COMPASS: TEMPEST Discrete DOC Data QAQC

September 2025

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 = "07/03/25"
  Run_by = "Stephanie J. Wilson"
  Script_run_by = "Stephanie J. Wilson"
  run_notes = "Run by Watershed Science at SERC, no duplicates included"

  #file path and name for summary file
    raw_file_name = "tmp_doc_raw_data_2025/TMP_202509.txt"
  #file path and name for the all peaks file
    raw_allpeaks_name = "tmp_doc_raw_data_2025/TMP_202509_allpeaks.txt"
  #file path and name for processed data after QAQC
    processed_file_name = "tmp_doc_processed_data_2025/TMP_PW_DOC_Processed_202509.csv"

#check standard concentrations - Update if running different checks:
   chk_std_c = 30
   chk_std_n = 2.5

#Log path
    Log_path = "tmp_doc_raw_data_2025/COMPASS_TMP_TOCTN_QAQClog_2025.csv"
```

Setup

Pull in active porewater tracking inventory sheet

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

Import Data Functions

Import Sample Data

```
## Import Sample Data
```

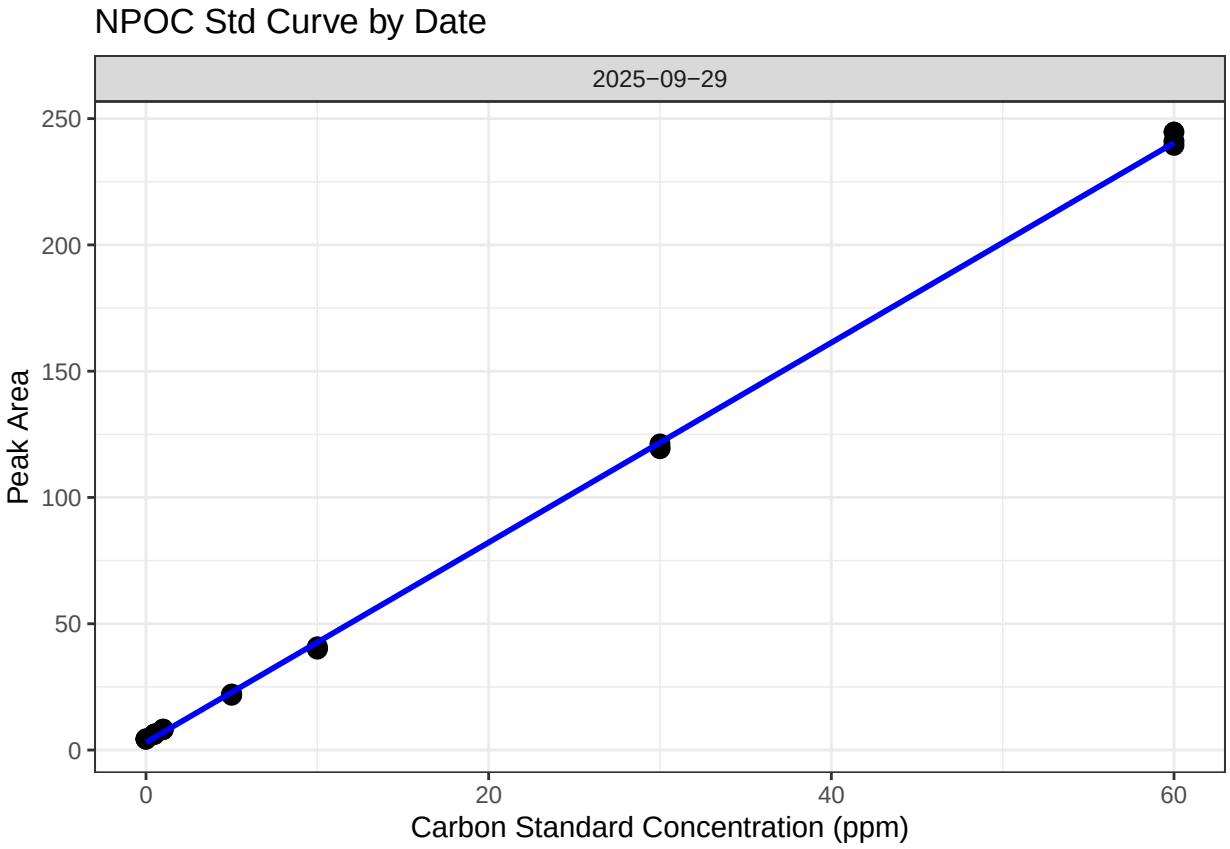
```
## New names:
## * `` -> `...14`
```

