Reproducible data analysis: thesis data analysis and visualization

Pacific Maritime forWater Masters Project (NSERC forWater)

Hannah J McSorley

2020-01-31

# General Introduction / context

Surface water is the primary source of drinking water for over 85% of the Canadian population and in the province of British Columbia, approximately 80% of drinking water originates from forested headwaters (Pike et al. 2010). Forests offer a variety of ecosystem services (e.g. biodiversity) and also slow and filter runoff, resulting in high quality source water supply (Dudley and Stolton 2003). Surface water quality varies over time and space due to climate, weather, and physical characteristics of the watershed (such as topography, land cover and geology) with terrestrial sediments, nutrients, and organic material (e.g. leaf litter) entering surface water through erosion and precipitation events (Pike et al. 2010; Johnson et al. 1997; Delpla and Rodriguez 2016; Health Canada 2019; Yang et al. 2015).

In Canada, all source water is treated to meet Health Canada drinking water quality guidelines, which specify allowable levels of biological, physical and chemical parameters that are safe for human use and consumption (British Columbia Ministry of Environment 2017). Drinking water treatment processes vary from chlorination alone (at the simplest level) to multiple stages of chemically assisted filtration in combination with advanced oxidative processes. Water treatment protocols differ between communities based on source water quality, infrastructure capabilities, budget, and regional water quality regulations [REFS]. Chlorination remains the most widely used method of inactivating potentially harmful microorganisms, whether it stands alone as the primary and sole treatment protocol, or is used in combination with other processes [HealthCanada (2006); HealthLinkBC2018].

When natural source water is chlorinated, chemical reactions with natural organic matter (NOM) can form a variety of chlorinated organic compounds which are broadly classified as disinfection by-products (DBPs). (Richardson et al. 2007; Delpla and Rodriguez 2016; Health Canada 2019; Yang et al. 2015; Hua, Reckhow, and Abusallout 2015; Eaton, A. D., Clesceri, L. S., Greenberg, A. E., Franson 2000). A number of DBPs are included in Health Canada’s drinking water quality guidelines, and have maximum allowable concentrations in treated water due to their potential or known health affects (i.e. genotoxicity and carcinogenicity) (Richardson et al. 2007; British Columbia Ministry of Environment 2017).

The molecular composition and physical structure of NOM influence its reactivity, therefore different types of NOM have different DBP formation potentials (DBP-FP) (Delpla and Rodriguez 2016; Yang et al. 2015; Health Canada 2019). NOM is a diverse family of molecules, which contain nitrogen, silica, oxygen and hydrogen but are composed primarily of organic carbon, therefore organic carbon is often measured as a proxy for NOM concentration (Health Canada 2019). Total organic carbon (TOC) is composed of particulate and dissolved fractions (POC and DOC, respectively). Generally, DOC comprises a larger component of TOC relative to POC, and is more likely to generate DBPs during treatment [@ REFS]. Because NOM is a precursor for DBPs, source water quality guidelines in British Columbia specify that TOC should remain below 4mg/L in source water (British Columbia Ministry of Environment 2017).

# Sample Analyses

Water samples were collected and transported via coolers (on ice) to UBC’s EcoHydrology Lab for analysis of dissolved organic carbon (DOC) concentrations and indicators of NOM character. Samples were also measured for phosphate concentration using a field portable colourmetric test kit (HACH); each sample had phosphate concentrations below detectable limits (XXXXXX mg/L).

## Quantifying dissolved organic carbon (DOC)

For quantification of DOC, samples were analyzed for non-purgeable organic carbon (NPOC) via High-Temperature Combustion Method (5310-B) on a Shimadzu TOC-V (Eaton, A. D., Clesceri, L. S., Greenberg, A. E., Franson 2000). For this method, samples were filtered with 0.45-micron PES filters, acidified to bring pH below 2, then sparged with hydrocarbon-free air to drive off inorganic carbon. Following sparging, samples were combusted to convert all organic carbon to carbon dioxide which was measured with a non-dispersive infrared gas detector to quantify the DOC content of the sample. This method represents a direct method of measuring DOC, as all natural organic carbon in the sample is measured. Although small volatile organic compounds would be removed in the sparging process, most NOM compounds are of higher molecular weight and it is unlikely that NOM DOC analytes would be lost (Eaton, A. D., Clesceri, L. S., Greenberg, A. E., Franson 2000; Matilainen et al. 2011).

