-17-581-2020>
- Bracken, L. J., & Croke, J. (2007). The concept of hydrological connectivity and its contribution to understanding runoff-dominated geomorphic systems. *Hydrological Processes*, *21*(13), 1749–1763. <https://doi.org/10.1002/hyp.6313>
- Brereton, R. G. (2003). *and Chemical Plant* (Vol. 8).

- Butman, D., Stackpoole, S., Stets, E., McDonald, C. P., Clow, D. W., & Striegl, R. G. (2016). Aquatic carbon cycling in the conterminous United States and implications for terrestrial carbon accounting. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(1), 58–63.
<https://doi.org/10.1073/pnas.1512651112>
- Chen, Q., Kruger, U., Meronk, M., & Leung, A. Y. T. (2004). Synthesis of T2 and Q statistics for process monitoring. *Control Engineering Practice*, *12*(6), 745–755.
<https://doi.org/10.1016/j.conengprac.2003.08.004>
- Chi, J., Nilsson, M. B., Laudon, H., Lindroth, A., Wallerman, J., Fransson, J. E. S., Kljun, N., Lundmark, T., Ottosson Löfvenius, M., & Peichl, M. (2020). The Net Landscape Carbon Balance—Integrating terrestrial and aquatic carbon fluxes in a managed boreal forest landscape in Sweden. *Global Change Biology*, *26*(4), 2353–2367.
<https://doi.org/10.1111/gcb.14983>
- Codden, C. J., Snauffer, A. M., Mueller, A. V., Edwards, C. R., Thompson, M., Tait, Z., & Stubbins, A. (2020). Predicting dissolved organic carbon concentration in a dynamic salt marsh creek via machine learning. *Limnology and Oceanography: Methods*.
<https://doi.org/10.1002/lom3.10406>
- Cole, J. J., Prairie, Y. T., Caraco, N. F., McDowell, W. H., Tranvik, L. J., Striegl, R. G., Duarte, C. M., Kortelainen, P., Downing, J. A., Middelburg, J. J., & Melack, J. (2007). Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget.

Ecosystems, 10(1), 171–184. <https://doi.org/10.1007/s10021-006-9013-8>

Creed, I. F., & Band, L. E. (1998). Export of nitrogen from catchments within a temperate forest: Evidence for a unifying mechanism regulated by variable source area dynamics. *Water Resources Research*, 34(11), 3105–3120. <https://doi.org/10.1029/98WR01924>

Creed, I. F., Hwang, T., Lutz, B., & Way, D. (2015). Climate warming causes intensification of the hydrological cycle, resulting in changes to the vernal and autumnal windows in a northern temperate forest. *Hydrological Processes*, 29(16), 3519–3534. <https://doi.org/10.1002/hyp.10450>

Creed, I. F., McKnight, D. M., Pellerin, B. A., Green, M. B., Bergamaschi, B. A., Aiken, G. R., Burns, D. A., Findlay, S. E. G., Shanley, J. B., Striegl, R. G., Aulenbach, B. T., Clow, D. W., Laudon, H., McGlynn, B. L., McGuire, K. J., Smith, R. A., & Stackpoole, S. M. (2015). The river as a chemostat: Fresh perspectives on dissolved organic matter flowing down the river continuum. *Canadian Journal of Fisheries and Aquatic Sciences*, 72(8), 1272–1285. <https://doi.org/10.1139/cjfas-2014-0400>

Cruciani, G., Baroni, M., Clementi, S., Costantino, G., Riganelli, D., & Skagerberg, B. (1992). Predictive ability of regression models. Part I: Standard deviation of prediction errors (SDEP). *Journal of Chemometrics*, 6(6), 335–346. <https://doi.org/10.1002/cem.1180060604>

De Vries, S., & Ter Braak, C. J. F. (1995). Prediction error in partial least squares regression :

a critique on the deviation used in The Unscrambler. *Chemometrics and Intelligent Laboratory Systems*, 30, 239–245.