```
## # A tibble: 3 x 4
##   sample_name        npoc_raw tdn_raw run_datetime
##   <chr>                 <dbl>   <dbl> <chr>
## 1 TMP_SW_D5_20250926     15.7   0.605 10/1/2025 10:45:08 PM
## 2 TMP_SW_H3_20250929     17.7   0.643 10/1/2025 11:05:35 PM
## 3 TMP_SW_H6_20250926     20.8   0.868 10/1/2025 11:33:43 PM
```

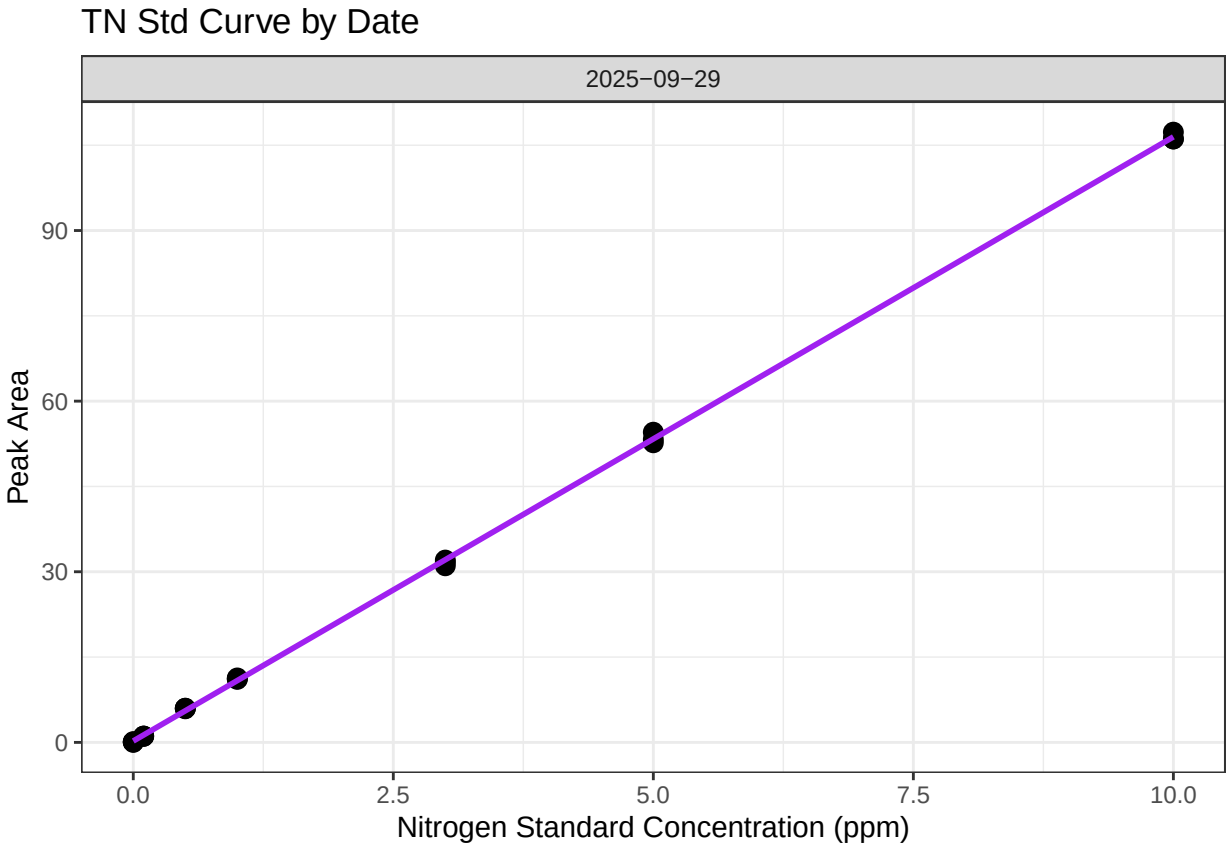
Assessing standard Curves

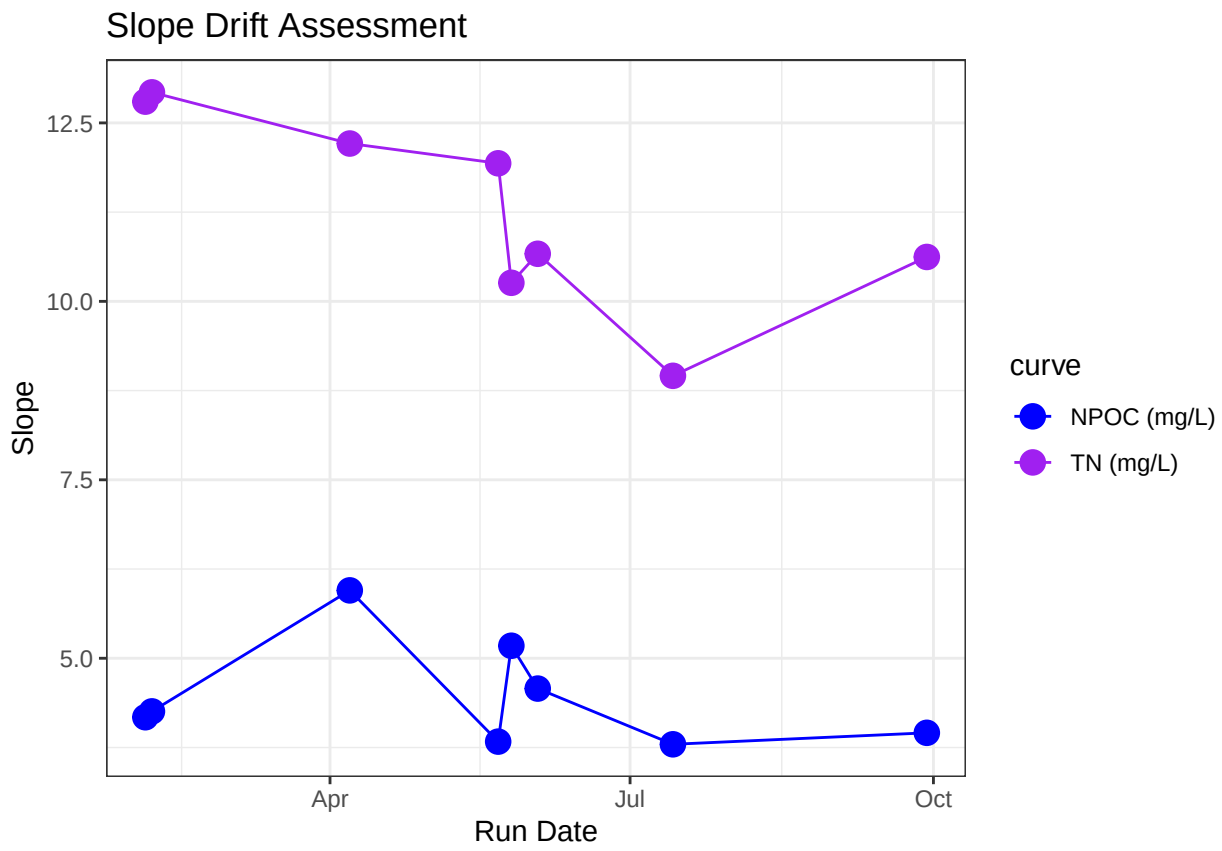
```
## Assess the Standard Curve
```

