Notebook 4: Modeling Stream Biogeochemistry

Dina Kaskina

July 12, 2022

# Fundamentals of Hydrology: Modeling stream biogeochemistry

## Introduction

### Stream Biogeochemistry

**Question**: Explain how the water and carbon cycle are interconnected in streams and rivers?

Generally, the water cycle is a physical phenomena and it is evaporated from the atmosphere or from the surface of plants by transpiration while the carbon cycle is mediated by photosynthesis and respiration by living organisms. The movement of water drives the carbon cycle in a way that streams and rivers provide an important intersection between terrestrial and marine biospheres.Thus, both cycles interact directly where carbon is transported, dissolved or suspended in running water which makes as a driving force moving biogeochemical constituents including carbon across streams and rivers.

**Question**: Describe the main forms in which carbon moves in streams and rivers.

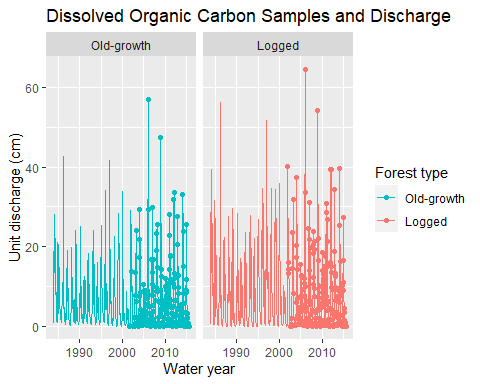
The main forms in which carbon moves in streams and rivers are CO2 (carbon dioxide), DOC (dissolved organic carbon), DIC (dissolved inorganic carbon), POC (particulate organic carbon) and PIC (particulate inorganic carbon). Different kinds of carbon (C) are transported across streams and rivers. The leaves, roots of trees absorb carbon dioxide (CO2) from the air and create particulate organic carbon (POC). After entering the stream, dissolved inorganic carbon (DIC) may either escape into the atmosphere (CO2 evaded) or remain as dissolved inorganic carbon and be transported downstream. Particulate organic carbon may break down to become dissolved organic carbon (DOC) or microbial communities may transform it into dissolved inorganic carbon (DIC) (Corson-Rikert et.al., 2016).

Today, we are going to explore bio-geochemical data. These data span several decades and are part of a Long Term Ecological Research Forest in Oregon. The [H.J.Andrews LTER](https://andrewsforest.oregonstate.edu/) is located in the Oregon Cascades. But before looking at these data, we need to load some packages that we installed in the previous session

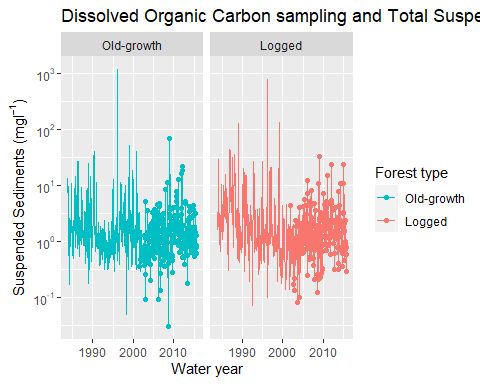
We will use the same data set analyzed in the stream flow section.

Our target variable for this analysis will be Dissolved Organic Carbon (DOC), let’s take a look at the sampling conditions as related to discharge, suspended sediments, and dissolved silica (non-reactive tracer).

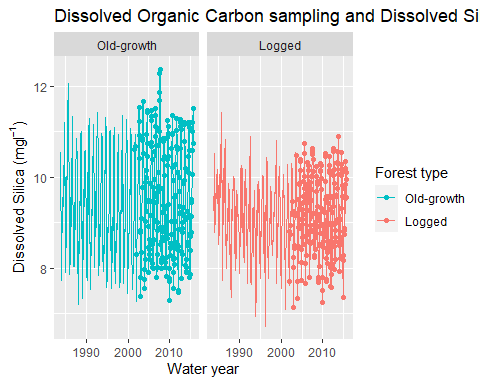
#### Dissolved Organic Carbon Samples and Discharge



