

COMPASS: TEMPEST Discrete DOC Data QAQC

October 2024

2025-10-02

Run Information

```
#identify which section you are in
cat("Run Information")

## Run Information

#a link to the Gitbook or whatever protocol you are using for this analysis
#steph will add this soon

#anything that needs to be changed do this in the first chunk
Date_Run = "10/13/24"
Run_by = "Stephanie J. Wilson"
Script_run_by = "Stephanie J. Wilson"
run_notes = "The TN went down during this run and there were no peaks for samples,
  and there was not enough volume to rerun - making TN NAs for these samples"
#waiting to hear about samples from SW that don't have metadata

#file path and name for summary file
raw_file_name = "tmp_doc_raw_data_2024/TMP_202410.txt"
#file path and name for the all peaks file
raw_allpeaks_name = "tmp_doc_raw_data_2024/TMP_202410_allpeaks.txt"
#file path and name for processed data after QAQC
processed_file_name = "tmp_doc_processed_data_2024/TMP_PW_DOC_Processed_202410.csv"

#check standard concentrations - Update if running different checks:
chk_std_c = 50
chk_std_n = 2

#Log path
Log_path = "tmp_doc_raw_data_2024/COMPASS_TMP_TOCTN_QAQClog_2024.csv"
```

Setup

Pull in active porewater tracking inventory sheet

```
## File already exists. No download needed.
```

Import Data Functions

Import Sample Data

```
## Import Sample Data
```

```
## New names:
```

```
## * ' ' -> '...14'
```

```
## # A tibble: 6 x 4
```

```
##   sample_name      npoc_raw  tdn_raw run_datetime
##   <chr>          <dbl>    <dbl> <chr>
## 1 TMP_20241003_CON_C6      68.8 -0.00364 10/13/2024 7:22:16 AM
## 2 TMP_20241003_SW_B4      27.4  0.00553 10/13/2024 8:01:42 AM
## 3 TMP_20241003_SW_C3      31.8  0.00192 10/13/2024 8:48:18 AM
## 4 TMP_20241007_SW_C6      29.7 -0.00157 10/13/2024 9:33:43 AM
## 5 TMP_20241007_SW_D5      15.0  0.00854 10/13/2024 10:17:39 AM
## 6 TMP_20241007_SW_E3      11.6  0.0227  10/13/2024 10:59:44 AM
```