Downing, B. D., Pellerin, B. A., Bergamaschi, B. A., Saraceno, J. F., & Kraus, T. E. C. (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*, 10(OCTOBER), 767–775.
<https://doi.org/10.4319/lom.2012.10.767>

Dunn, K. (2019). Process Improvement using Data. *Http://Learnche.Org/Pid*, April, 381.
learnche.mcmaster.ca/pid

Etheridge, J. R., Birgand, F., Osborne, J. A., Osburn, C. L., Burchell, M. R., & Irving, J. (2014). Using in situ ultraviolet-visual spectroscopy to measure nitrogen, carbon, phosphorus, and suspended solids concentrations at a high frequency in a brackish tidal marsh. *Limnology and Oceanography: Methods*, 12(1 JAN), 10–22.
<https://doi.org/10.4319/lom.2014.12.10>

Faber, K., & Kowalski, B. R. (1997). Propagation of measurement errors for the validation of predictions obtained by principal component regression and partial least squares. *Journal of Chemometrics*, 11(3), 181–238. [https://doi.org/10.1002/\(SICI\)1099-128X\(199705\)11:3<181::AID-CEM459>3.0.CO;2-7](https://doi.org/10.1002/(SICI)1099-128X(199705)11:3<181::AID-CEM459>3.0.CO;2-7)

Faber, N. M., & Rajkó, R. (2007). How to avoid over-fitting in multivariate calibration-The conventional validation approach and an alternative. *Analytica Chimica Acta*, 595(1-

2 SPEC. ISS.), 98–106. <https://doi.org/10.1016/j.aca.2007.05.030>

Fichot, C. G., & Benner, R. (2011). A novel method to estimate DOC concentrations from CDOM absorption coefficients in coastal waters. *Geophysical Research Letters*, *38*(3), n/a-n/a. <https://doi.org/10.1029/2010GL046152>

Franke, D., Hamilton, M. W., & Ziegler, S. E. (2012). Variation in the photochemical lability of dissolved organic matter in a large boreal watershed. *Aquatic Sciences*, *74*(4), 751–768. <https://doi.org/10.1007/s00027-012-0258-3>

Friedlingstein, P., Sullivan, M. O., Jones, M. W., Andrew, R. M., & Hauck, J. (2020). *Global Carbon Budget 2020*. *2020*, 3269–3340.

Grayson, R. P., & Holden, J. (2016). Improved automation of dissolved organic carbon sampling for organic-rich surface waters. *Science of the Total Environment*, *543*, 44–51. <https://doi.org/10.1016/j.scitotenv.2015.10.149>

Green, S. A., & Blough, N. V. (1994). Optical absorption and fluorescence properties of chromophoric dissolved organic matter in natural waters. *Limnology and Oceanography*, *39*(8), 1903–1916. <https://doi.org/10.4319/lo.1994.39.8.1903>

Helms, J. R., Stubbins, A., Ritchie, J. D., Minor, E. C., Kieber, D. J., Mopper, K., John, R. H., Stubbins, A., Ritchie, J. D., Minor, E. C., Kieber, D. J., & Mopper, K. (2009). Erratum: Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter (*Limnology and*

Oceanography 53 955-969). *Limnology and Oceanography*, 54(3), 1023.
<https://doi.org/10.4319/lo.2009.54.3.1023>

Hinton, A. M. J., Schiff, S. L., English, M. C., Biogeochemistry, S., & May, N. (1998). Sources and Flowpaths of Dissolved Organic Carbon during Storms in Two Forested Watersheds of the Precambrian Shield Published by: Springer Stable URL : <http://www.jstor.org/stable/1469533> REFERENCES Linked references are available on JSTOR for this arti. *Biogeochemistry*, 41(2), 175–197.

Høy, M., Steen, K., & Martens, H. (1998). Review of partial least squares regression prediction error in Unscrambler. *Chemometrics and Intelligent Laboratory Systems*, 44(1–2), 123–133. [https://doi.org/10.1016/S0169-7439\(98\)00163-4](https://doi.org/10.1016/S0169-7439(98)00163-4)

Huber, E., & Frost, M. (1998). Light scattering by small particles. *Journal of Water Supply: Research and Technology - AQUA*, 47(2), 87–94.
<https://doi.org/10.2166/aqua.1998.14>

Inserillo, E. A., Green, M. B., Shanley, J. B., & Boyer, J. N. (2017). Comparing catchment hydrologic response to a regional storm using specific conductivity sensors. *Hydrological Processes*, 31(5), 1074–1085. <https://doi.org/10.1002/hyp.11091>

Jackson, J. E. (1991). A User's Guide to Principal Components. In *The Journal of the Operational Research Society* (Vol. 43, Issue 6). <https://doi.org/10.2307/2583020>

Jackson, J. E., & Mudholkar, G. S. (1979). Control procedures for residuals associated with

principal component analysis. *Technometrics*, 21(3), 341–349.
<https://doi.org/10.1080/00401706.1979.10489779>

Jennings, E., Järvinen, M., Allott, N., Arvola, L., Moore, K., Naden, P., Aonghusa, C. N., Nöges, T., & Weyhenmeyer, G. A. (2010). Impacts of Climate on the Flux of Dissolved Organic Carbon from Catchments. In G. George (Ed.), *The Impact of Climate Change on European Lakes* (pp. 199–220). Springer Netherlands.
https://doi.org/10.1007/978-90-481-2945-4_12