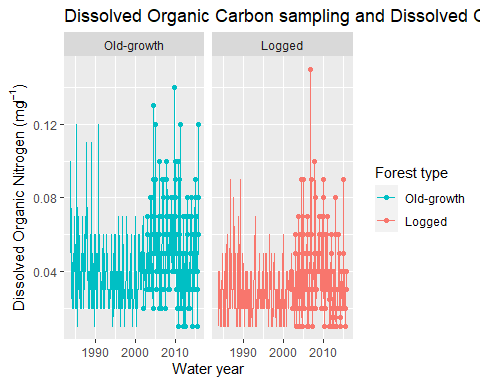
#### Dissolved Organic Carbon sampling and Total Suspended Solids



#### Dissolved Organic Carbon sampling and Dissolved Silica



#### Dissolved Organic Carbon sampling and Dissolved Organic Nitrogen



**To do:** Add captions to the figures above and describe how representative are the DOC samples in terms of the variation of discharge, sediment transport, dissolved silica, and dissolved organic nitrogen.

As can be seen from “Dissolved Organic Carbon Samples and Discharge” graph, compared to old-growth in the logged area the higher DOC values are associated with higher discharge.(The points on both graphs represent how DOC values match with predicted values). As for “Dissolved Organic Carbon sampling and Total Suspended Solids” graph, it is clear that by 1990 the DOC values associated with suspended sediments were higher in logged area rather than in old-growth one while in 1995 DOC was higher in old-growth. Further, “Dissolved Organic Carbon sampling and Dissolved Silica” graph indicates that DOC values associated with Silica concentrations were much higher in old-growth (in the range of 11-12.6 mg/L) compared to logged area (in the range of 10-11.3 mg/L). With regard to “Dissolved Organic Carbon sampling and Dissolved Organic Nitrogen” graph, DOC samples associated with Dissolved Organic Nitrogen clearly show an increase in concentrations in old-growth rather than in logged area.

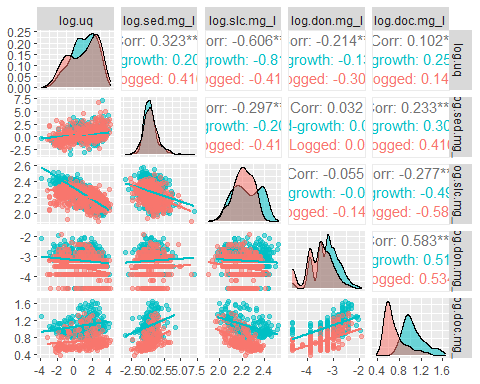
### Predicting Dissolved Organic Carbon in Headwater streams

Our selection of discharge, sediment transport, dissolved silica, and dissolved organic nitrogen in the plots above is not random. All of this variables affect the concentrations of DOC directly or indirectly.

**To do**: Look for examples of relationships between Q, TSS, DSi, and DON with DOC in headwater streams in the literature. Write a short paragraph (no more than 4 sentences about each of the relationships-please include the bibliographic reference)

The study by Jantze et.al.(2018) demonstrates the importance of studying lateral carbon transport in combination with hydrological flow paths which included measurements of stream discharge and concentrations of both dissolved organic carbon (DOC), dissolved inorganic carbon (DIC), and carbon dioxide (CO2) for 32 subcatchments of the Abiskojokka catchment in northern Sweden.Researchers of this study found that DOC, DIC and CO2 concentrations showed significant variability (p < 0.05) relative to discharge and other parameters. In another study by Townsend-Small et.al.(2011), the relationship between not only seasonal water discharge and dissolved organic carbon but also of nitrogen (DOC and DON) and nutrient concentrations in the upper Kuparuk River, Arctic Alaska was discovered. In particular, DOC and DON concentrations were positively correlated with discharge while nitrate (NO3 −) and silicate were negatively correlated with discharge throughout the study. Yet, discharge-specific DOC and DON concentrations decreased over the summer whereas discharge-specific concentrations of NO3 − and silicate increased. Both of the studies clearly show the relationship between the studied parameters and provide valuable information on watershed biogeochemistry.