Assessing standard Curves

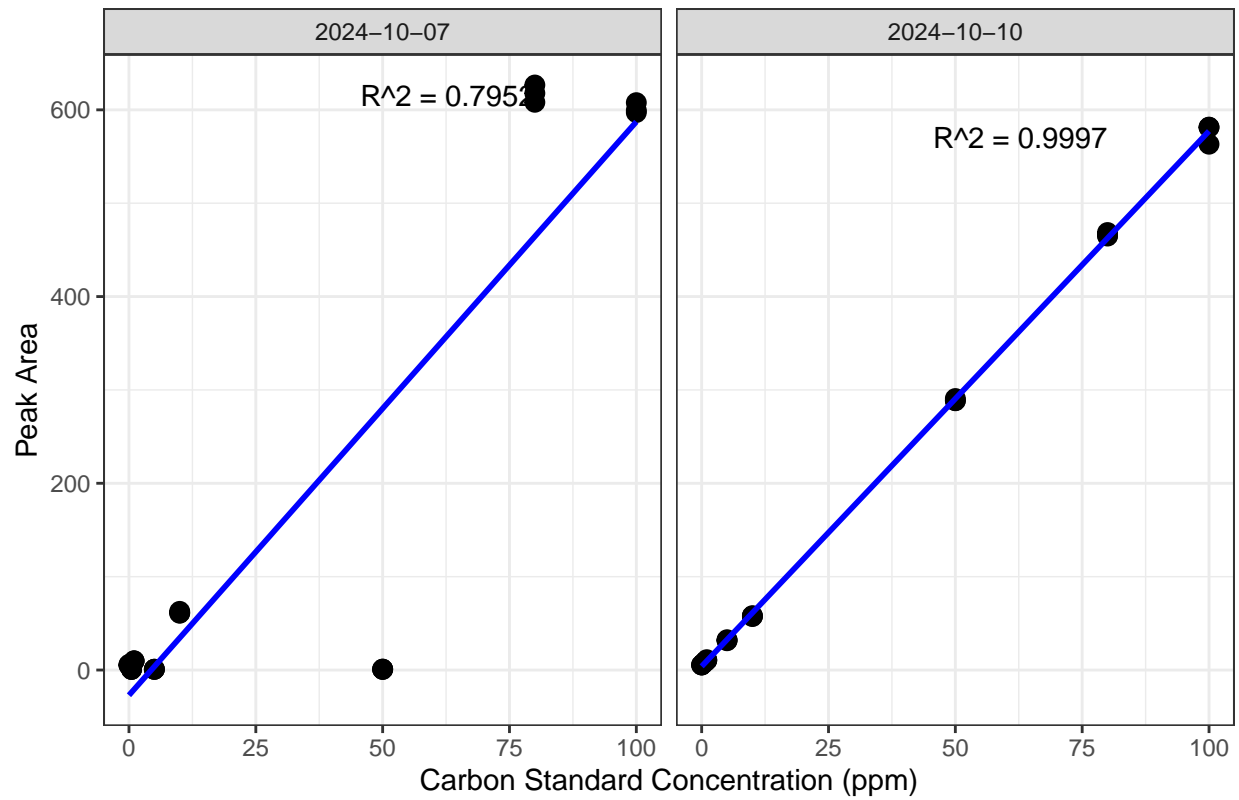
```
## Assess the Standard Curve
```

```
## New names:
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

