EMP Continuous WQ Metrics - 2023

Ted Flynn

2024-02-16

Process and plot Wagner rating data from C-EMP sondes.

Read in Raw Data

Import the raw data from C-EMP and perform basic QA-QC. Check station completeness and screen for duplicates.

```r
# Import C-EMP data files
df_CEMP <- read_csv(file = here("04_Other",
                                "performance-metrics",
                                "data",
                                "2023_wagner_newer.csv"),
                    show_col_types = FALSE)

## Remove date time column (will only consider rating by date)
df_CEMP <- df_CEMP %>% select(-visit_time)

# Check for missing dates
df_CEMP %>% filter(is.na(visit_date)) # No missing dates

# Select only needed columns
df_CEMP <- df_CEMP %>% select(site, visit_date, analyte_name, rating)

# Check for station completeness
unique(df_CEMP$site)
summary(unique(df_CEMP$site)) # Lists all 15 sites

# Check for duplicated lines with
df_CEMP %>% janitor::get_dupes(site:rating) # 1 duplicated row

# Retain only distinct rows (remove duplicates)
df_CEMP <- distinct(df_CEMP)
```

Process Sonde Data

Calculate sonde rating intervals and summarize.

```r
# Get ratings to sort in order good -> bad
rating.order <- c("Excellent","Good","Fair","Poor","MAL")
```

```r
df_CEMP$rating <- factor(as.character(df_CEMP$rating),
                         levels = rating.order)

df_CEMP_w <- pivot_wider(df_CEMP, names_from = analyte_name, values_from = rating)

# Sort samples by date (early to late)
# Add previous exchange data plus one day (to prevent overlap in time series
# figures) as end point for range of rating data

df_CEMP_w <- df_CEMP_w %>%
  group_by(site) %>%
  arrange(visit_date) %>%
  mutate(start_date = lag(visit_date) + days(1), .before = visit_date) %>%
  ungroup()

df_CEMP_w <- df_CEMP_w %>%
  mutate(start_date = if_else(is.na(start_date), ymd("2023-01-01"), start_date))

# Calculate number of days between sonde exchanges
df_CEMP_w <- df_CEMP_w %>%
  mutate(days_rated = as.integer(visit_date - start_date))
```

Expand rating data from intervals to daily

Only data rating intervals are given but we need a daily value to properly graph the results. Also remove turbidity data because it is not useful to apply Wagner method to this data.

```r
# Create tibble with exchange intervals for each station
df_CEMP_l <- pivot_longer(df_CEMP_w,
                          names_to = "Parameter",
                          values_to = "Rating",
                          cols = where(is.factor))

df_CEMP_l <- df_CEMP_l %>% select(-days_rated)

# Fill in ratings data for each individual day
df_CEMP_c <- df_CEMP_l %>%
  mutate(across(ends_with("_date"), ymd)) %>%
  mutate(Date = map2(start_date,
                     visit_date, ~ seq.Date(.x, .y, by = "1 day")),
         .before = start_date) %>%
  unnest(Date) %>%
  select(!ends_with("_date"))

# Convert date formats and add Julian Date
df_CEMP_c <- df_CEMP_c %>%
  mutate(Julian = yday(Date)) %>%
  ungroup()

# Remove turbidity data and data after 1/1/2024
df_CEMP_c <- df_CEMP_c %>%
  filter(Parameter != "Turbidity") %>%
  filter(Date <= "2024-01-01")
```

Results
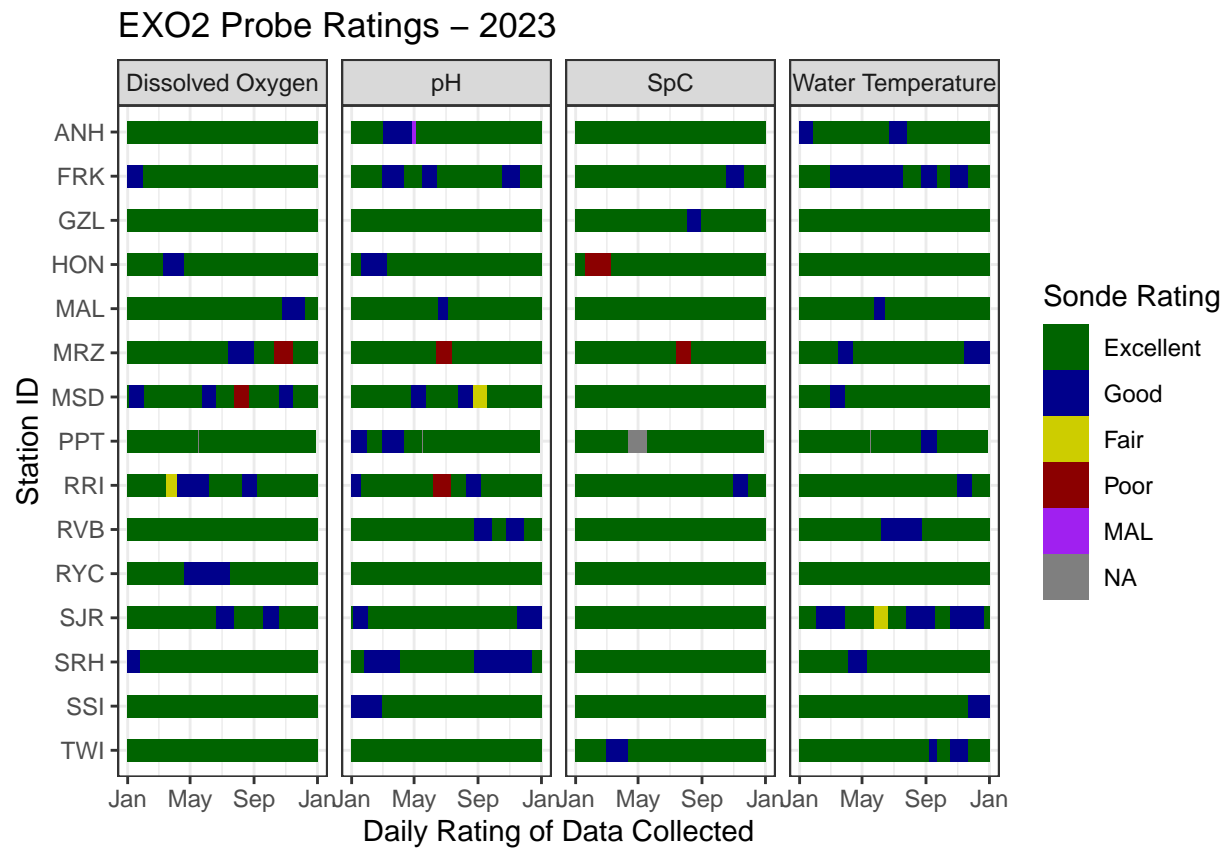
Plot sonde rating for all stations by probe.