## Qualifying dissolved organic matter (proxy measures of DOC)

Spectral properties of sample NOM were analyzed using a “Spectro::lyser” spectrophotometer (S::can, Vienna, Austria) which measures turbidity and the chromophoric portion of organic matter to estimate concentrations of total organic carbon (TOC), dissolved organic carbon (DOC), as well as nitrate-nitrogen (NO3--N). A caveat to interpreting the Spectrolyser results is that, for NOM to be detected by UV-Vis absorption the molecules must absorb UV or Visible light (not all molecules do). UV-Vis absorption occurs only if the applied energy (light) can be absorbed by the molecule; in general, this required the presence of aromatic bonds (i.e. conjugated pi-bond systems in the molecule, a.k.a chromophore). Therefore, UV-Vis absorption is proportional to the molecule’s degree of aromaticity and is a measure of chromophoric DOM (CDOM).

The Spectrolyser outputs a fingerprint file containing absorbance values at all of the wavelengths monitored. These absorbance values can provide information about the character of NOM. For example, the ratio of the slope between 275-295nm and the slope from 350-400nm (“slope ratio (SR)”), is commonly used as an indicator of molecular weight. Based on full scan data from the Spectrolyser (250-700nm), several indices of NOM character were determined {*in progress*}:

* SUVA254 — Specific UV absorbance at 254nm is the ratio of UV absorption (spectral absorbance coefficient, SAC, m-1) at 254nm, normalized to DOC concentration (mgL-1). SUVA254 correlates strongly with DOM aromaticity (Helms et al. 2008).
* Spectral slopes — it has been shown that UV spectral slopes are useful semi quantitative indicators for assessing NOM molecular weights (Helms et al. 2008):
* *Spectral slopes 1 & 2 — the slope of spectral intensity over the shorter wavelength range of 275-295nm (1) and longer wavelength range of 350-400nm (2).* Slope Ratio — The ratio of spectral slopes 1 and 2, correlates well with CDOM molecular weight (Helms et al. 2008)

# Data Visualization & Summaries

## Leech DOC concentrations overview

Table 1 and Figure 1 show DOC concentrations were highest at Weeks Creek, the headwaters monitoring point for a sub-basin which included Weeks Lake and surrounding wetlands. On average, Chris Creek DOC concentration was 51% lower than Weeks creek. Below the confluence of Weeks and Chris Creek, the head of Leech River had a mean DOC concentration slightly less than the average of the two headwater tributaries. West Leech River had higher DOC than Cragg Creek by about 20%. Below the confluences of West Leech and Cragg Creek, the Leech Tunnel site had DOC concentrations that were 54% lower than the average at the Leech head, and closer to those of Cragg Creek.