```
## New names:
## ‘geom_smooth()‘ using formula = ’y ~ x’
## * ‘‘ -> ‘...18‘
```




```
## ‘geom_smooth()‘ using formula = ’y ~ x’
```


TN Std Curve by Date
2025−09−29
Peak Area
90
60
30
0
0.0
2.5
5.0
7.5
10.0
Nitrogen Standard Concentration (ppm)




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

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

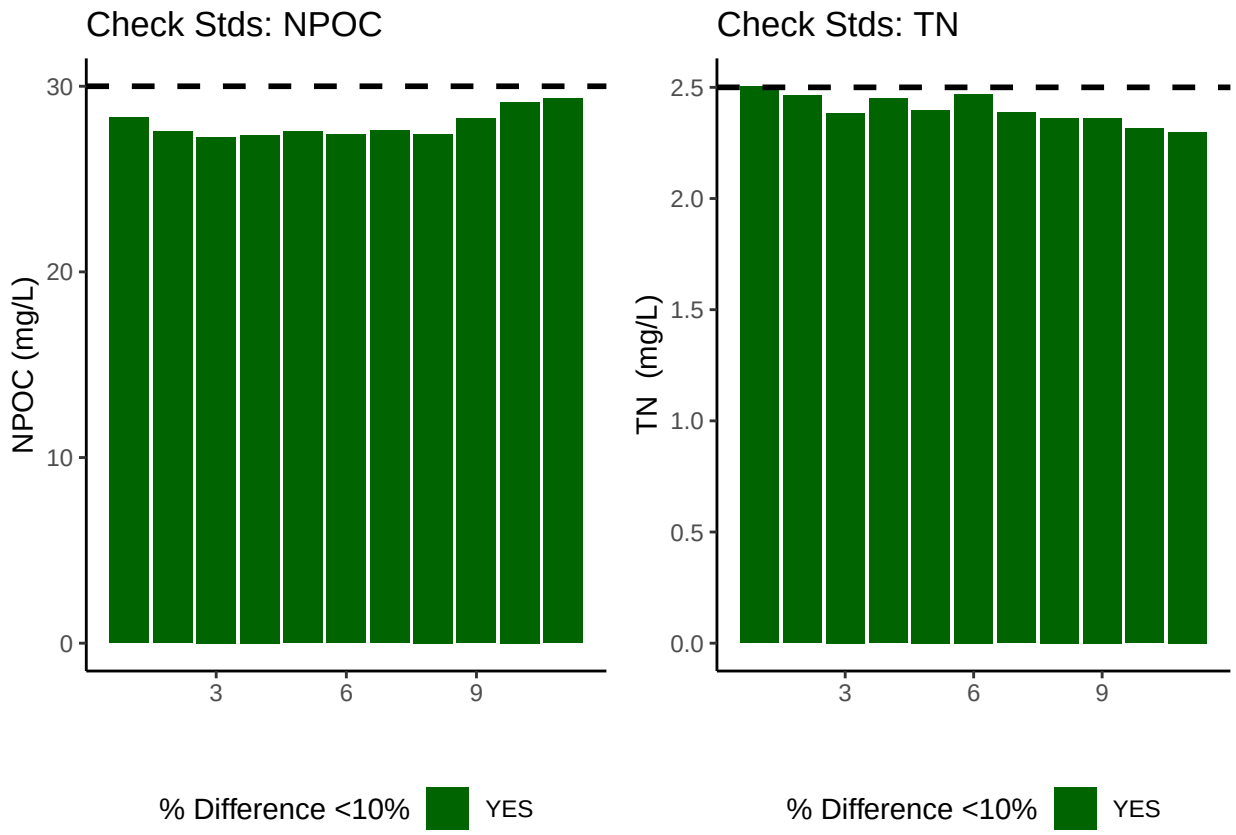