Let’s now explore these relationships in our data set:



**To do**: Add a caption to the figure and compare the relationships observed between DOC and other variables across forest types. Also, how do these relationships compared with what has been reported in the literature?

In the above listed graph,the relationships between DOC and other parameters across the two forest types demonstrates the correrlation in terms of significance levels. For example, the relationship between sedimentation and silica is negative linear (-0.297) whereas the relationship between dissolved organic nitrogen and dissolved organic carbon is moderately correlated and it is stronger (0.583). Furthermore, the relationship between discharge and silica shows the strongest correlation (-0.606\*\*\*).All considered results are statistically significant.

The correlation coefficients calculated above, are not independent from each other. That is, the effect of the other variables has not been accounted for before estimating values. If we wanted to predict DOC concentrations based on this relationships, our starting point would be to build simple linear regression models.

**Question**: What does it mean “regression” in the context of statistical analyses?

Regression is a statistical method that helps us to examine and comprehend the relationship between two or more important variables. Regression analysis is carried out using a process that makes it easier to determine which aspects are crucial, which may be disregarded, and how they interact. In regression, usually there is one dependent variable and one or more independent variables (source: www.statisticshowto.com).

A simple linear regression model is expressed as

y = + \*x +

Where: y = response variable (the one we are interested in predicting, in our case DOC) x = predictor (the variable we are using to predict the response, e.g., Q, or TSS, or DSi) = Intercept (the value of our response variable when the predictor = 0) = Slope (the rate of change of our response variable per unit change in our predictor)

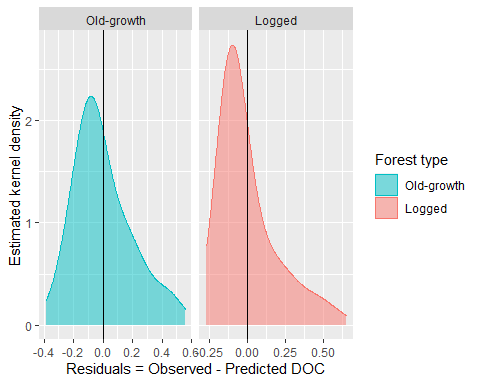
Let’s start with a simple linear regression of DOC against Q:

uq\_model = lm(log.doc.mg\_l~log.uq + ws.f, hjm)  
summary(uq\_model)

##   
## Call:  
## lm(formula = log.doc.mg\_l ~ log.uq + ws.f, data = hjm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.38356 -0.12763 -0.06197 0.08904 0.65124   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.045117 0.013644 76.601 < 2e-16 \*\*\*  
## log.uq 0.022970 0.005386 4.265 2.45e-05 \*\*\*  
## ws.fLogged -0.365056 0.018464 -19.772 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1939 on 440 degrees of freedom  
## (691 пропущенное наблюдение удалено)  
## Multiple R-squared: 0.476, Adjusted R-squared: 0.4736   
## F-statistic: 199.9 on 2 and 440 DF, p-value: < 2.2e-16

Let’s take a look at the residuals:

res\_plot <- ggplot(uq\_model, aes(x= uq\_model$residuals, color = ws.f, fill = ws.f))+  
 geom\_density(bw=0.085, alpha = 0.5)+  
 geom\_vline(xintercept = 0)+  
 facet\_wrap(~ws.f, scales = "free\_x")+  
 xlab("Residuals = Observed - Predicted DOC")+  
 ylab("Estimated kernel density")+  
 scale\_color\_manual(values = c("#00BFC4","#F8766D"), name = "Forest type")+  
 scale\_fill\_manual(values = c("#00BFC4","#F8766D"), name = "Forest type")  
res\_plot