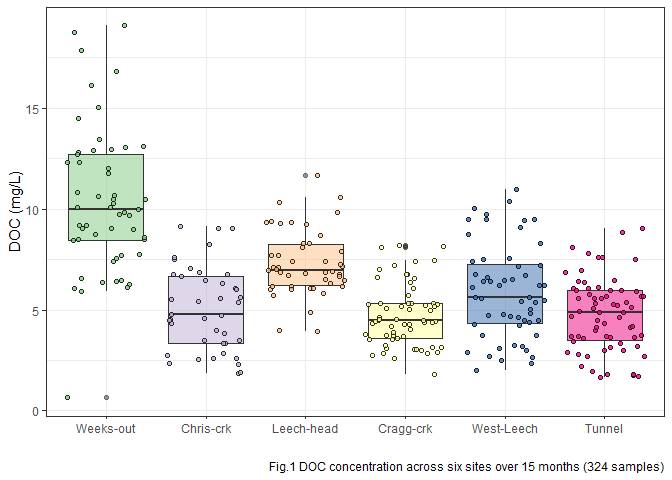
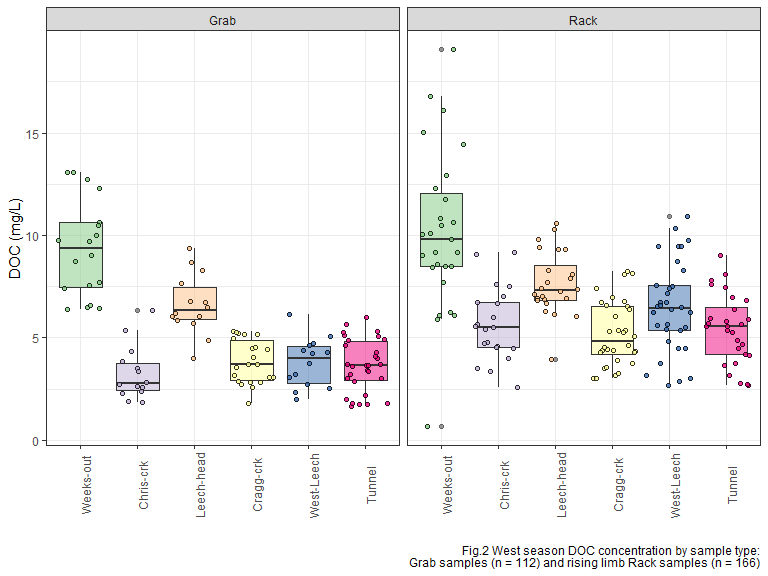


Table 1: summary of DOC concentrations at the six main sites

|  |  |  |  |
| --- | --- | --- | --- |
| site | mean DOC (mg/L) | std.dev. (± mg/L) | RSD (%) |
| Weeks-out | 10.4 | 3.6 | 34.4 |
| Chris-crk | 5.1 | 2.1 | 40.9 |
| Leech-head | 7.3 | 1.6 | 22.1 |
| Cragg-crk | 4.7 | 1.5 | 32.4 |
| West-Leech | 5.9 | 2.3 | 39.2 |
| Tunnel | 4.8 | 1.8 | 37.7 |

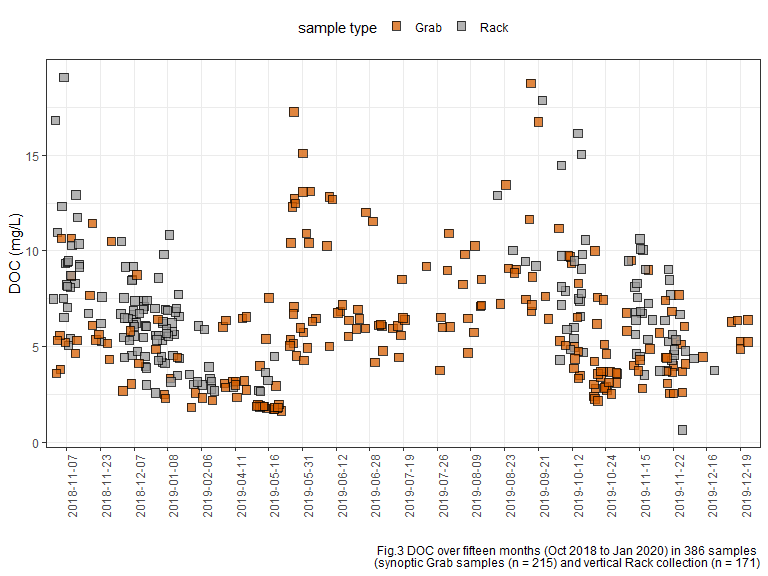
At the six sub-basin monitoring sites, vertical racks were installed to collect samples as the rivers rose in response to precipitation events.