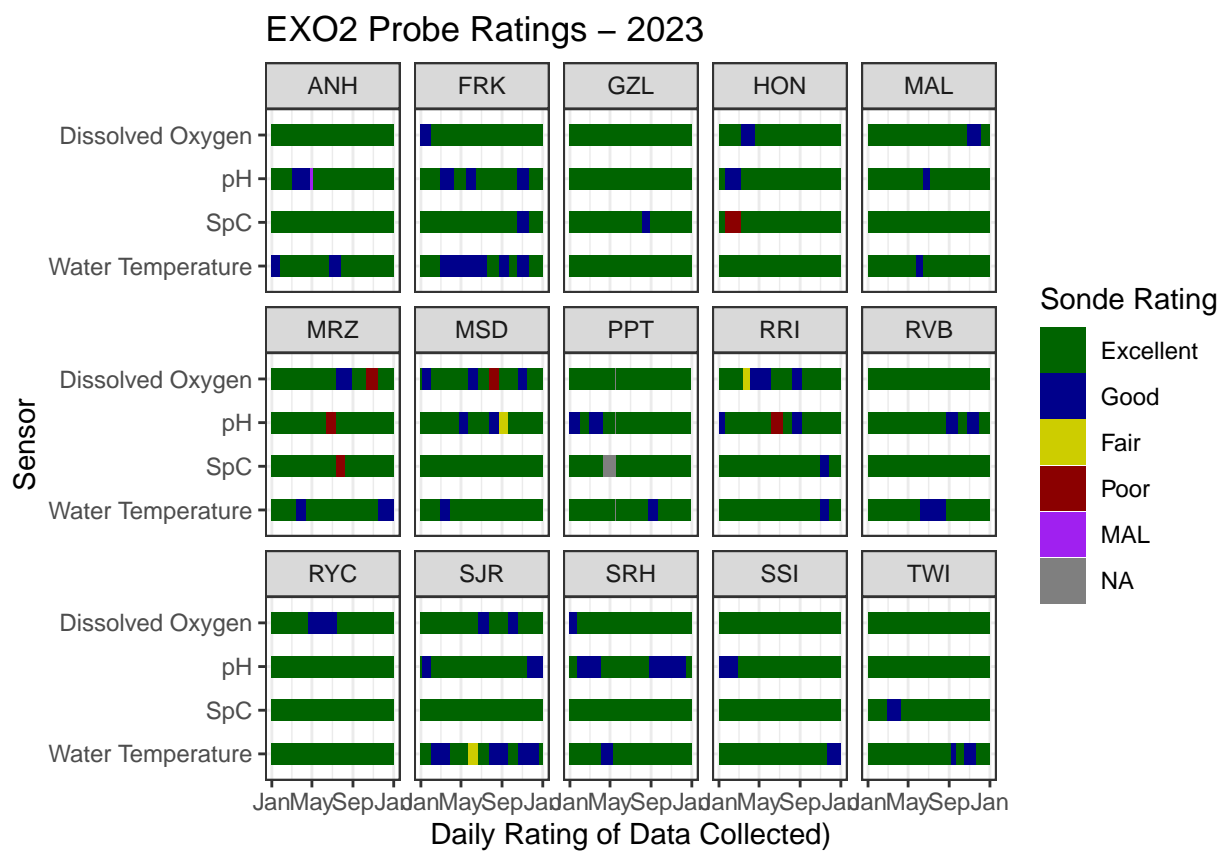Figure 1: C-EMP Sonde Ratings.

Plot sonde rating for all probes at each station.

Figure 2: C-EMP Sonde Ratings for each station.