ECOL 610: Analysis of NEON Data - Part 1

Group - Santa Rita Experimental Range (SRER)

27 September, 2022

Contents

Group Members **1**

Setup **1**

Introduction **2**

Load Data **2**

Assignment **4**

- Carbon fluxes: Plots of GPP, Re, and NEE over time . . . 4
  - Daily . . . 5
  - Smoothed Daily . . . 6
  - Evaluation . . . 8
- Environmental variables . . . 8
  - Evaluation . . . 11
- Plots of C fluxes vs. Environmental variables . . . 11
  - Evaluation . . . 14
- Scatter Plot Matrices (SPLOMS) . . . 15

Group Members

- Emily Swartz
- Shahriar Shah Heydari
- Stephanie Cardinalli
- George Woolsey

Setup

First, load in the needed packages. Install the packages if needed.

```r
library(tidyverse)
library(lubridate)
library(viridis)
library(RColorBrewer)
library(scales)
library(latex2exp)
library(psych)
remove(list=ls())
# set year range
min_yr <- 2020
max_yr <- 2021
# what is your site name?
site <- "Santa Rita Experimental Range"
```

Introduction

Net ecosystem exchange (NEE) is defined, by convention, as CO_2 flux from the ecosystem to the atmosphere. It corresponds to a negative carbon input to ecosystems. NEE is defined in this way because atmospheric scientists, who originated the term, seek to document net sources of CO_2 to the atmosphere (i.e., NEE) that account for rising atmospheric CO_2 concentration. Therefore, CO_2 input to the ecosystem is a negative NEE.

Chapin, F. S., Matson, P. A., Mooney, H. A., & Vitousek, P. M. (2002). Principles of terrestrial ecosystem ecology. p.208

Key point: CO_2 **input to the ecosystem is a negative NEE**

$$NEE = R_E - GPP$$

$$GPP = R_E - NEE$$

$$NEP = GPP - R_E$$

Load Data

Information about these variables can be found in the metadata file

```r
# create title date range
if(min_yr==max_yr){
  yr_lab <- paste0(min_yr)
}else{
  yr_lab <- paste0(min_yr, "-", max_yr)
}

# load 30 min data
  # I put both CPER and SRER data in this directory
  # all data will be loaded in the same R data set
f_list <- list.files(path = "../data/", pattern="*30 min.csv")
for (i in 1:length(f_list)){
```

```
  nm <- stringr::word(f_list[i], start = 1, sep = "30 min") %>%
    stringr::str_replace_all(pattern = "[[:punct:]]", replacement = "") %>%
    stringr::str_trim() %>%
    stringr::str_squish()
  temp <- read.csv(paste0("../data/", f_list[i])) %>%
    dplyr::mutate(neon_site_name = nm) %>%
    dplyr::relocate(neon_site_name) %>%
    dplyr::rename_with(~ tolower(
      gsub(" ", "_",
         str_trim(gsub("\\s+", " ", .x))
      )
    ))
  if(i==1){
    dta_30min <- temp
  }else{
    dta_30min <- dplyr::union_all(dta_30min, temp)
  }
  remove(temp)
}
# load daily data
f_list <- list.files(path = "../data/", pattern="*daily.csv")
for (i in 1:length(f_list)){
  nm <- stringr::word(f_list[i], start = 1, sep = "daily") %>%
    stringr::str_replace_all(pattern = "[[:punct:]]", replacement = "") %>%
    stringr::str_trim() %>%
    stringr::str_squish()
  temp <- read.csv(paste0("../data/", f_list[i])) %>%
    dplyr::mutate(neon_site_name = nm) %>%
    dplyr::relocate(neon_site_name) %>%
    dplyr::rename_with(~ tolower(
      gsub(" ", "_",
         str_trim(gsub("\\s+", " ", .x))
      )
    ))
  if(i==1){
    dta_1day <- temp
  }else{
    dta_1day <- dplyr::union_all(dta_1day, temp)
  }
  remove(temp)
}
# create dates and record counts
# 1-day
dta_1day <- dta_1day %>%
  dplyr::mutate(
    date_id = lubridate::make_date(year = year, month = month, day = day)
    , week = lubridate::week(date_id)
    , has_gpp = ifelse(!is.na(gpp), 1, 0)
    , season =
        dplyr::case_when(
          month %in% c(1:3, 12) ~ "Winter"
          , month %in% c(4:5) ~ "Spring"
          , month %in% c(6:8) ~ "Summer"
```

```
        , month %in% c(9:11) ~ "Autumn"
        , TRUE  ~ "Other")
  ) %>%
  dplyr::group_by(neon_site_name, week, year) %>%
  dplyr::mutate(is_full_week = sum(has_gpp)==7) %>%
  dplyr::ungroup()
# 30-min
dta_30min <- dta_30min %>%
  dplyr::mutate(
    date_id = lubridate::make_date(year = year, month = month, day = day)
    , time_id = lubridate::make_datetime(year = year, month = month, day = day
                        , hour = floor(hour)
                        , min = (hour-floor(hour))*60
                )
    , week = lubridate::week(date_id)
    , has_gpp = ifelse(!is.na(gpp), 1, 0)
    , season =
        dplyr::case_when(
          month %in% c(1:3, 12) ~ "Winter"
          , month %in% c(4:5) ~ "Spring"
          , month %in% c(6:8) ~ "Summer"
          , month %in% c(9:11) ~ "Autumn"
          , TRUE  ~ "Other")
  ) %>%
  dplyr::group_by(neon_site_name, week, year) %>%
  dplyr::mutate(is_full_week = sum(has_gpp)==24*2*7) %>%
  dplyr::ungroup()
```