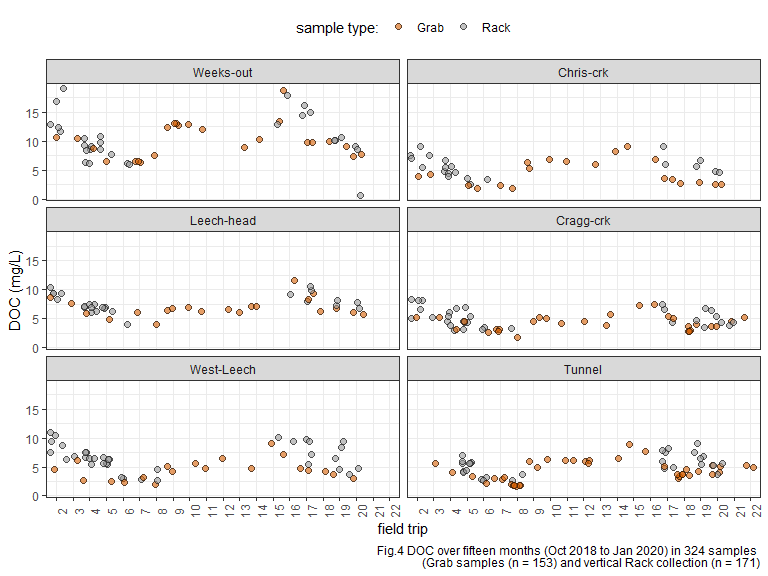
These vertical racks collected samples on the rising hydrograph limb, which has been shown to have higher DOC concentration that non-storm flow (Yang et al. 2015; Raymond et al. 2016, 2010). As the vertical racks require a hydrologic response to collect samples, they are only applicable during the wet season. Comparing wet season Grab samples to those collected on Racks confirms that the rack samplers were able to capture more samples, and many during periods with higher DOC concentrations (Fig.2).



## Sampling campaign

Four hundred water sample were collected and analyzed for DOC over a 16 month study period (October 2018 to February 2020). The synoptic grab-sampling program was enhanced by the collection of stormflow on the six vertical sampling racks. In total, 171 samples were collected on vertical Racks, 215 Grab samples were collected at 15 synoptically sampled sites (which included the six installation sites), and an additional 14 Grab samples were collected opportunistically at one-off sites across the water supply area. Figure 3 shows DOC concentrations across the 15 synoptically sampled sites over the study period (n=386).

 Over the study period, DOC concentrations decreased throughout the wet season, and increased again during the dry summer months (Fig.3). This seasonal DOC concentration pattern was clear at each of the six installation locations in the LWSA (Fig.4).



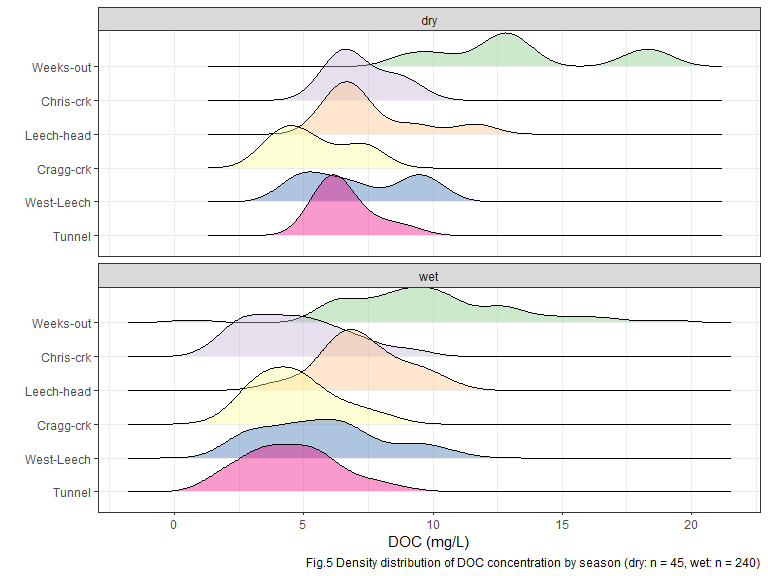
## Seasonal examination of DOC & DOM dynamics

In this section, a categorical season ID was used to assess temporal variations in DOC. Seasons were separated based on months of the year: the wet season was October through May, and the dry season was June until October.

* Precipitation data will be employed to operationally separate season (data obtained January 31, 2020 from CRD).