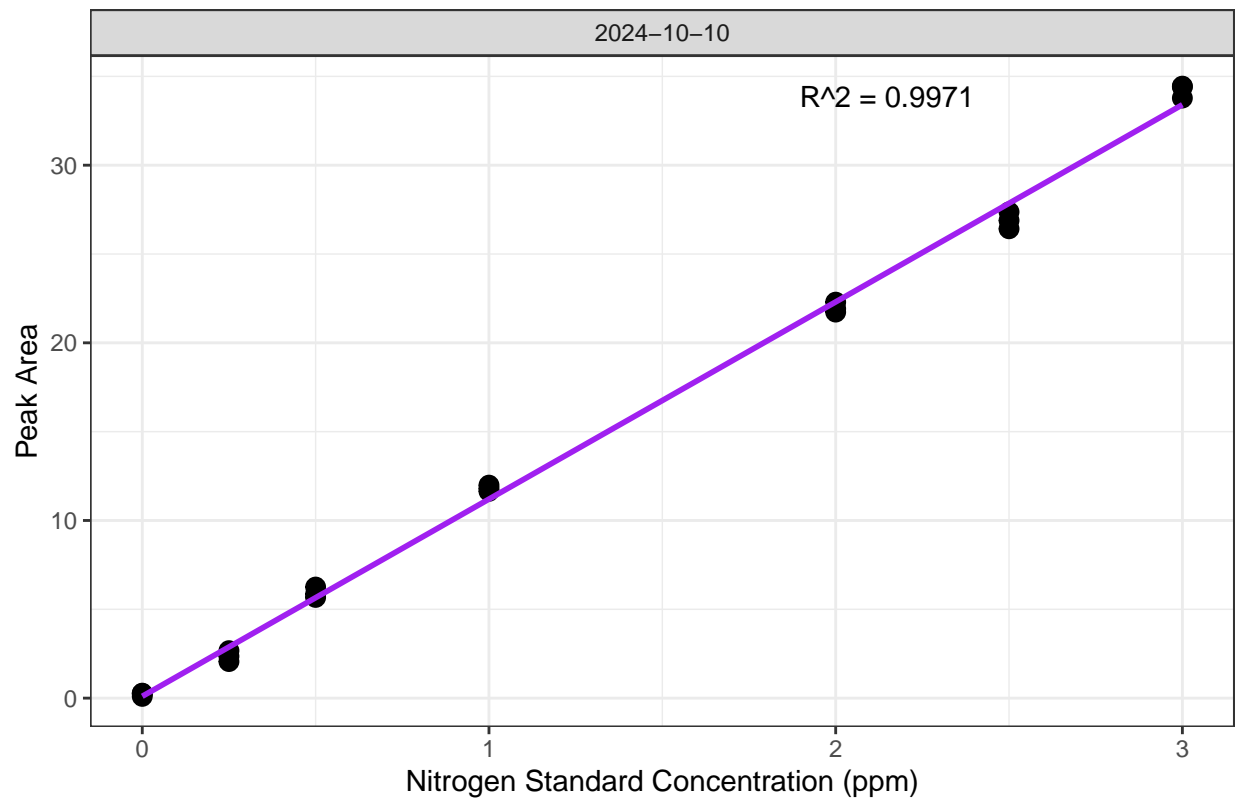
```
## * '' -> '...18'
```

NPOC Std Curve by Date



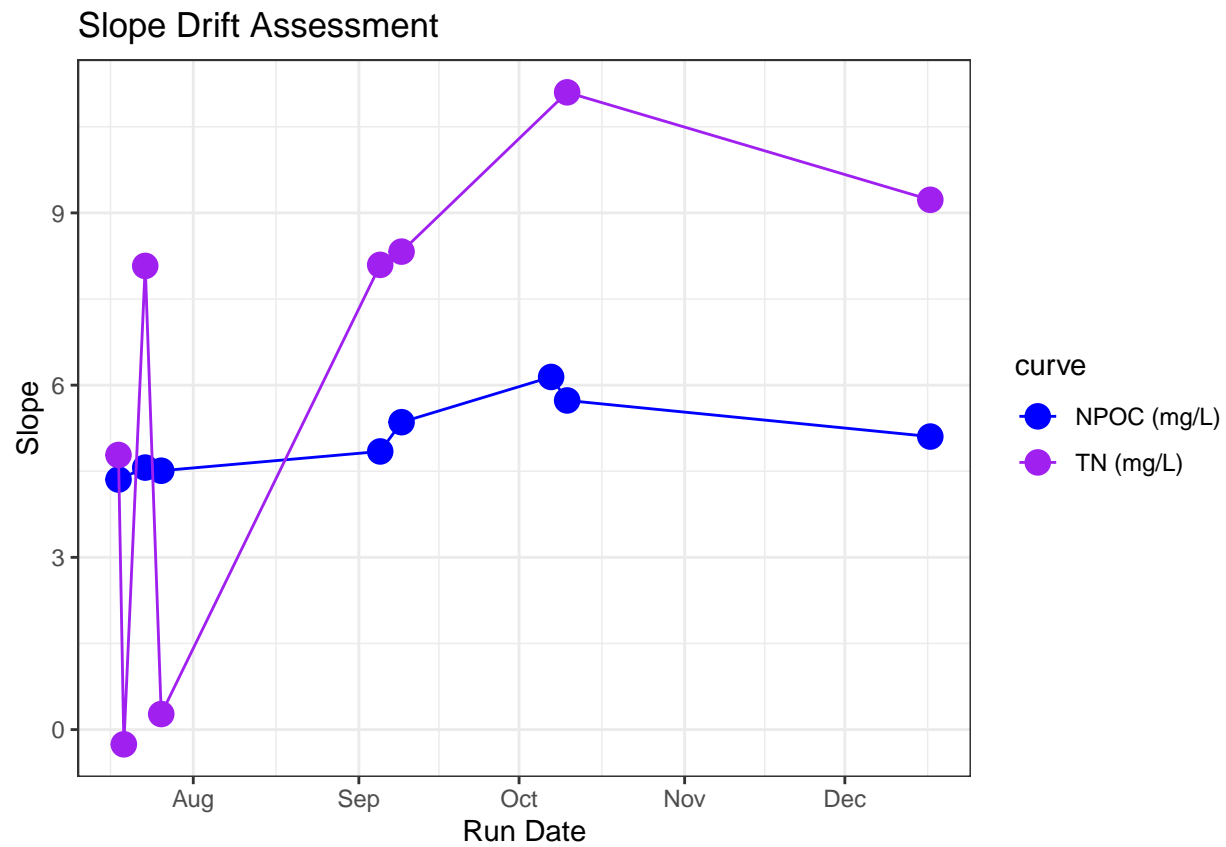
```
## 'geom_smooth()' using formula = 'y ~ x'
```

TN Std Curve by Date