Jeong, J. J., Bartsch, S., Fleckenstein, J. H., Matzner, E., Tenhunen, J. D., Lee, S. D., Park, S. K., & Park, J. H. (2012). Differential storm responses of dissolved and particulate organic carbon in a mountainous headwater stream, investigated by high-frequency, in situ optical measurements. *Journal of Geophysical Research: Biogeosciences*, 117(3), 1–13. <https://doi.org/10.1029/2012JG001999>

Jollymore, A., Johnson, M. S., & Hawthorne, I. (2012). Submersible UV-Vis spectroscopy for quantifying stream water organic carbon dynamics: Implementation and challenges before and after forest harvest in a headwater stream. *Sensors*, 12(4), 3798–3813. <https://doi.org/10.3390/s120403798>

Juhls, B., Paul Overduin, P., Hölemann, J., Hieronymi, M., Matsuoka, A., Heim, B., & Fischer, J. (2019). Dissolved organic matter at the fluvial-marine transition in the Laptev Sea using in situ data and ocean colour remote sensing. *Biogeosciences*, 16(13), 2693–2713. <https://doi.org/10.5194/bg-16-2693-2019>

- Kaistha, N., & Upadhyaya, B. R. (2001). Incipient fault detection and isolation of field devices in nuclear power systems using principal component analysis. *Nuclear Technology*, 136(2), 221–230. <https://doi.org/10.13182/NT01-A3240>
- Kaplan, L. A., & Cory, R. M. (2016). Dissolved Organic Matter in Stream Ecosystems: Forms, Functions, and Fluxes of Watershed Tea. Forms, Functions, and Fluxes of Watershed Tea. In *Stream Ecosystems in a Changing Environment*. Elsevier Inc. <https://doi.org/10.1016/B978-0-12-405890-3.00006-3>
- Kobayashi, D. (1986). Separation of a snowmelt hydrograph by stream conductance. *Journal of Hydrology*, 84(1–2), 157–165. [https://doi.org/10.1016/0022-1694\(86\)90049-1](https://doi.org/10.1016/0022-1694(86)90049-1)
- Lal, R. (2005). Forest soils and carbon sequestration. *Forest Ecology and Management*, 220(1–3), 242–258. <https://doi.org/10.1016/j.foreco.2005.08.015>
- Langergraber, G., Fleischmann, N., & Hofstädter, F. (2003). A multivariate calibration procedure for UV/VIS spectrometric quantification of organic matter and nitrate in wastewater. *Water Science and Technology*, 47(2), 63–71. <https://doi.org/10.2166/wst.2003.0086>
- Laudon, H., Berggren, M., Ågren, A., Buffam, I., Bishop, K., Grabs, T., Jansson, M., & Köhler, S. (2011). Patterns and Dynamics of Dissolved Organic Carbon (DOC) in Boreal Streams: The Role of Processes, Connectivity, and Scaling. *Ecosystems*, 14(6), 880–893. <https://doi.org/10.1007/s10021-011-9452-8>

- Laudon, H., & Buffam, I. (2008). Impact of changing DOC concentrations on the potential distribution of acid sensitive biota in a boreal stream network. *Hydrology and Earth System Sciences*, 12(2), 425–435. <https://doi.org/10.5194/hess-12-425-2008>
- Laudon, H., Buttle, J., Carey, S. K., McDonnell, J., McGuire, K., Seibert, J., Shanley, J., Soulsby, C., & Tetzlaff, D. (2012). Cross-regional prediction of long-term trajectory of stream water DOC response to climate change. *Geophysical Research Letters*, 39(17), 4–9. <https://doi.org/10.1029/2012GL053033>
- Laudon, H., Westling, O., Lofgren, S., & Bishop, K. (2001). Modeling preindustrial ANC and pH during the spring flood in northern Sweden. *Biogeochemistry*, 54(2), 171–195. <https://doi.org/10.1023/A:1010614631588>
- Lee, E. J., Yoo, G. Y., Jeong, Y., Kim, K. U., Park, J. H., & Oh, N. H. (2015). Comparison of UV-VIS and FDOM sensors for in situ monitoring of stream DOC concentrations. *Biogeosciences*, 12(10), 3109–3118. <https://doi.org/10.5194/bg-12-3109-2015>
- Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of ‘goodness-of-fit’ measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35(1), 233–241. <https://doi.org/10.1029/1998WR900018>
- Li, P., & Hur, J. (2017). Utilization of UV-Vis spectroscopy and related data analyses for dissolved organic matter (DOM) studies: A review. *Critical Reviews in Environmental Science and Technology*, 47(3), 131–154. <https://doi.org/10.1080/10643389.2017.1309186>