Assignment

For one year of data (your choice!) for CPER and your site please create the following plots and use them to answer the below questions:

- Carbon fluxes: Plots of GPP, Re, and NEE over time
- Environmental variables: Plots of air temperature, soil temperature, soil moisture and PAR/PPFD over time
- Plots of GPP, Re, NEE vs. Environmental variables and color by day of year (ask Katie or your group mates if you need help!)

Then answer the following questions about each graph:

Identify outliers (make sure data look ok) Describe patterns and hypothesize what is driving them.

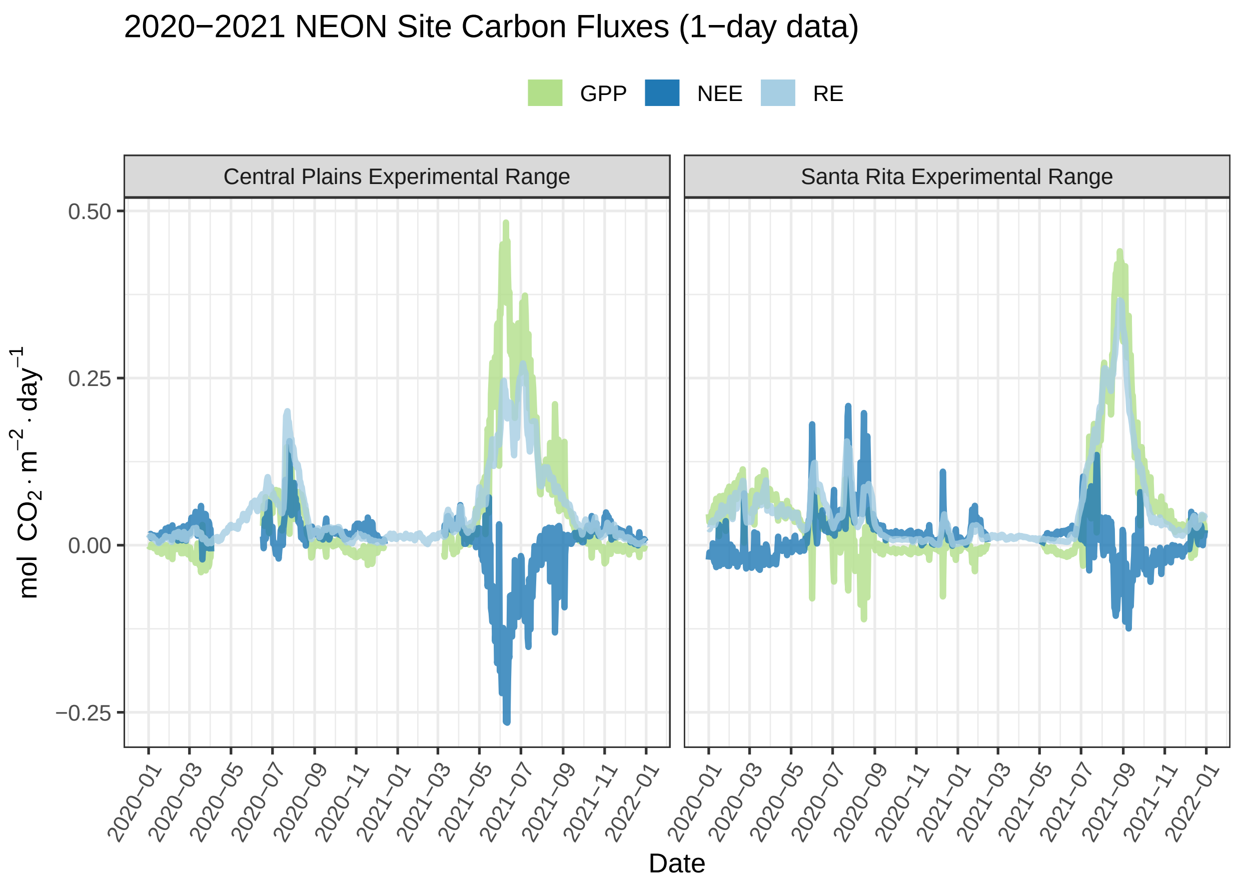
Submit your code, plots, and the answers to these questions in an RMarkdown PDF. As always, let us know if you have questions. This is due Tuesday, September 27th at 9:30am (class start).