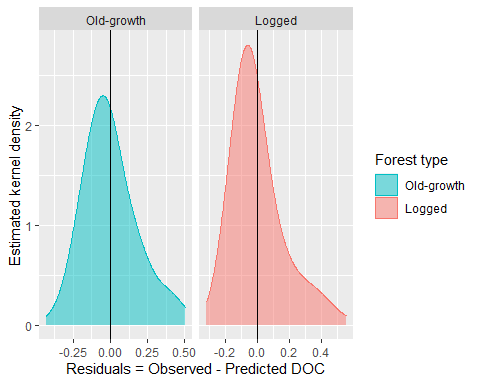
Let’s look now at the regression of DOC against Dsi

dsi\_model = lm(log.doc.mg\_l~log.slc.mg\_l + ws.f, hjm)  
summary(dsi\_model)

##   
## Call:  
## lm(formula = log.doc.mg\_l ~ log.slc.mg\_l + ws.f, data = hjm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.43099 -0.10957 -0.03391 0.07006 0.55418   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.12008 0.15847 19.69 <2e-16 \*\*\*  
## log.slc.mg\_l -0.91543 0.07035 -13.01 <2e-16 \*\*\*  
## ws.fLogged -0.39335 0.01618 -24.31 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1681 on 440 degrees of freedom  
## (691 пропущенное наблюдение удалено)  
## Multiple R-squared: 0.606, Adjusted R-squared: 0.6042   
## F-statistic: 338.4 on 2 and 440 DF, p-value: < 2.2e-16

Let’s take a look at the residuals again:

dsi\_res\_plot <- ggplot(uq\_model, aes(x=dsi\_model$residuals, color = ws.f, fill = ws.f))+  
 geom\_density(bw=.085,alpha = 0.5)+  
 geom\_vline(xintercept = 0)+  
 facet\_wrap(~ws.f, scales = "free\_x")+  
 xlab("Residuals = Observed - Predicted DOC")+  
 ylab("Estimated kernel density")+  
 scale\_color\_manual(values = c("#00BFC4","#F8766D"), name = "Forest type")+  
 scale\_fill\_manual(values = c("#00BFC4","#F8766D"), name = "Forest type")  
dsi\_res\_plot

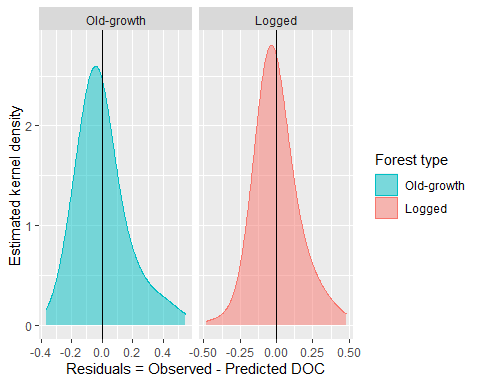
 Let’s build now a regression model including both Q and DSi

m\_model\_0 <- lm(log.doc.mg\_l~log.uq+log.slc.mg\_l+ws.f, hjm)  
summary(m\_model\_0)

##   
## Call:  
## lm(formula = log.doc.mg\_l ~ log.uq + log.slc.mg\_l + ws.f, data = hjm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.47955 -0.09611 -0.02166 0.05958 0.54635   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.835210 0.236749 20.423 <2e-16 \*\*\*  
## log.uq -0.062986 0.006864 -9.177 <2e-16 \*\*\*  
## log.slc.mg\_l -1.656859 0.103387 -16.026 <2e-16 \*\*\*  
## ws.fLogged -0.406222 0.014905 -27.254 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1542 on 439 degrees of freedom  
## (691 пропущенное наблюдение удалено)  
## Multiple R-squared: 0.6694, Adjusted R-squared: 0.6672   
## F-statistic: 296.3 on 3 and 439 DF, p-value: < 2.2e-16

Let’s look at the residuals:

m0\_res\_plot <- ggplot(m\_model\_0, aes(x =m\_model\_0$residuals, color = ws.f, fill = ws.f))+  
 geom\_density(bw=0.085, alpha = 0.5)+  
 geom\_vline(xintercept = 0)+  
 facet\_wrap(~ws.f, scales = "free\_x")+  
 xlab("Residuals = Observed - Predicted DOC")+  
 ylab("Estimated kernel density")+  
 scale\_color\_manual(values = c("#00BFC4","#F8766D"), name = "Forest type")+  
 scale\_fill\_manual(values = c("#00BFC4","#F8766D"), name = "Forest type")  
m0\_res\_plot