- Lindroth, A., Grelle, A., & Morén, A. S. (1998). Long-term measurements of boreal forest carbon balance reveal large temperature sensitivity. *Global Change Biology*, 4(4), 443–450. <https://doi.org/10.1046/j.1365-2486.1998.00165.x>
- Maloney, K. O., Morris, D. P., Moses, C. O., & Osburn, C. L. (2005). The role of iron and dissolved organic carbon in the absorption of ultraviolet radiation in humic lake water. *Biogeochemistry*, 75(3), 393–407. <https://doi.org/10.1007/s10533-005-1675-3>
- McGlynn, B. L., & McDonnell, J. J. (2003). Role of discrete landscape units in controlling catchment dissolved organic carbon dynamics. *Water Resources Research*, 39(4). <https://doi.org/10.1029/2002WR001525>
- Mehmood, T., Liland, K. H., Snipen, L., & Sæbø, S. (2012). A review of variable selection methods in Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems*, 118, 62–69. <https://doi.org/10.1016/j.chemolab.2012.07.010>
- Montgomery, D. C., & Runger, G. C. (2018). *Applied Statistics and Probability for Engineers* (Seventh ed). John Wiley & Sons.
- Mujica, L. E., Rodellar, J., Fernández, A., & Güemes, A. (2011). Q-statistic and t2-statistic pca-based measures for damage assessment in structures. *Structural Health Monitoring*, 10(5), 539–553. <https://doi.org/10.1177/1475921710388972>
- Muller, F. L. L., & Tankéré-Muller, S. P. C. (2012). Seasonal variations in surface water

- chemistry at disturbed and pristine peatland sites in the Flow Country of northern Scotland. *Science of the Total Environment*, 435–436, 351–362.
<https://doi.org/10.1016/j.scitotenv.2012.06.048>
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology*, 10(3), 282–290.
[https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Neff, J. C., & Asner, G. P. (2001). Dissolved organic carbon in terrestrial ecosystems: Synthesis and a model. *Ecosystems*, 4(1), 29–48.
<https://doi.org/10.1007/s100210000058>
- Olefeldt, D., Roulet, N., Giesler, R., & Persson, A. (2013). Total waterborne carbon export and DOC composition from ten nested subarctic peatland catchments-importance of peatland cover, groundwater influence, and inter-annual variability of precipitation patterns. *Hydrological Processes*, 27(16), 2280–2294.
<https://doi.org/10.1002/hyp.9358>
- Pace, M. L., Reche, I., Cole, J. J., Fernández-Barbero, A., Mazuecos, I. P., & Prairie, Y. T. (2012). pH change induces shifts in the size and light absorption of dissolved organic matter. *Biogeochemistry*, 108(1–3), 109–118. <https://doi.org/10.1007/s10533-011-9576-0>
- Peacock, M., Evans, C. D., Fenner, N., Freeman, C., Gough, R., Jones, T. G., & Lebron, I. (2014). UV-visible absorbance spectroscopy as a proxy for peatland dissolved organic

carbon (DOC) quantity and quality: Considerations on wavelength and absorbance degradation. *Environmental Sciences: Processes and Impacts*, 16(6), 1445–1461.

<https://doi.org/10.1039/c4em00108g>

Pereira, H. R., Meschiatti, M. C., Pires, R. C. de M., & Blain, G. C. (2018). On the performance of three indices of agreement: An easy-to-use r-code for calculating the willmott indices. *Bragantia*, 77(2), 394–403. <https://doi.org/10.1590/1678-4499.2017054>

Rantakari, M., Mattsson, T., Kortelainen, P., Piirainen, S., Finér, L., & Ahtiainen, M. (2010). Organic and inorganic carbon concentrations and fluxes from managed and unmanaged boreal first-order catchments. *Science of the Total Environment*, 408(7), 1649–1658. <https://doi.org/10.1016/j.scitotenv.2009.12.025>

Raymond, P. A., & Saiers, J. E. (2010). Event controlled DOC export from forested watersheds. *Biogeochemistry*, 100(1), 197–209. <https://doi.org/10.1007/s10533-010-9416-7>

Raymond, P. A., Saiers, J. E., & Sobczak, W. V. (2016). Hydrological and biogeochemical controls on watershed dissolved organic matter transport: Pulse- shunt concept. *Ecology*, 97(1), 5–16. <https://doi.org/10.1890/14-1684.1>