Carbon fluxes: Plots of GPP, Re, and NEE over time

Compare annual time-trend of GPP, NEE, R_E for sites:

Daily

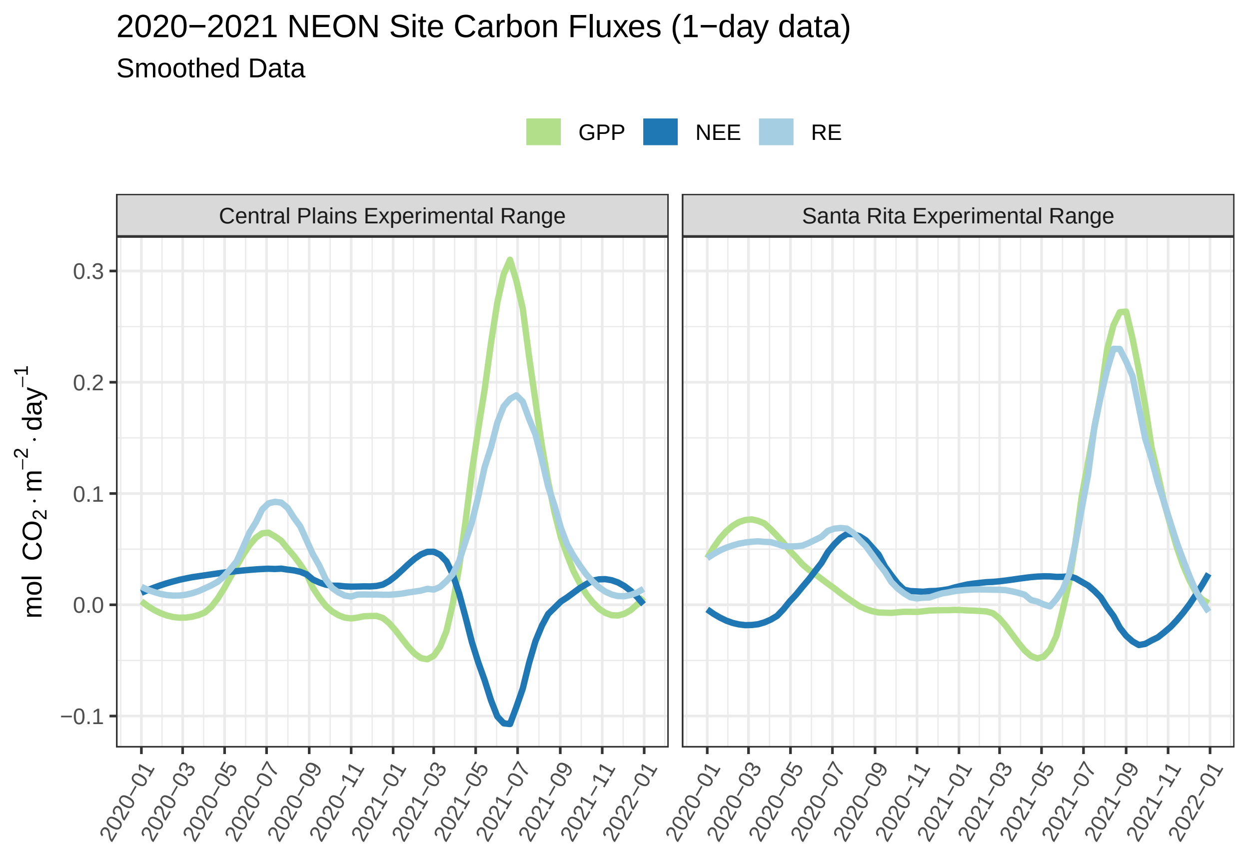
```r
# plot daily gpp, nee, re
dta_1day %>%
  dplyr::filter(
    # neon_site_name == site
    year %in% c(min_yr:max_yr)
  ) %>%
ggplot(., aes(x = date_id)) +
  geom_line(
    aes(y = gpp, color = "GPP")
    , lwd = 1.2
    , alpha = 0.8
  ) +
  geom_line(
    aes(y = nee, color = "NEE")
    , lwd = 1.2
    , alpha = 0.8
  ) +
  geom_line(
    aes(y = re, color = "RE")
    , lwd = 1.2
    , alpha = 0.8
  ) +
  facet_wrap(~neon_site_name) +
  scale_color_brewer(type = "qual", palette = "Paired", direction = -1) +
  scale_x_date(date_breaks = "2 month", date_labels = "%Y-%m") +
  xlab("Date") +
  ylab(latex2exp::TeX("$mol \\; CO_{2} \\cdot m^{-2} \\cdot day^{-1}$")) +
  labs(
    title = paste0(yr_lab, " NEON Site Carbon Fluxes (1-day data)")
  ) +
  theme_bw() +
  theme(
    legend.position = "top"
    , legend.direction = "horizontal"
    , legend.title = element_blank()
    , axis.text.x = element_text(angle = 60, hjust=1)
  ) +
  guides(color = guide_legend(override.aes = list(size = 5)))
```