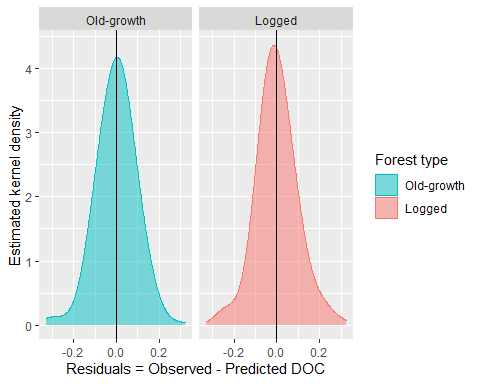
Let’s add all our predictor variables:

doc\_mod <- lm(log.doc.mg\_l ~ log.uq + q.cv + log.ptm + pt.cv + log.sed.mg\_l + (log.slc.mg\_l + log.don.mg\_l) \* ssn + ws.f + log.utn.mg\_l, data = dplyr::filter(hjm,is.na(hjm$doc.mg\_l)==FALSE))  
summary(doc\_mod)

##   
## Call:  
## lm(formula = log.doc.mg\_l ~ log.uq + q.cv + log.ptm + pt.cv +   
## log.sed.mg\_l + (log.slc.mg\_l + log.don.mg\_l) \* ssn + ws.f +   
## log.utn.mg\_l, data = dplyr::filter(hjm, is.na(hjm$doc.mg\_l) ==   
## FALSE))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.32683 -0.05513 0.00064 0.05008 0.32383   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.102e+00 2.329e-01 17.611 < 2e-16 \*\*\*  
## log.uq -3.590e-02 7.449e-03 -4.819 2.04e-06 \*\*\*  
## q.cv 9.962e-04 1.211e-04 8.229 2.58e-15 \*\*\*  
## log.ptm 1.520e-02 8.067e-03 1.885 0.06018 .   
## pt.cv -4.293e-04 7.872e-05 -5.454 8.55e-08 \*\*\*  
## log.sed.mg\_l 7.606e-03 6.142e-03 1.238 0.21630   
## log.slc.mg\_l -1.038e+00 1.014e-01 -10.233 < 2e-16 \*\*\*  
## log.don.mg\_l 2.209e-01 2.251e-02 9.811 < 2e-16 \*\*\*  
## ssnWinter -1.944e-01 3.534e-01 -0.550 0.58255   
## ssnSpring -1.699e+00 3.435e-01 -4.948 1.10e-06 \*\*\*  
## ssnSummer -2.513e+00 4.359e-01 -5.764 1.63e-08 \*\*\*  
## ws.fLogged -3.425e-01 1.048e-02 -32.665 < 2e-16 \*\*\*  
## log.utn.mg\_l 4.362e-01 1.674e-01 2.606 0.00949 \*\*   
## log.slc.mg\_l:ssnWinter -2.628e-01 1.703e-01 -1.543 0.12368   
## log.slc.mg\_l:ssnSpring 4.429e-01 1.543e-01 2.871 0.00430 \*\*   
## log.slc.mg\_l:ssnSummer 7.658e-01 1.770e-01 4.326 1.91e-05 \*\*\*  
## log.don.mg\_l:ssnWinter -2.013e-01 2.782e-02 -7.236 2.32e-12 \*\*\*  
## log.don.mg\_l:ssnSpring -1.936e-01 2.810e-02 -6.888 2.16e-11 \*\*\*  
## log.don.mg\_l:ssnSummer -2.255e-01 3.249e-02 -6.939 1.56e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.09196 on 407 degrees of freedom  
## (17 пропущенных наблюдений удалены)  
## Multiple R-squared: 0.8863, Adjusted R-squared: 0.8813   
## F-statistic: 176.2 on 18 and 407 DF, p-value: < 2.2e-16