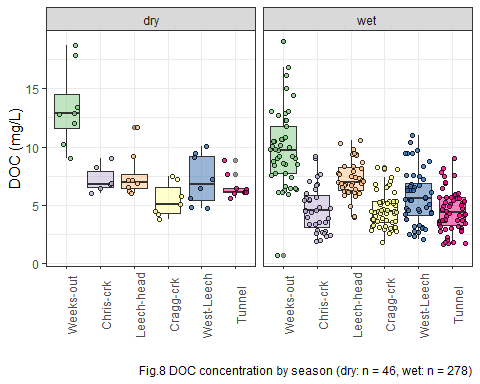
These seasonal plots show data from both Rack and Grab samples in order to include the full available range of concentrations.

In the wet season, DOC concentrations generally increased during stormflow (relative to between-storm peaks) *{data not shown here}*, and progressively dropped over the course of the wet season. During dry season baseflow, DOC concentration was elevated compared to late wet season DOC (Fig.5).



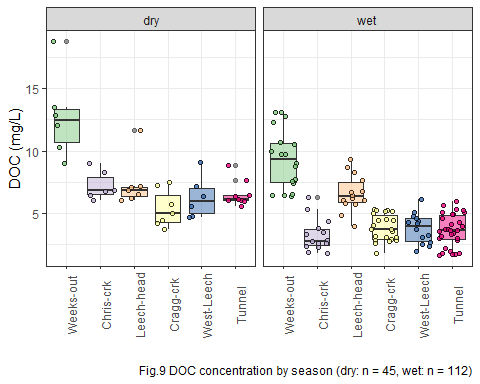
At each of the six sub-basin monitoring locations, mean DOC was higher during the dry season than during the wet season (Table 2, Fig.8). There was less variation in dry season DOC concentrations at each site also (Table 2), which could be due to actual reduction in variance in the absence of stormflow pulses and also due to fewer dry season samples (n = 46) compared to wet season samples (n = 278).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| season | site | mean DOC (mg/L) | std.dev. (± mg/L) | RSD (%) |
| Weeks-out | dry | 13.4 | 3.4 | 25.3 |
| Weeks-out | wet | 9.8 | 3.4 | 34.2 |
| Chris-crk | dry | 7.2 | 1.2 | 16.0 |
| Chris-crk | wet | 4.7 | 2.0 | 42.0 |
| Leech-head | dry | 7.6 | 1.9 | 25.3 |
| Leech-head | wet | 7.3 | 1.6 | 21.6 |
| Cragg-crk | dry | 5.4 | 1.5 | 27.2 |
| Cragg-crk | wet | 4.7 | 1.5 | 33.0 |
| West-Leech | dry | 7.1 | 2.1 | 29.9 |
| West-Leech | wet | 5.7 | 2.3 | 40.4 |
| Tunnel | dry | 6.6 | 1.0 | 15.6 |
| Tunnel | wet | 4.5 | 1.7 | 38.7 |



To remove the variance of stormflow DOC concentrations, Table 2-B and Figure 9 shows seasonal concentrations based only on Grab samples (excludes 166 stormflow samples collected by Racks).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| season | site | mean DOC (mg/L) | std.dev. (± mg/L) | RSD (%) |
| Weeks-out | dry | 12.7 | 3.4 | 26.6 |
| Weeks-out | wet | 9.3 | 2.4 | 25.4 |
| Chris-crk | dry | 7.2 | 1.2 | 16.0 |
| Chris-crk | wet | 3.3 | 1.3 | 39.9 |
| Leech-head | dry | 7.3 | 1.9 | 26.5 |
| Leech-head | wet | 6.6 | 1.5 | 22.2 |
| Cragg-crk | dry | 5.4 | 1.5 | 27.2 |
| Cragg-crk | wet | 3.8 | 1.1 | 27.8 |
| West-Leech | dry | 6.3 | 1.7 | 26.6 |
| West-Leech | wet | 3.8 | 1.2 | 31.5 |
| Tunnel | dry | 6.6 | 1.0 | 15.6 |
| Tunnel | wet | 3.6 | 1.3 | 35.6 |



### Plots of DOC & proxy

DOC was quantified as non-purgeable organic carbon (NPOC) on the Shimadzu TOC auto analyzer, and samples were also measured on a Spectrolyser spectrophotometer. A spectrophotometer can only measure the fraction of organic matter that is able to absorb UV or Visible light, and therefore DOC measured on the Spectrolyser was a measure of CDOM and an indicator of aromaticity (molecular character).