```
## Warning: Removed 15 rows containing missing values or values outside the scale range
## ('geom_point()').
```

```
## Warning: Removed 15 rows containing missing values or values outside the scale range
## ('geom_line()').
```



```
## [1] "NPOC Curve r2 GOOD"
```

```
## [1] "TN Curve r2 GOOD"
```

Assess Check Standards

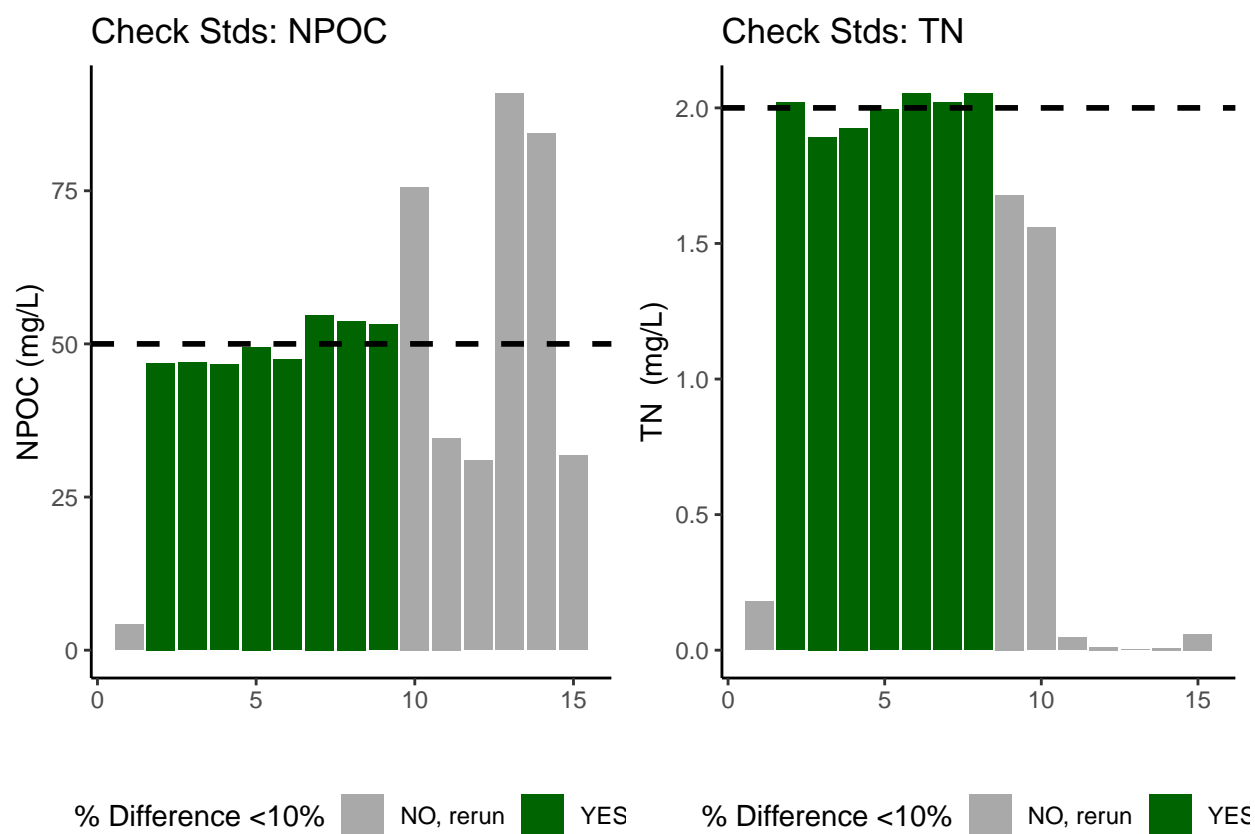
```
## Assess the Check Standards
```