And explore the residuals:

doc\_res\_plot <- ggplot(doc\_mod, aes(x = doc\_mod$residuals, color = ws.f, fill = ws.f))+  
 geom\_density(bw = 0.05, alpha = 0.5)+  
 geom\_vline(xintercept = 0)+  
 xlab("Residuals = Observed - Predicted DOC")+  
 ylab("Estimated kernel density")+  
 scale\_color\_manual(values = c("#00BFC4","#F8766D"), name = "Forest type")+  
 scale\_fill\_manual(values = c("#00BFC4","#F8766D"), name = "Forest type")+  
 facet\_wrap(~ws.f)  
doc\_res\_plot



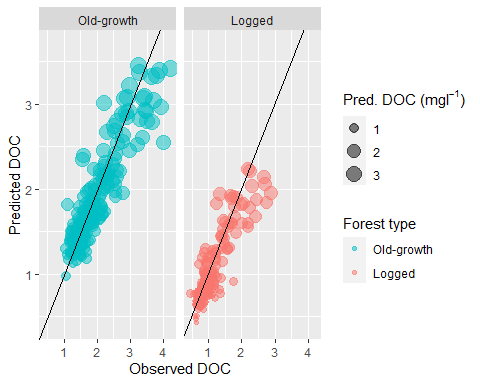
Let’s now predict DOC across our entire data set:

hjm$p.doc.mg\_l <- exp(predict.lm(doc\_mod,newdata = hjm,na.action = na.pass))  
cm <- lm(doc.mg\_l~p.doc.mg\_l,hjm)  
hjm$p.doc.mg\_l <- (cm$coefficients[1] + hjm$p.doc.mg\_l)/cm$coefficients[2]  
hjm$log.p.doc.mg\_l <- log10(hjm$p.doc.mg\_l)

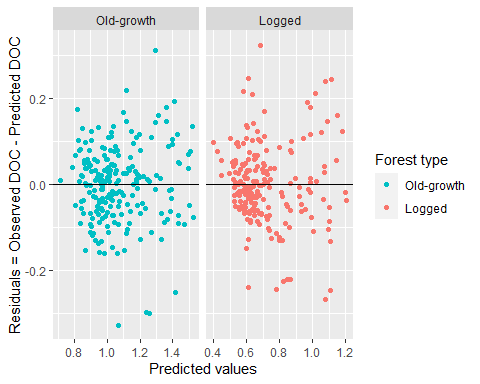
And let’s take a look at the observed vs. predicted relationship

obs\_prd\_doc <- ggplot(hjm, aes(x=doc.mg\_l,y=p.doc.mg\_l,color = ws.f, size =p.doc.mg\_l))+  
 geom\_point(alpha = 0.5)+  
 geom\_abline()+  
 scale\_color\_manual(values = c("#00BFC4","#F8766D"), name = "Forest type")+  
 scale\_size(name= expression(paste("Pred. DOC (mg",l^{-1}, ")")))+  
 xlab("Observed DOC")+  
 ylab("Predicted DOC")+  
 facet\_wrap(~ws.f)  
obs\_prd\_doc

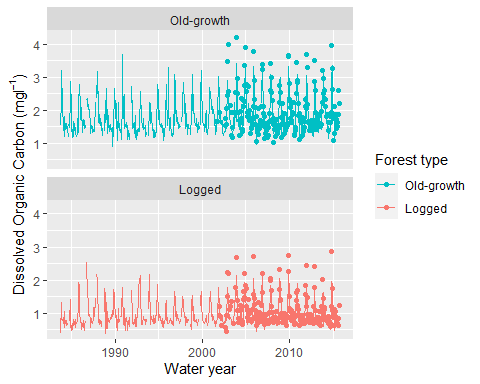
## Warning: Removed 708 rows containing missing values (geom\_point).