Assess Check Standards

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

```
## New names:
## * `` -> `...14`
```

```
## [1] "Carbon Check Standard RSD within Range"
```

```
## [1] "Nitrogen Check Standard RSD within Range"
```




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

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

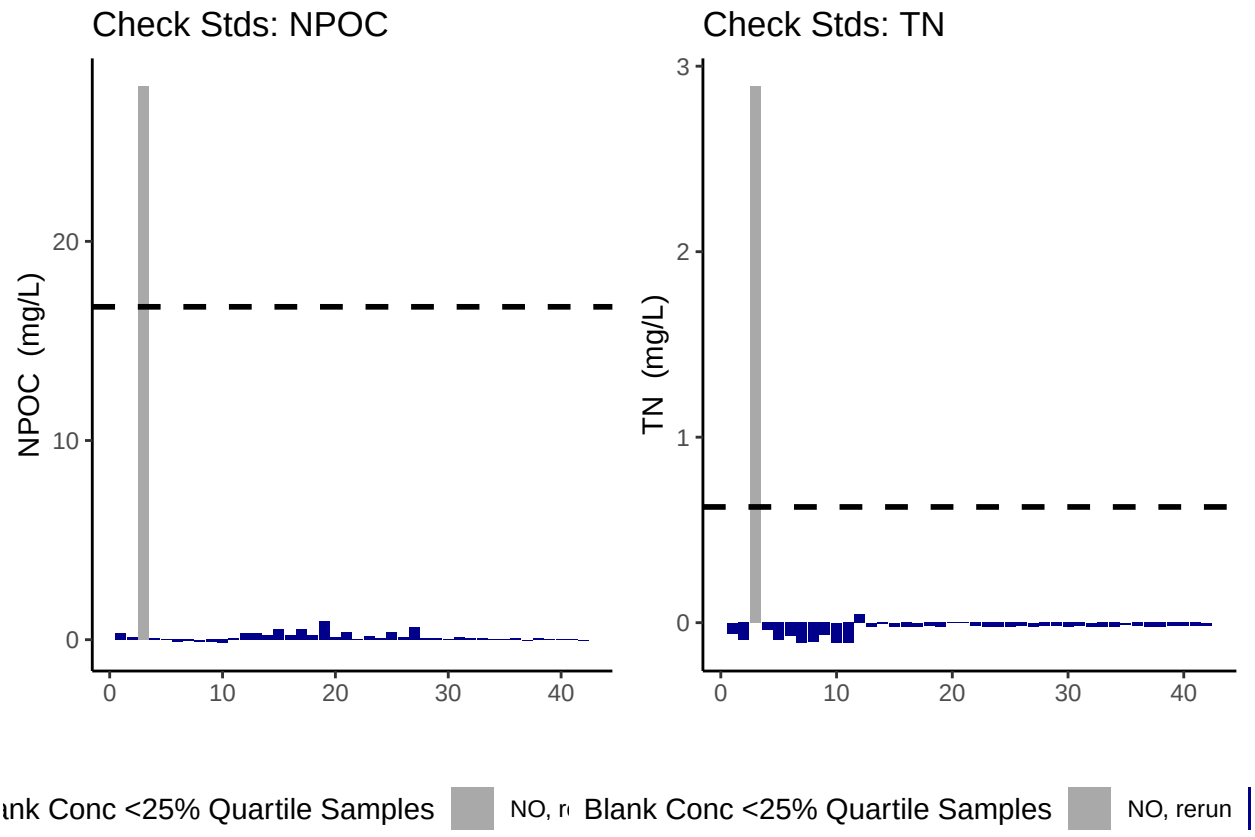
Assess Blanks