Figure 6 shows results from all samples comparing DOC concentrations (as NPOC, direct measure) to CDOM concentrations (UV-Vis, proxy measure of DOC). There was a strong seasonal separation of the relationship between these two measurements. These results indicate a seasonally-driven physiochemical difference in the aquatic NOM in this drinking water supply area.

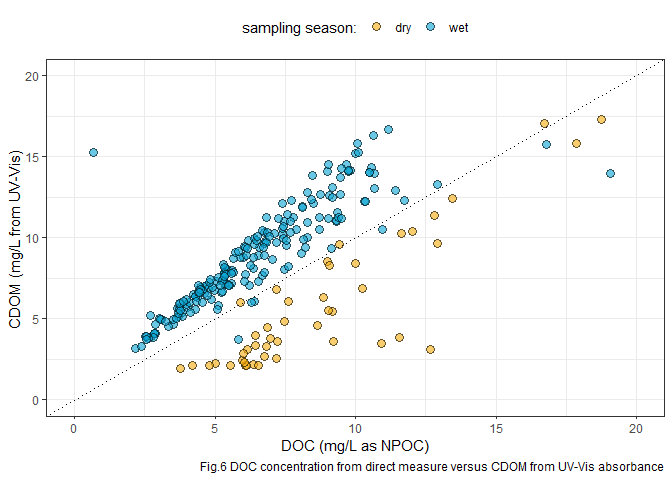
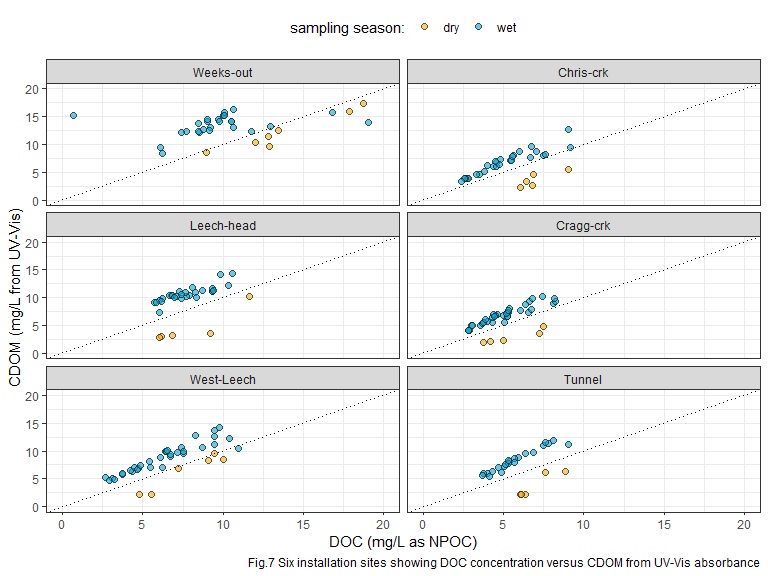


Figure 7 shows the same DOC-CDOM comparison from Figure 6, isolated to the six sub-basin monitoring sites. The seasonal relationship appears similar for the six sites.



In progress: analysis of spectral fingerprint data for slope ratio indices, which will help to elucidate seasonal DOM character shifts.

# References

British Columbia Ministry of Environment. 2017. “Source Drinking Water Quality Guidelines: Guideline Summary.” Victoria, B.C.: Prov. B.C. <https://www2.gov.bc.ca/gov/content/governments/organizational-structure/ministries-organizations/ministries/environment-climate-change>.

Delpla, Ianis, and Manuel J. Rodriguez. 2016. “Experimental disinfection by-product formation potential following rainfall events.” *Water Research* 104: 340–48. <https://doi.org/10.1016/j.watres.2016.08.031>.