Rosecrans, C. Z., Nolan, B. T., & Gronberg, J. A. M. (2017). Prediction and visualization of redox conditions in the groundwater of Central Valley, California. *Journal of Hydrology*, 546, 341–356. <https://doi.org/10.1016/j.jhydrol.2017.01.014>

- Ruhala, S. S., & Zarnetske, J. P. (2017). Using in-situ optical sensors to study dissolved organic carbon dynamics of streams and watersheds: A review. *Science of the Total Environment*, 575, 713–723. <https://doi.org/10.1016/j.scitotenv.2016.09.113>
- Senar, O. E., Webster, K. L., & Creed, I. F. (2018). Catchment-Scale Shifts in the Magnitude and Partitioning of Carbon Export in Response to Changing Hydrologic Connectivity in a Northern Hardwood Forest. *Journal of Geophysical Research: Biogeosciences*, 123(8), 2337–2352. <https://doi.org/10.1029/2018JG004468>
- Shalizi, C. (2015). *Lecture 10 : F -Tests , R 2 , and Other Distractions*. October, 1–24.
- Shalizi, Cosma. (2013). Advanced data analysis from an elementary point of view. *Book Manuscript*, 801. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.371.4613&rep=rep1&type=pdf>
- Simonsson, M., Kaiser, K., Danielsson, R., Andreux, F., & Ranger, J. (2005). Estimating nitrate, dissolved organic carbon and DOC fractions in forest floor leachates using ultraviolet absorbance spectra and multivariate analysis. *Geoderma*, 124(1–2), 157–168. <https://doi.org/10.1016/j.geoderma.2004.04.010>
- Spencer, R. G. M., Bolton, L., & Baker, A. (2007). Freeze/thaw and pH effects on freshwater dissolved organic matter fluorescence and absorbance properties from a number of UK locations. *Water Research*, 41(13), 2941–2950. <https://doi.org/10.1016/j.watres.2007.04.012>

- Strohmeier, S., Knorr, K. H., Reichert, M., Frei, S., Fleckenstein, J. H., Peiffer, S., & Matzner, E. (2013). Concentrations and fluxes of dissolved organic carbon in runoff from a forested catchment: Insights from high frequency measurements. *Biogeosciences*, *10*(2), 905–916. <https://doi.org/10.5194/bg-10-905-2013>
- Tank, S. E., Fellman, J. B., Hood, E., & Kritzberg, E. S. (2018). Beyond respiration: Controls on lateral carbon fluxes across the terrestrial-aquatic interface. *Limnology and Oceanography Letters*, *3*(3), 76–88. <https://doi.org/10.1002/lol2.10065>
- Thurman, E. M. (1985). Organic Geochemistry of Natural Waters. In *MARTIN US NIJHOFF IDR W. JUNK PUBLISHERS*. Springer Netherlands. <https://doi.org/10.1007/978-94-009-5095-5>
- Tipping, E., Hilton, J., & James, B. (1988). Dissolved organic matter in Cumbrian lakes and streams. *Freshwater Biology*, *19*(3), 371–378. <https://doi.org/10.1111/j.1365-2427.1988.tb00358.x>
- Tiwari, T., Laudon, H., Beven, K., & Ågren, A. (2014). Downstream changes in DOC: Inferring contributions in the face of model uncertainties. *Water Resources Research*, *50*(1), 514–525. <https://doi.org/10.1002/2013WR014275>
- Vaughan, M. C. H., Bowden, W. B., Shanley, J. B., Vermilyea, A., & Schroth, A. W. (2019). Shining light on the storm: in-stream optics reveal hysteresis of dissolved organic matter character. *Biogeochemistry*, *143*(3), 275–291. <https://doi.org/10.1007/s10533-019-00561-w>

- Vaughan, M. C. H., Bowden, W. B., Shanley, J. B., Vermilyea, A., Sleeper, R., Gold, A. J., Pradhanang, S. M., Inamdar, S. P., Levia, D. F., Andres, A. S., Birgand, F., & Schroth, A. W. (2017). High-frequency dissolved organic carbon and nitrate measurements reveal differences in storm hysteresis and loading in relation to land cover and seasonality. *Water Resources Research*, 53(7), 5345–5363. <https://doi.org/10.1002/2017WR020491>
- Villacorte, L. O., Ekowati, Y., Neu, T. R., Kleijn, J. M., Winters, H., Amy, G., Schippers, J. C., & Kennedy, M. D. (2015). Characterisation of algal organic matter produced by bloom-forming marine and freshwater algae. *Water Research*, 73, 216–230. <https://doi.org/10.1016/j.watres.2015.01.028>
- Villegas, T., Fuente, M. J., & Rodríguez, M. (2010). Principal component analysis for fault detection and diagnosis. Experience with a pilot plant. *International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics - Proceedings*, 147–152.
- Wallin, M. B., Grabs, T., Buffam, I., Laudon, H., Ågren, A., Öquist, M. G., & Bishop, K. (2013). Evasion of CO₂ from streams - The dominant component of the carbon export through the aquatic conduit in a boreal landscape. *Global Change Biology*, 19(3), 785–797. <https://doi.org/10.1111/gcb.12083>
- Webb, J. R., Santos, I. R., Maher, D. T., & Finlay, K. (2019). The Importance of Aquatic Carbon Fluxes in Net Ecosystem Carbon Budgets: A Catchment-Scale Review.