```
## Assess Blanks
```

```
## New names:
## * `` -> `...14`
```

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

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




```
## carbon blanks:
```

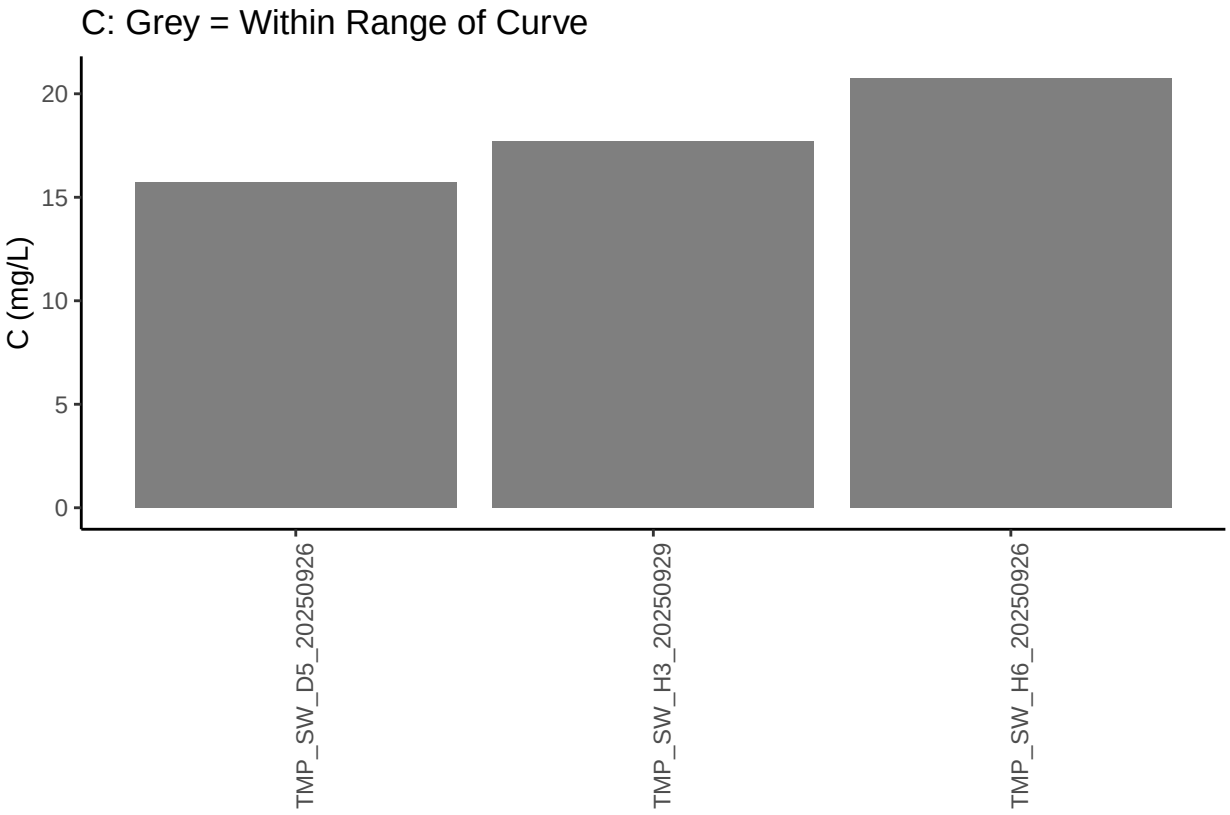
```
## [1] 0.7973995
```

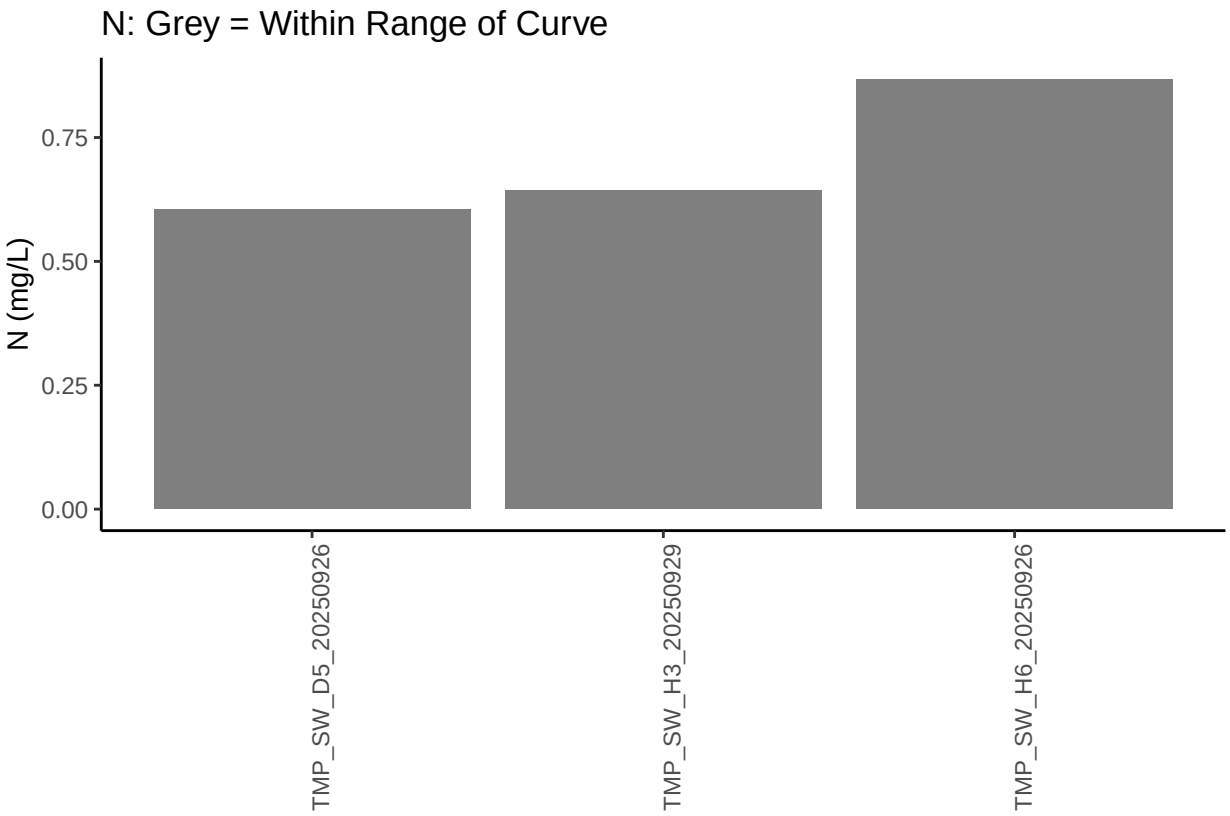
```
## nitrogen blanks:
```