```
## New names:
```

```
## * ' ' -> '...14'
```

```
## [1] "Carbon CHECK STANDARD RSD TOO HIGH - REASSESS"
```

```
## [1] "Nitrogen CHECK STANDARD RSD TOO HIGH - REASSESS"
```



```
## [1] "<60% of Carbon Check Standards are within range of the expected concentration - REASSESS"
```

```
## [1] "<60% of Nitrogen Check Standards are within range of the expected concentration - REASSESS"
```

Assess Blanks

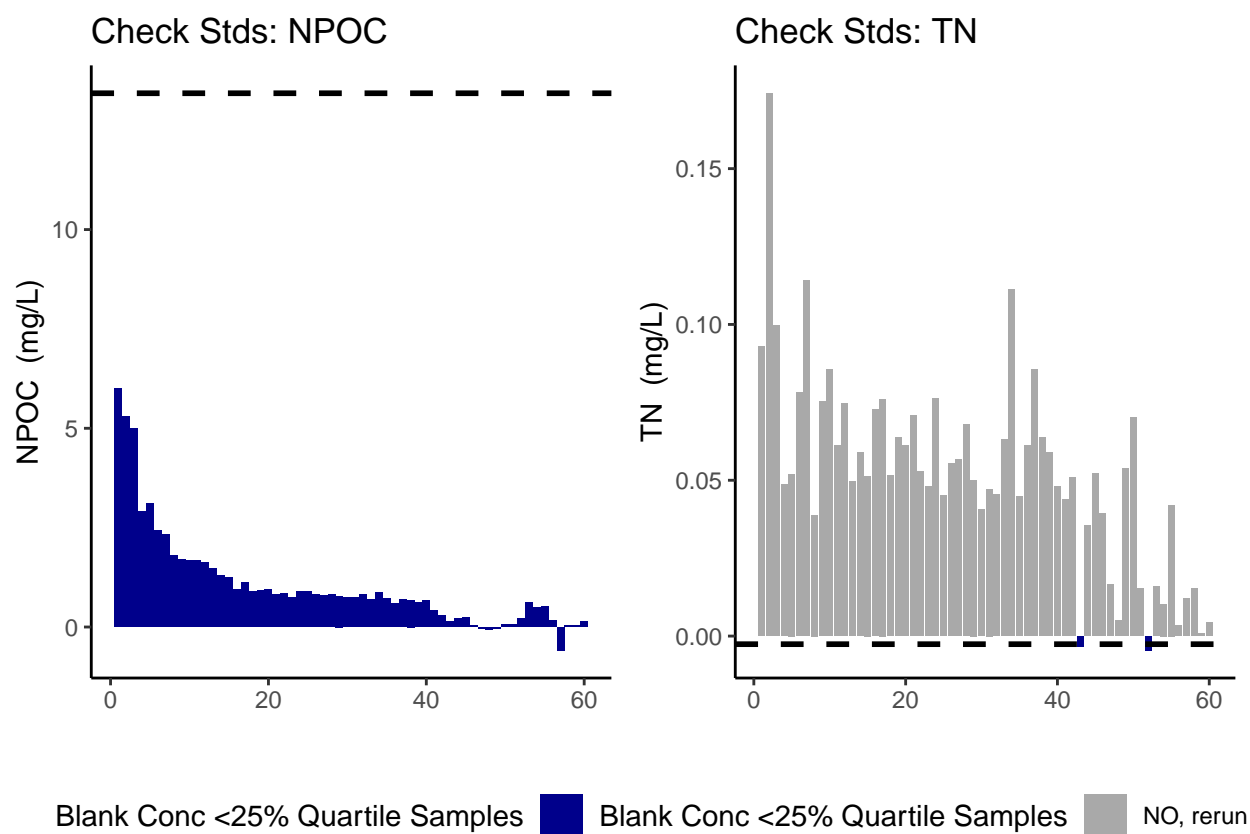
```
## Assess Blanks
```

```
## New names:
```

```
## * ' ' -> '...14'
```

```
## [1] ">60% of Carbon Blank concentrations are below the lower 25% quartile of samples"
```

```
## [1] "<60% of Nitrogen blanks are higher than the lower 25% quartile of samples - REASSESS"
```