Ecosystems, 22(3), 508–527. <https://doi.org/10.1007/s10021-018-0284-7>

Weishaar, J. L., Aiken, G. R., Bergamaschi, B. A., Fram, M. S., Fujii, R., & Mopper, K. (2003).

Evaluation of specific ultraviolet absorbance as an indicator of the chemical composition and reactivity of dissolved organic carbon. *Environmental Science and Technology*, 37(20), 4702–4708. <https://doi.org/10.1021/es030360x>

Werner, B. J., Musolff, A., Lechtenfeld, O. J., De Rooij, G. H., Oosterwoud, M. R., &

Fleckenstein, J. H. (2019). High-frequency measurements explain quantity and quality of dissolved organic carbon mobilization in a headwater catchment. *Biogeosciences*, 16(22), 4497–4516. <https://doi.org/10.5194/bg-16-4497-2019>

Willmott, C. J., Robeson, S. M., & Matsuura, K. (2012). A refined index of model performance. *International Journal of Climatology*, 32(13), 2088–2094. <https://doi.org/10.1002/joc.2419>

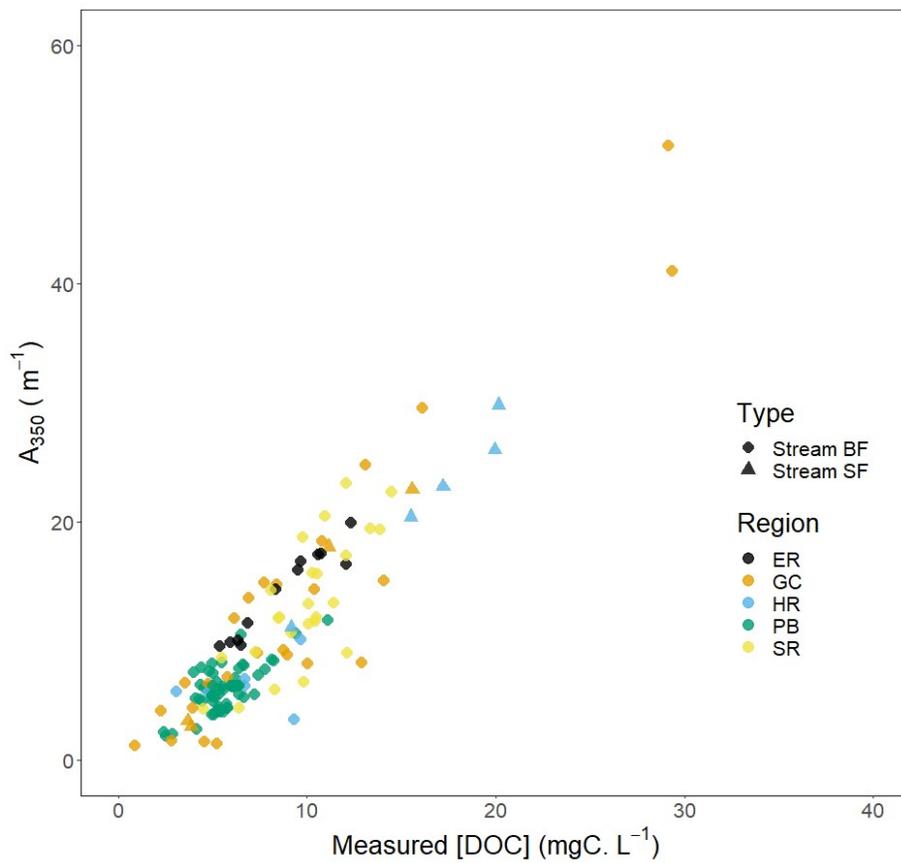
Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics.