```
## [1] 0.03649595
```

Sample Flagging

```
## Sample Flagging
```



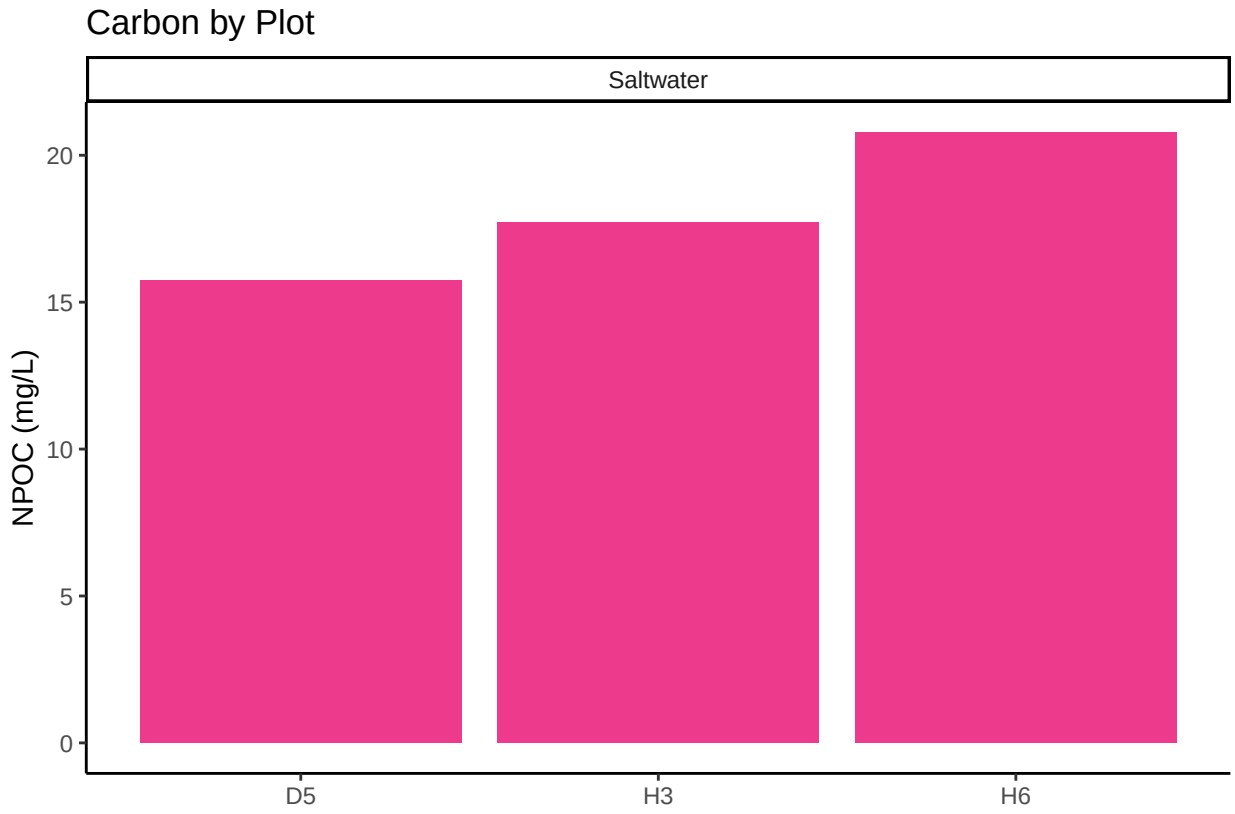

N: Grey = Within Range of Curve
N (mg/L)
0.75
0.50
0.25
0.00
TMP_SW_D5_20250926
TMP_SW_H3_20250929
TMP_SW_H6_20250926

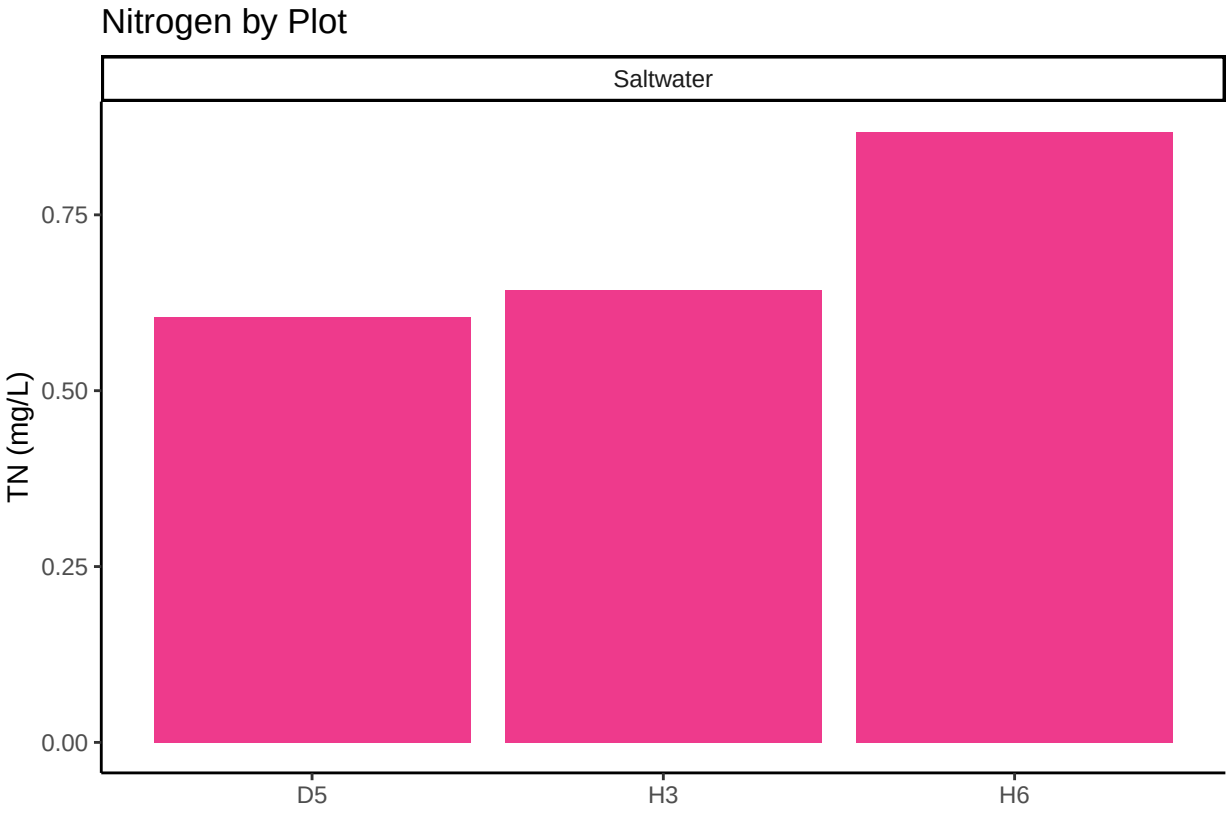
Visualize Data by Plot