```
## carbon blanks:
```

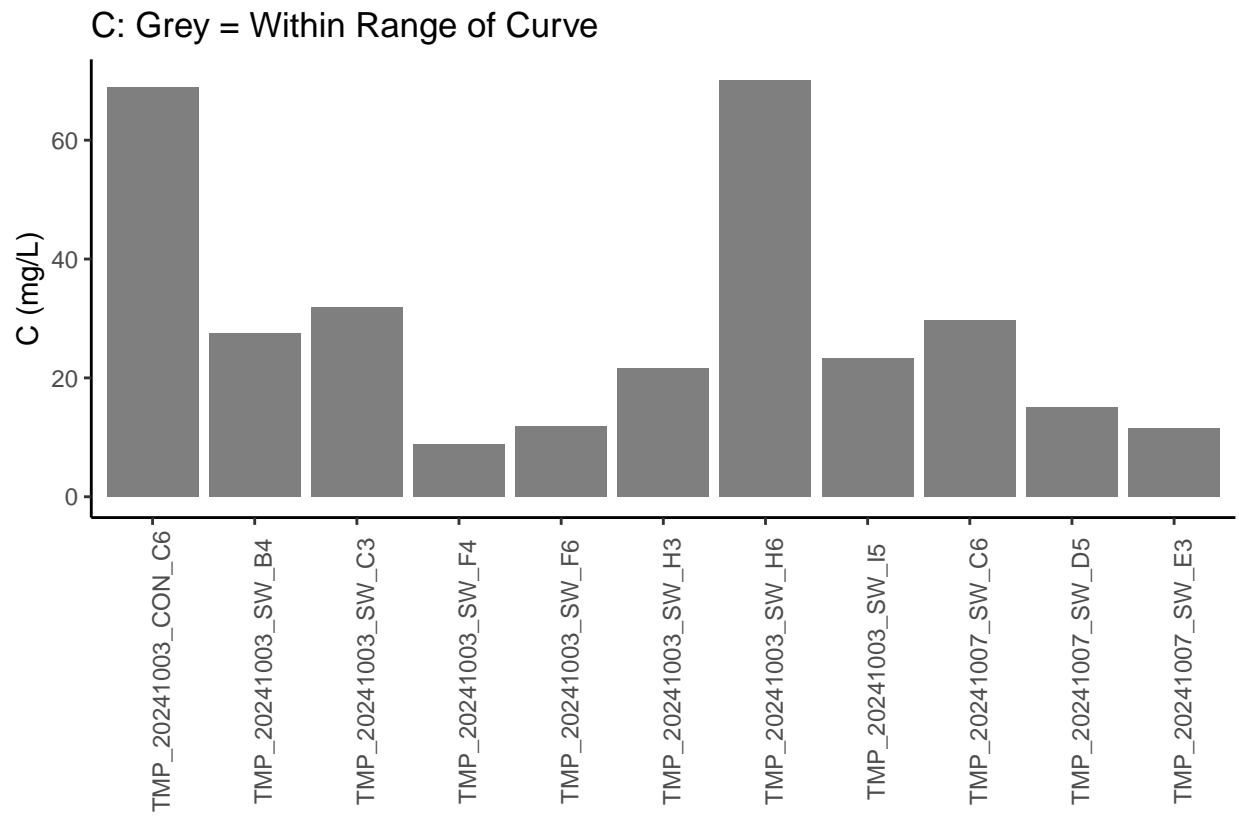
```
## [1] 1.04306
```

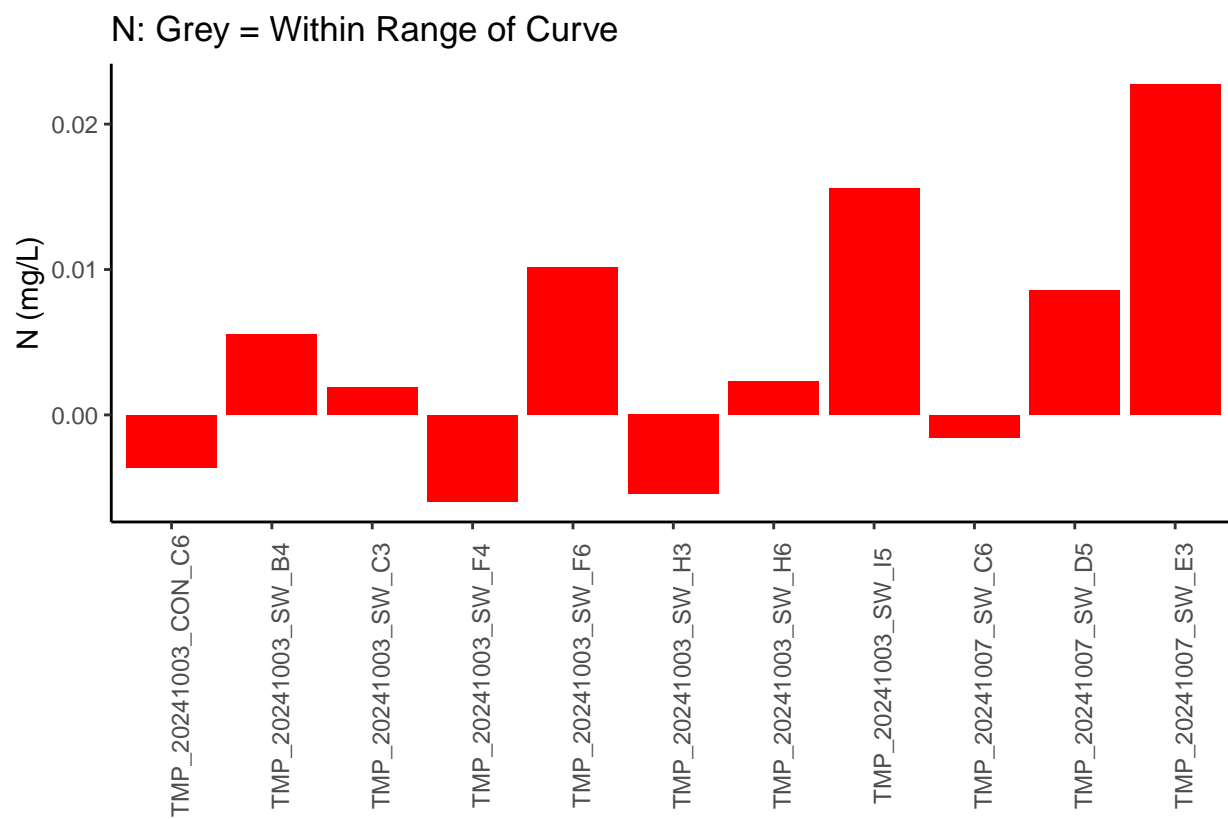
```
## nitrogen blanks:
```