Chemometrics and Intelligent Laboratory Systems, 58(2), 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1)

Wold, S., & Trygg, J. (2004). The PLS method -- partial least squares projections to latent structures -- and its applications in industrial RDP (research , development , and production). *PLS in Industrial RPD for Prague*, 1(June), 1–44. <https://doi.org/10.1109/JMEMS.2011.2159097>

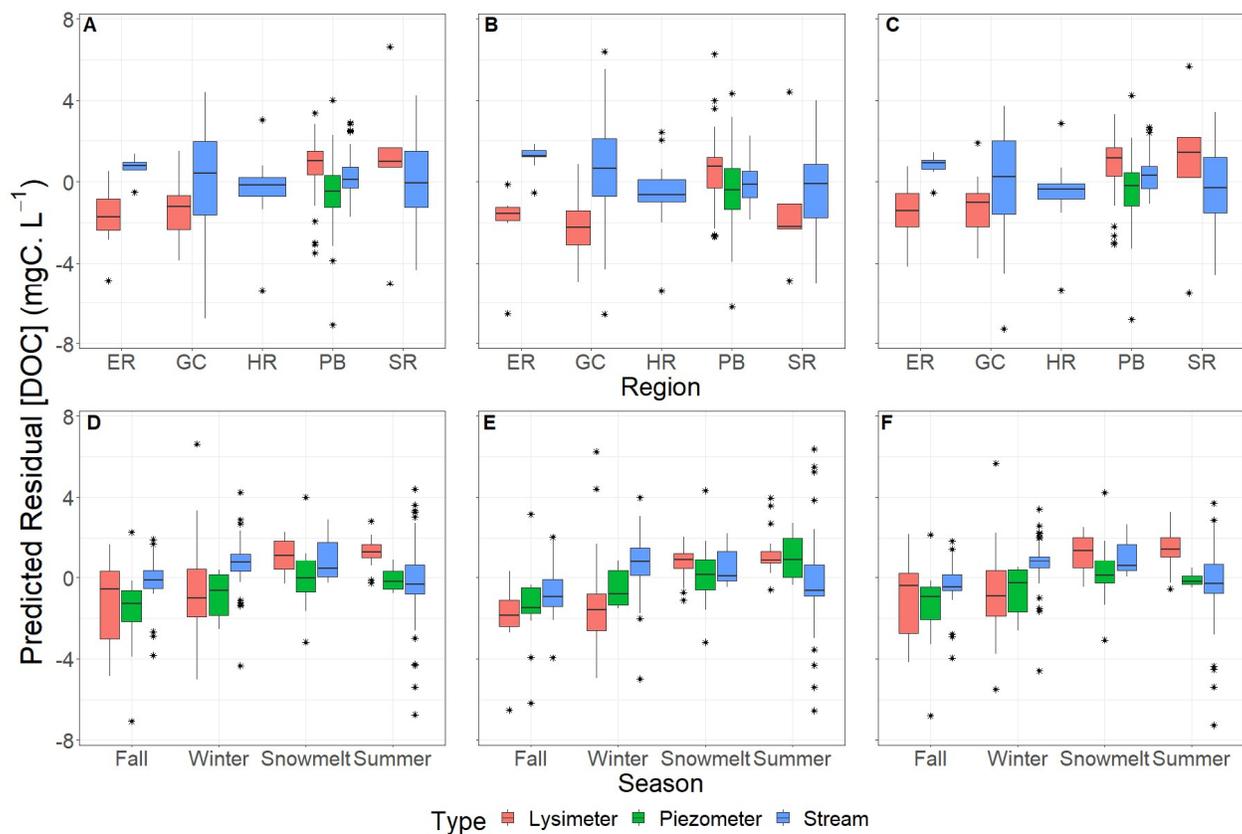
- Worrall, F., Burt, T., & Adamson, J. (2004). Can climate change explain increases in DOC flux from upland peat catchments? *Science of the Total Environment*, 326(1–3), 95–112. <https://doi.org/10.1016/j.scitotenv.2003.11.022>
- Wymore, A. S., Leon, M. C., Shanley, J. B., & McDowell, W. H. (2019). Hysteretic response of solutes and turbidity at the event scale across forested tropical montane watersheds. *Frontiers in Earth Science*, 7(May). <https://doi.org/10.3389/feart.2019.00126>
- Zhu, X., Chen, L., Pumpanen, J., Keinänen, M., Laudon, H., Ojala, A., Palviainen, M., Kiirikki, M., Neitola, K., & Berninger, F. (2020). Assessment of a portable UV–Vis spectrophotometer’s performance for stream water DOC and Fe content monitoring in remote areas. *Talanta*, 224(December). <https://doi.org/10.1016/j.talanta.2020.121919>
- Ziegler, S. E., Benner, R., Billings, S. A., Edwards, K. A., Philben, M., Zhu, X., & Laganière, J. (2017). Climate warming can accelerate carbon fluxes without changing soil carbon stocks. *Frontiers in Earth Science*, 5(February), 1–12. <https://doi.org/10.3389/feart.2017.00002>

Appendix



Appendix 1: Relationship between stream water DOC and A₃₅₀

The relationship between absorbance at 350 nm (A₃₅₀) and dissolved organic carbon concentration ([DOC]) across stream samples across NL-BELT and the PBEWA. NL-BELT regions include Eagle River (ER), Grand Codroy (GC), Humber River (HR) and Salmon River (SR).