```
## Visualize Data
```

```
##   Site_Code Plot Grid_Square     Date
## 1       TMP   SW          D5 20250926
## 2       TMP   SW          H3 20250929
## 3       TMP   SW          H6 20250926
```

```
##   Site_Code Plot Grid_Square     Date        sample_name npoc_raw tdn_raw
## 1       TMP   SW          D5 20250926 TMP_SW_D5_20250926    15.73  0.6051
## 2       TMP   SW          H3 20250929 TMP_SW_H3_20250929    17.70  0.6428
## 3       TMP   SW          H6 20250926 TMP_SW_H6_20250926    20.77  0.8676
##            run_datetime npoc_flag tdn_flag
## 1 10/1/2025 10:45:08 PM
## 2 10/1/2025 11:05:35 PM
## 3 10/1/2025 11:33:43 PM
```




Nitrogen by Plot
Saltwater
TN (mg/L)
0.75
0.50
0.25
0.00
D5
H3
H6

Convert data from mg/L to uMoles/L

Add in/check metadata

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

```
## # A tibble: 3 x 2
##   sample_name        metadata_recorded
##   <chr>              <lgl>
## 1 TMP_SW_D5_20250926 TRUE
## 2 TMP_SW_H3_20250929 TRUE
## 3 TMP_SW_H6_20250926 TRUE
```

Export Processed Data

```
## Export Processed Data
```

```
## # A tibble: 3 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~ SW    D5          15 DOC         SW_D5_~ 2025~     15.7   1311.
## 2 COMPASS: TEMP~ SW    H3          15 DOC         SW_H3_~ 2025~     17.7   1475
## 3 COMPASS: TEMP~ SW    H6          15 DOC         SW_H6_~ 2025~     20.8   1731.
## # i 12 more variables: npoc_flag <chr>, tdn_mgL <dbl>, tdn_uM <dbl>,
## #   tdn_flag <chr>, Analysis_runtime <chr>, Run_notes <chr>,
## #   Evacuation_date_YYYMMDD <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 dups$sample_name <- -gsub(”_dup”, ””, as.character(dups$sample_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(QAdups$npoc_raw)df2dups <- QAdups$npoc_raw_dup

df2sds <- -apply(df2, 1, sd)df2mean <- apply(df2, 1, mean)

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

df3 <- as.data.frame(QAdups$tdn_raw)df3dups <- QAdups$tdn_raw_dup

df3sds < -apply(df3, 1, sd)df3mean <- apply(df3, 1, mean)

QAdups$tdn_dups_cv < -(df3sds/df3mean) * 100QAdups$tdn_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(QAdups$npoc_dups_cv_flag == "YES")/nrow(QAdups))*100n_dups_percent < -(sum(QAdups$tdn_dups_cv_flag == "YES")/nrow(QAdups))*100

#report out if flags indicate need for rerun ifelse(c_dups_percent >= chks_flag, ">60% of Carbon 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_raw$npoc_flag < -ifelse(dat_raw$npoc_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$tdn_flag < -ifelse(dat_raw$tdn_flag != "", paste0(dat_raw$tdn_flag,"; TN dups out of range"), "TN dups out of range") }

```
```