```
## [1] 0.052526
```

Sample Flagging

Sample Flagging





Visualize Data by Plot

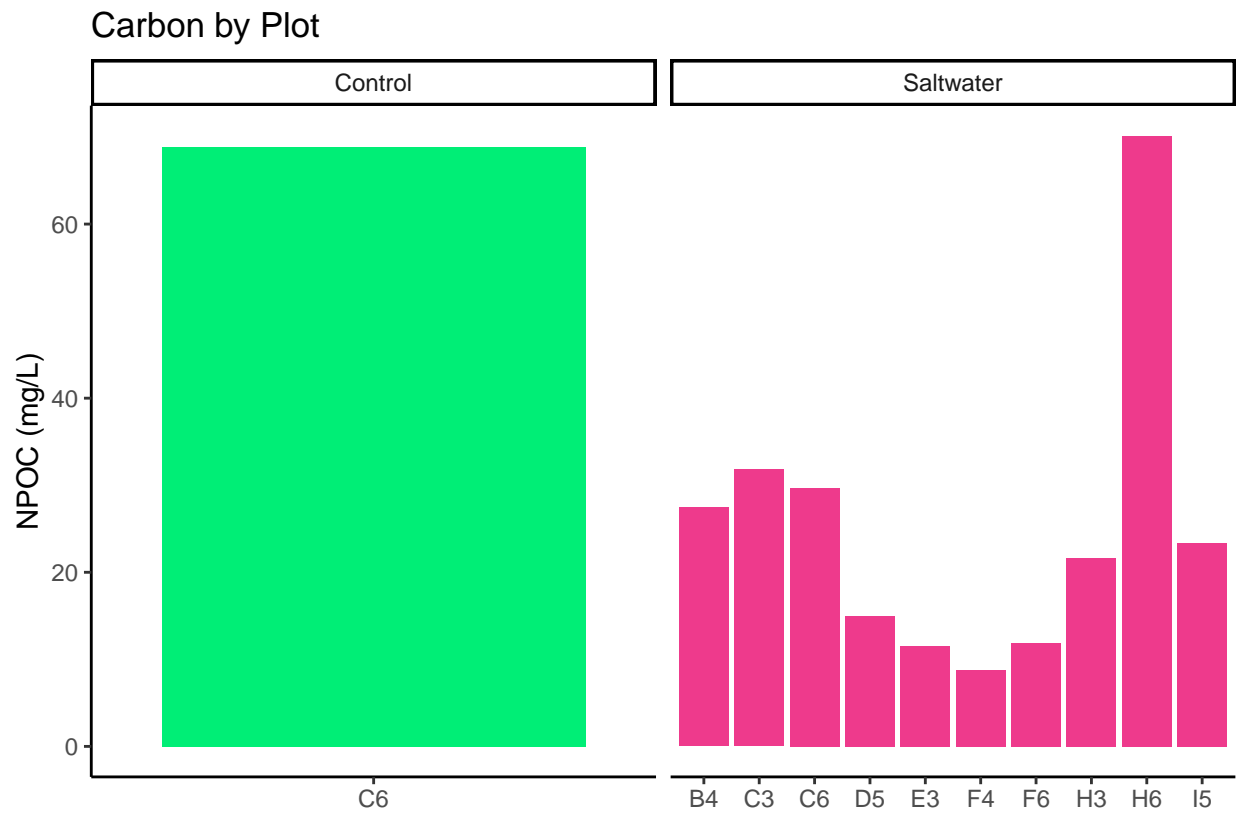
Visualize Data

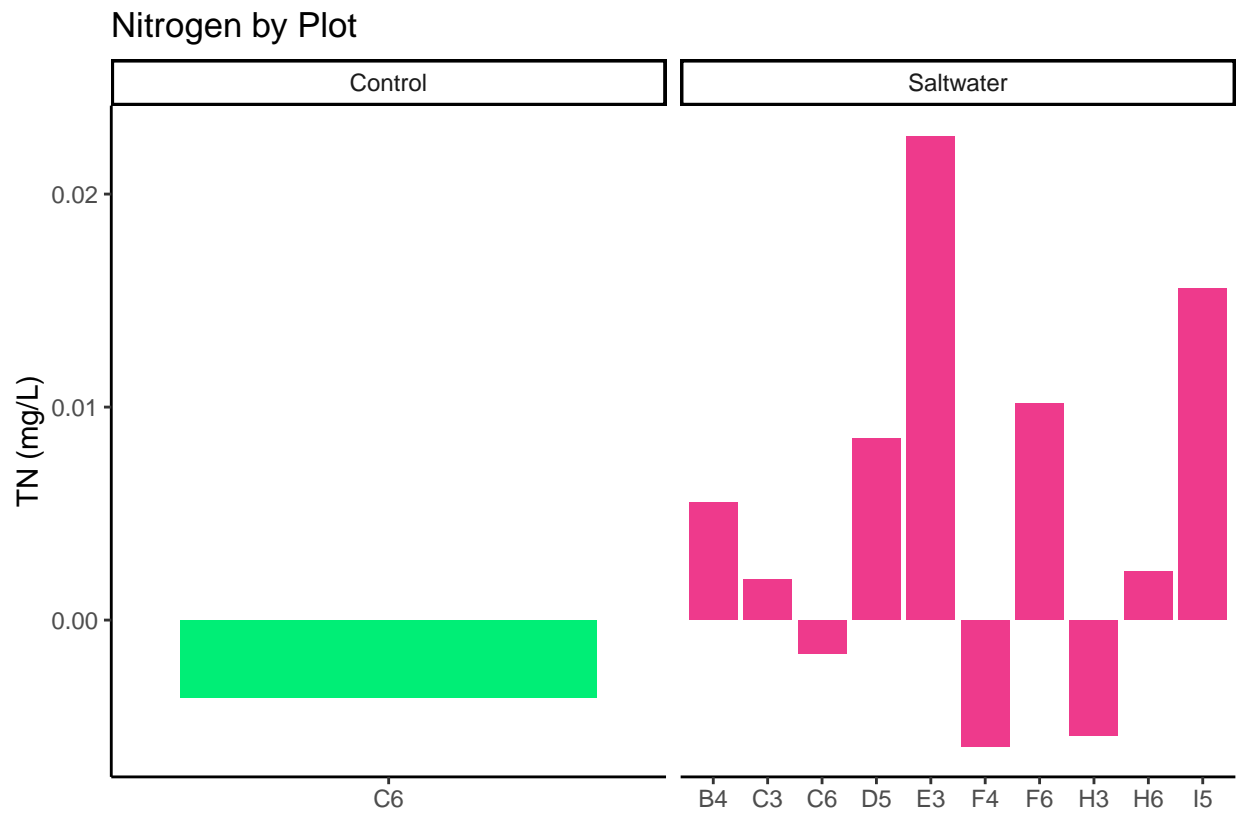
##	Site_Code	Date	Plot	Grid_Square
## 1	TMP	20241003	CON	C6
## 2	TMP	20241003	SW	B4
## 3	TMP	20241003	SW	C3
## 4	TMP	20241007	SW	C6
## 5	TMP	20241007	SW	D5
## 6	TMP	20241007	SW	E3

##	Site_Code	Date	Plot	Grid_Square	sample_name	npoc_raw	tdn_raw
## 1	TMP	20241003	CON	C6	TMP_20241003_CON_C6	68.83	-0.00364
## 2	TMP	20241003	SW	B4	TMP_20241003_SW_B4	27.44	0.00553
## 3	TMP	20241003	SW	C3	TMP_20241003_SW_C3	31.81	0.00192
## 4	TMP	20241007	SW	C6	TMP_20241007_SW_C6	29.69	-0.00157
## 5	TMP	20241007	SW	D5	TMP_20241007_SW_D5	15.02	0.00854
## 6	TMP	20241007	SW	E3	TMP_20241007_SW_E3	11.55	0.02272

##	run_datetime	npoc_flag
## 1	10/13/2024 7:22:16 AM	NPOC checks out of range
## 2	10/13/2024 8:01:42 AM	NPOC checks out of range
## 3	10/13/2024 8:48:18 AM	NPOC checks out of range
## 4	10/13/2024 9:33:43 AM	NPOC checks out of range
## 5	10/13/2024 10:17:39 AM	NPOC checks out of range
## 6	10/13/2024 10:59:44 AM	NPOC checks out of range

##	tdn_flag
## 1	TN checks out of range; TN blanks out of range; blank is > 25% of sample value
## 2	TN checks out of range; TN blanks out of range; blank is > 25% of sample value
## 3	TN checks out of range; TN blanks out of range; blank is > 25% of sample value
## 4	TN checks out of range; TN blanks out of range; blank is > 25% of sample value
## 5	TN checks out of range; TN blanks out of range; blank is > 25% of sample value
## 6	TN checks out of range; TN blanks out of range; blank is > 25% of sample value





Convert data from mg/L to uMoles/L

Add in/check metadata