Appendix 2: Model residuals by season, site and sample type

Model estimate residual distributions across NL-BELT and the PBEWA regions (top, A, B, C) and seasons (bottom, D, E, F) for the Avagyan (left, A, D), Fichot & Benner (middle, B, E) and PLS-plain (right, C, F) models across sample types. One-way ANOVAs testing each of the model residuals to regions or seasons found season to be significant in the Fichot & Benner residuals ($p < .001$). The asterisks refer to extreme outliers outside 1.5 times the interquartile range.

Appendix 3: Detailed NL-BELT site information

NL-BELT site locations. Passive pan lysimeters are located in terrestrial sites. The Pynn's Brook Experimental Watershed Area terrestrial site contains passive piezometers and groundwater wells in addition to passive pan lysimeters.

Region	Site Type	Site	Latitude	Longitude	Elevation (m)
Eagle River	Terrestrial	Muddy Pond	53°33'1.08"N	56°59'12.79"W	145
Eagle River	Terrestrial	Sheppard's Ridge	53°33'24.60"N	56°56'2.34"W	170
Eagle River	Terrestrial	Harry's Pond	53°35'12.48"N	56°53'20.69"W	136
Eagle River	Stream	Desperation Brook	53°31'57.48"N	57°4'1.99"W	46
Eagle River	Stream	Marcus Brook	53°35'28.52"N	56°53'5.72"W	13
Eagle River	Stream	Suckers Brook	53°33'31.75"N	56°55'8.38"W	39
Eagle River	Spring	Spring	53°35'4.90"N	56°53'1.39"W	107
Salmon River	Terrestrial	Hare Bay	51°15'21.28"N	56°8'17.76"W	31
Salmon River	Terrestrial	Tuckamore	51°9'50.68"N	56°0'15.14"W	16
Salmon River	Terrestrial	Catch-a-Feeder	51°5'21.27"N	56°12'15.58"W	38
Salmon River	Stream	Flat Brook	50°59'6.90"N	6°29'12.14"W	250
Salmon River	Stream	Wilcox Brook	51°0'35.64"N	6°26'31.47"W	158
Salmon River	Stream	Red Ale Brook	51°2'10.54"N	6°19'46.01"W	78
Salmon River	Spring	Trailer Park Spring	51°3'33.59"N	6°37'24.39"W	172
Salmon River	Spring	Beer Cooler Spring	51°2'1.79"N	6°36'49.91"W	211
Humber River	Terrestrial	Camp 10	49°4'11.08"N	7°38'34.29"W	171
Humber River	Terrestrial	Camp 9	49°3'6.77"N	7°37'28.72"W	45

Humber River	Terrestrial	Caribou Pass	49° 4'22.27"N	7°37'42.06"W	113
Humber River	Terrestrial	Pynn's Brook Experimental Watershed Area	49° 5'2.72"N	57°23'8.68"W	180
Humber River	Stream	Laughing Brook	49° 1'47.52"N	57°39'50.46" W	39
Humber River	Stream	Lower Horseshoe	49° 5'23.7"N	57°22'48.9"W	175
Humber River	Stream	Upper Horseshoe	49° 4' 9.11"N	57°20'58.15" W	245
Humber River	Stream	Kettle Creek	49° 4'39.96"N	7°23'55.21"W	156
Humber River	Stream	Pynn's Brook	49° 4'43.87"N	57°31'2.71"W	143
Grand Codroy	Terrestrial	O'Regan's	47°53'36.34" N	59°10'28.31" W	100
Grand Codroy	Terrestrial	Maple Ridge	48° 0'28.25"N	8°55'14.44"W	165
Grand Codroy	Terrestrial	Slug Hill	48° 0'38.61"N	8°54'16.34"W	215
Grand Codroy	Stream	Morris Brook	48° 4'24.69"N	8°54'57.21"W	265
Grand Codroy	Stream	North Branch	48° 0'48.12"N	8°54'42.57"W	166
Grand Codroy	Stream	O'Regan's Brook	47°53'23.37" N	59° 9'48.60"W	74