doc\_res\_all <- ggplot(doc\_mod, aes(x = doc\_mod$fitted.values, y = doc\_mod$residuals, color = ws.f))+  
 # geom\_density(alpha = 0.5)+  
 geom\_point()+  
 geom\_hline(yintercept = 0)+  
 scale\_color\_manual(values= c("#00BFC4","#F8766D"), name = "Forest type")+  
 xlab("Predicted values")+  
 ylab("Residuals = Observed DOC - Predicted DOC")+  
 facet\_wrap(~ws.f, scales="free\_x")  
doc\_res\_all

 Let’s take a look at the entire time series for DOC and the observed values

pdoc\_ts <- ggplot(hjm,aes(x=as.Date(dt),y=p.doc.mg\_l, color = ws.f, fill = ws.f))+  
 geom\_line()+  
 geom\_point(data=dplyr::filter(hjs,is.na(hjs$doc.mg\_l)==FALSE),aes(x=as.Date(dt),y=doc.mg\_l))+  
 scale\_color\_manual(values = c("#00BFC4","#F8766D"), name = "Forest type")+  
 scale\_fill\_manual(values = c("#00BFC4","#F8766D"), name = "Forest type")+  
 xlab("Water year")+  
 ylab(expression(paste("Dissolved Organic Carbon (mg",l^{-1}, ")")))+  
 facet\_wrap(~ws.f, nrow = 2)  
pdoc\_ts

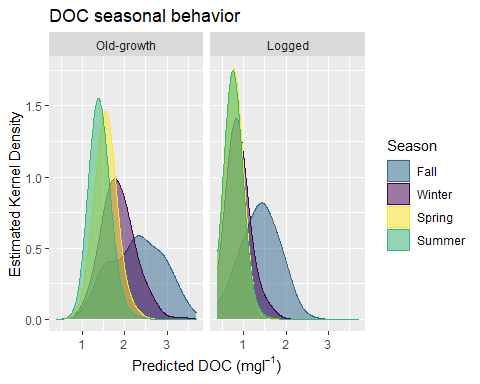
 ### Processes affecting Dissolved Organic Carbon concentrations in Headwater streams

In our model, we have formally tested hypothesis about the relationship between DOC and multiple variables (Q, P, TSS, Seasonality, etc.). Now is time to use our reconstructed historical time series to explore different processes affecting DOC concentrations in headwater streams.

#### Seasonality

ssn\_pdoc <- ggplot(hjm, aes(x=p.doc.mg\_l, color = ssn, fill = ssn))+  
 geom\_density(bw = 0.2,alpha = 0.5)+  
 scale\_color\_manual(values = c("#31688EFF","#440154FF","#FDE725FF","#35B779FF"),name ="Season")+  
 scale\_fill\_manual(values = c("#31688EFF","#440154FF","#FDE725FF","#35B779FF"),name ="Season")+  
 ggtitle ("DOC seasonal behavior")+  
 xlab(expression(paste("Predicted DOC (mg",l^{-1}, ")")))+  
 ylab("Estimated Kernel Density")+  
 facet\_wrap(~ws.f)  
ssn\_pdoc

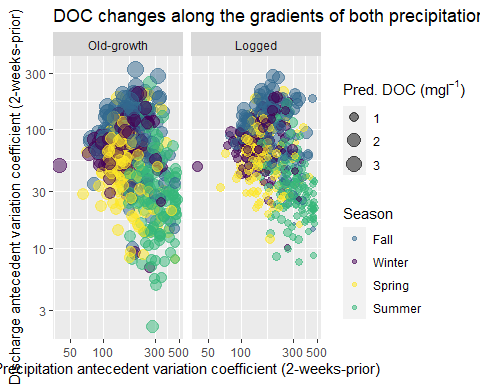
## Warning: Removed 49 rows containing non-finite values (stat\_density).

 **To do**: Add a caption to the figure and describe the differences in doc seasonal behavior between the two forest types. The graph shows that the likelihood of DOC concentrations are comparatively higher in Fall season of logged area than in old-growth while in summer chances of having lower concentration of DOC are higher in logged area than in old-growth.