```
## Check Sample IDs with Metadata
```

```
## # A tibble: 11 x 2
##   sample_name      metadata_recorded
##   <chr>           <lgl>
## 1 TMP_C_C6_20241003 TRUE
## 2 TMP_SW_B4_20241003 TRUE
## 3 TMP_SW_C3_20241003 TRUE
## 4 TMP_SW_C6_20241007 TRUE
## 5 TMP_SW_D5_20241007 FALSE
## 6 TMP_SW_E3_20241007 FALSE
## 7 TMP_SW_F4_20241003 TRUE
## 8 TMP_SW_F6_20241003 TRUE
## 9 TMP_SW_H3_20241003 TRUE
## 10 TMP_SW_H6_20241003 TRUE
## 11 TMP_SW_I5_20241003 TRUE
```

Export Processed Data

```
## Export Processed Data
```

```
## # A tibble: 6 x 21
##   Project      plot grid Depth_cm sample_type Vial_ID date npoc_mgL npoc_uM
##   <chr>      <chr> <chr>   <dbl> <chr>      <chr>  <chr>   <dbl>   <dbl>
## 1 COMPASS: TEMP~ C    C6      15 DOC      C_C6_D~ 2024~    68.8    5736.
## 2 COMPASS: TEMP~ SW    B4      15 DOC      SW_B4_~ 2024~    27.4    2287.
## 3 COMPASS: TEMP~ SW    C3      15 DOC      SW_C3_~ 2024~    31.8    2651.
## 4 COMPASS: TEMP~ SW    C6      15 DOC      SW_C6_~ 2024~    29.7    2474.
## 5 <NA>        <NA> <NA>     15 <NA>      <NA>    <NA>    15.0    1252.
## 6 <NA>        <NA> <NA>     15 <NA>      <NA>    <NA>    11.6    963.
## # i 12 more variables: npoc_flag <chr>, tdn_mgL <dbl>, tdn_uM <dbl>,
## #   tdn_flag <chr>, Analysis_runtime <chr>, Run_notes <chr>,
## #   Evacuation_date_YYYYMMDD <dbl>, Collection_Date_YYYYMMDD <dbl>,
## #   Collection_Start_Time_24hrs <dbl>, Collection_End_Time_24hrs <dbl>,
## #   EST_EDT <chr>, Volume_mL <dbl>
```

```
#end
```

Assess Duplicates - No duplicates on this run

```

“{#r Check Duplicates, echo=FALSE}
cat("Assess Duplicates")
#Take a look at the raw data #head(dat_raw)

#pull out any rows that have "dup" in the sample_name column dups <- dat_raw %>%
select(!c(npoc_flag, tdn_flag)) %>% filter(str_detect(sample_name, "dup")) #have to change this to match
data

#create a new dataframe and remove dups from sample dataframe dat_raw2 <- dat_raw %>%
filter(!str_detect(sample_name, "dup"))

#remove the dup from these IDs so we will have duplicate sample names dupssample_name <-
gsub("dup", "", as.character(dupssample_name)) dups <- dups[ ,c(4)] #remove the run date time
for colnames(dups) <- c('sample_name', 'npoc_raw_dup', 'tdn_raw_dup') head(dups)

QAdups <- merge(dat_raw2, dups) head(QAdups)

df2 <- as.data.frame(QAdupsnpoc_raw) df2dups <- QAdups$npoc_raw_dup
df2sds <- apply(df2, 1, sd) df2mean <- apply(df2, 1, mean)

QAdupsnpoc_dups_cv <- (df2sds/df2mean)*100 QAdupsnpoc_dups_cv_flag <- ifelse(QAdups$npoc_dups_cv
<10, 'YES', 'NO, rerun')

df3 <- as.data.frame(QAdupstdn_raw) df3dups <- QAdups$tdn_raw_dup
df3sds <- apply(df3, 1, sd) df3mean <- apply(df3, 1, mean)

QAdupstdn_dups_cv <- (df3sds/df3mean) * 100 QAdupstdn_dups_cv_flag <- ifelse(QAdups$tdn_dups_cv
<10, 'YES', 'NO, rerun')

head(QAdups)

#plot dups output as a bar graph to easily check - want any over 10% to be red need to work on this
C_dups <- ggplot(data = QAdups, aes(x = sample_name, y = npoc_dups_cv, fill = npoc_dups_cv_flag))
+ geom_bar(stat = 'identity') + theme_classic() + labs(x = "Sample ID", y = "CV of NPOC
Dups (%)") + scale_fill_manual(values = c("YES" = "darkgreen", "NO, rerun" = "red")) +
theme(legend.position = "none") + geom_hline(yintercept = 10, linetype = "dashed", color = "black", size = 1)
+ guides(fill = guide_legend(title = "CV Between Dups <10%")) + theme(axis.text.x = element_text(angle
= 90, hjust = 0.5))

N_dups <- ggplot(data = QAdups, aes(x = sample_name, y = tdn_dups_cv, fill = tdn_dups_cv_flag)) +
geom_bar(stat = 'identity') + theme_classic() + labs(x = "Sample ID", y = "CV of TN Dups (%)") +
scale_fill_manual(values = c("YES" = "darkgreen", "NO, rerun" = "red")) + theme(legend.position = "none")
+ geom_hline(yintercept = 10, linetype = "dashed", color = "black", size = 1) + guides(fill = guide_legend(title = "CV
Between Dups <10%")) + theme(axis.text.x = element_text(angle = 90, hjust = 0.5))

ggarrange(C_dups, N_dups, ncol = 2, nrow = 1)

#calculate the percent of check standards that are within the range based on the flag c_dups_percent <-
(sum(QAdupsnpoc_dups_cv_flag == "YES")/nrow(QAdups))*100 n_dups_percent <- (sum(QAdupstdn_dups_cv_flag
== "YES")/nrow(QAdups))*100

#report out if flags indicate need for rerun ifelse(c_dups_percent >= chks_flag, ">60% of Car-
bon Duplicates have a CV <10%", "<60% of Carbon Duplicates have a CB <10% - REASSESS")
ifelse(n_dups_percent >= chks_flag, ">60% of Nitrogen Duplicates have a CV <10%", "<60% of Nitrogen
Duplicates have a CB <10% - REASSESS")

#write out a flag to the sample dataframe if more than 60% of the dups have CVs out of range if
(c_dups_percent <= chks_flag) { dat_rawnpoc_flag <- ifelse(dat_rawnpoc_flag != "", paste0(dat_raw$npoc_flag, ";
NPOC dups out of range"), "NPOC dups out of range" ) }

```

```

if (n_dups_percent <= chks_flag) { # assuming you have tn_chks_percent similarly dat_raw $dn_{flag}$  <
-ifelse(dat $_{raw}$ tn_flag != "", paste0(dat_raw$tn_flag, "; TN dups out of range"), "TN dups out of range"
) }
""

```