Dudley, N, and S Stolton. 2003. “Running Pure: The importance of forest protected areas to drinking water.” World Bank / WWF Alliance for Forest Conservation; Sustainable Use. <http://scholar.google.com/scholar?hl=en{\&}btnG=Search{\&}q=intitle:Running+Pure{\#}1>.

Eaton, A. D., Clesceri, L. S., Greenberg, A. E., Franson, M. A. H. 2000. “Total Organic Carbon (TOC) - 5310 A.” In *Standard Methods for the Examination of Water and Wastewater*, 6th ed., 19–20. Washington, DC: American Public Health Association. <http://www.standardmethods.org/>.

Health Canada. 2019. “Guidance on Natural Organic Matter in Drinking Water.” <https://www.canada.ca/content/dam/hc-sc/documents/programs/consultation-organic-matter-drinking-water/NOM20190129-eng.pdf>.

HealthCanada. 2006. “Drinking Water Chlorination.” <https://www.canada.ca/en/health-canada/services/healthy-living/your-health/environment/drinking-water-chlorination.html>.

Helms, John R, Avon Stubbins, Jason D Ritchie, Elizabeth C Minor, and Kenneth Mopper. 2008. “Absorption Spectral Slopes and Slope Ratios as Indicators of Molecular Weight , Source , and Photobleaching of Chromophoric Dissolved Organic Matter Author ( s ): John R . Helms , Aron Stubbins , Jason D . Ritchie , Elizabeth C . Minor , David J . Kieber.” *Limnology and Oceanography* 53 (3): 955–69. <https://www.jstor.org/stable/40058211>.

Hua, Guanghui, David A. Reckhow, and Ibrahim Abusallout. 2015. “Correlation between SUVA and DBP formation during chlorination and chloramination of NOM fractions from different sources.” *Chemosphere* 130: 82–89. <https://doi.org/10.1016/j.chemosphere.2015.03.039>.

Johnson, Lucinda, Carl Richards, George Host, and John Arthur. 1997. “Landscape influences on water chemistry in Midwestern stream ecosystems.” *Freshwater Biology* 37: 193–208. <https://doi.org/doi:10.1046/j.1365-2427.1997.d01-539.x>.

Matilainen, Anu, Egil T. Gjessing, Tanja Lahtinen, Leif Hed, Amit Bhatnagar, and Mika Sillanpää. 2011. “An overview of the methods used in the characterisation of natural organic matter (NOM) in relation to drinking water treatment.” *Chemosphere* 83 (11): 1431–42. <https://doi.org/10.1016/j.chemosphere.2011.01.018>.

Pike, R., M. Feller, J. Stednick, K Rieberger, and M Carver. 2010. “Water Quality and Forest Management.” In *Compendium of Forest Hydrology and Geomorphology in British Columbia: Volume 2 of 2*, 400–439. <https://www.for.gov.bc.ca/hfd/pubs/docs/lmh/Lmh66/LMH66{\_}volume2of2.pdf>.

Raymond, Peter A, James E Saiers, Source Biogeochemistry, No September, Peter A Raymond, and James E Saiers. 2010. “Event controlled DOC export from forested watersheds.” *Biogeochemistry* 100 (1): 197–209. <https://doi.org/10.1007/sl0533-010-9416-7>.

Raymond, Peter A, James E Saiers, William V Sobczak, and E James. 2016. “Hydrological and biogeochemical controls on watershed dissolved organic matter transport: pulse-shunt concept.” *Ecology* 97 (1): 5–16. <https://www.jstor.org/stable/24702986>.

Richardson, Susan D., Michael J. Plewa, Elizabeth D. Wagner, Rita Schoeny, and David M. DeMarini. 2007. “Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: A review and roadmap for research.” *Mutation Research - Reviews in Mutation Research* 636 (1-3): 178–242. <https://doi.org/10.1016/j.mrrev.2007.09.001>.

Yang, Liyang, Jin Hur, Sonmin Lee, Soon Woong Chang, and Hyun Sang Shin. 2015. “Dynamics of dissolved organic matter during four storm events in two forest streams: source, export, and implications for harmful disinfection byproduct formation.” *Environmental Science and Pollution Research* 22 (12): 9173–83. <https://doi.org/10.1007/s11356-015-4078-6>.