#### Precipitation and Discharge antecedent variability

pqv\_pdoc <- ggplot(hjm, aes(x=pt.cv, y = q.cv, color = ssn, size = p.doc.mg\_l))+  
 geom\_point(alpha = 0.5)+  
 scale\_color\_manual(values = c("#31688EFF","#440154FF","#FDE725FF","#35B779FF"),name ="Season")+  
 scale\_size(name= expression(paste("Pred. DOC (mg",l^{-1}, ")")))+  
 scale\_x\_log10()+  
 scale\_y\_log10()+  
 ggtitle ("DOC changes along the gradients of both precipitation and discharge antecedent variability")+  
 xlab("Precipitation antecedent variation coefficient (2-weeks-prior)")+  
 ylab("Discharge antecedent variation coefficient (2-weeks-prior)")+  
 facet\_wrap(~ws.f)  
pqv\_pdoc

## Warning: Removed 49 rows containing missing values (geom\_point).

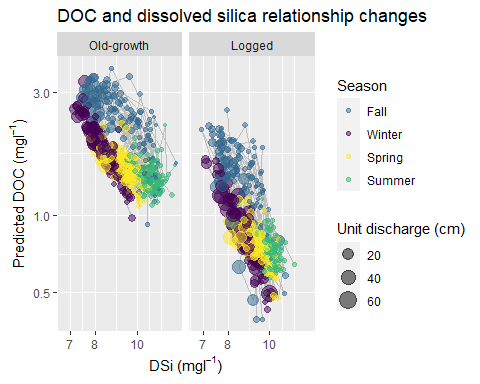


**To do**: Add a caption to the figure and describe the doc changes along the gradients of both precipitation and discharge antecedent variability. It is seen that DOC concentrations and discharge are higher in old-growth than in logged area in Fall season. As for summer, the lower discharge, the lower DOC concentrations.

#### DOC and dissolved Silica

dsi\_pdoc <- ggplot(hjm, aes(x=slc.mg\_l, y = p.doc.mg\_l, color = ssn, size = uq.cm))+  
 geom\_path(data = hjm,aes(slc.mg\_l,p.doc.mg\_l),size = 0.45, color = "gray",inherit.aes = FALSE)+  
 geom\_point(alpha = 0.5)+  
 scale\_color\_manual(values = c("#31688EFF","#440154FF","#FDE725FF","#35B779FF"),name ="Season")+  
 scale\_size("Unit discharge (cm)")+  
 scale\_x\_log10()+  
 scale\_y\_log10()+  
 ggtitle ("DOC and dissolved silica relationship changes")+  
 xlab(expression(paste("DSi (mg",l^{-1}, ")")))+  
 ylab(expression(paste("Predicted DOC (mg",l^{-1}, ")")))+  
 facet\_wrap(~ws.f)  
dsi\_pdoc

## Warning: Removed 49 rows containing missing values (geom\_point).



**To do**: Add a caption to the figure and summarize the findings from the last three figures in a short concluding paragraph.

To conclude, the figures show clear relationship between parameters. DOC concetrations clearly vary between seasons and provide extensive information on watershed biogeochemistry. The last graph shows hysterisis phenomena which clearly shows the counterwise movement.

References:

Corson-Rikert H.A., Wondzell S.M, Haggerty, R. and Santelmann, M.V. 2016. Carbon dynamics in the hyporheic zone of a headwater mountain stream in the Cascade Mountains, Oregon. Water Resources Research. 52: 7556–7576. <https://www.fs.usda.gov/treesearch/pubs/57038>.

E. J. Jantze, H. Laudon, H. E. Dahlke & S. W. Lyon (2015) Spatial Variability of Dissolved Organic and Inorganic Carbon in Subarctic Headwater Streams, Arctic, Antarctic, and Alpine Research, 47:3, 529-546, DOI: 10.1657/AAAR0014-044

Townsend-Small, A., McClelland, J.W., Max Holmes, R. et al. Seasonal and hydrologic drivers of dissolved organic matter and nutrients in the upper Kuparuk River, Alaskan Arctic. Biogeochemistry 103, 109–124 (2011). <https://doi.org/10.1007/s10533-010-9451-4>

www.statisticshowto.com [accessed 07/13/2022]

THE END!