Smoothed Daily

```
# plot daily gpp, nee, re
dta_1day %>%
  dplyr::filter(
    # neon_site_name == site
    year %in% c(min_yr:max_yr)
  ) %>%
ggplot(., aes(x = date_id)) +
  geom_smooth(
    aes(y = gpp, color = "GPP")
    , method = "loess"
    , span = 0.3
    , se = FALSE
    , lwd = 1.2
    , alpha = 0.8
  ) +
  geom_smooth(
    aes(y = nee, color = "NEE")
    , method = "loess"
    , span = 0.3
    , se = FALSE
    , lwd = 1.2
```

```
  , alpha = 0.8
) +
geom_smooth(
  aes(y = re, color = "RE")
  , method = "loess"
  , span = 0.3
  , se = FALSE
  , lwd = 1.2
  , alpha = 0.8
) +
facet_wrap(~neon_site_name) +
scale_color_brewer(type = "qual", palette = "Paired", direction = -1) +
scale_x_date(date_breaks = "2 month", date_labels = "%Y-%m") +
xlab("") +
ylab(latex2exp::TeX("$mol \\; CO_{2} \\cdot m^{-2} \\cdot day^{-1}$")) +
labs(
  title = paste0(yr_lab, " NEON Site Carbon Fluxes (1-day data)")
  , subtitle = "Smoothed Data"
) +
theme_bw() +
theme(
  legend.position = "top"
  , legend.direction = "horizontal"
  , legend.title = element_blank()
  , axis.text.x = element_text(angle = 60, hjust=1)
) +
guides(color = guide_legend(override.aes = list(size = 5)))
```



Evaluation

There was a wider range of variability in the daily carbon fluxes in 2021 for both CPER and SRER than in 2020. The variability in the carbon flux measurements at these sites in 2021 was largely driven by a sharp decline in NEE (i.e. increased CO_2 input to the ecosystem) in the summer and autumn relative to preceding levels. It is likely that the increased productivity during this time period was driven by increased water availability at these water-limited monitoring sites. Unfortunately, precipitation and soil water content data is not available for these sites during this same time period (see environmental section below).

Environmental variables

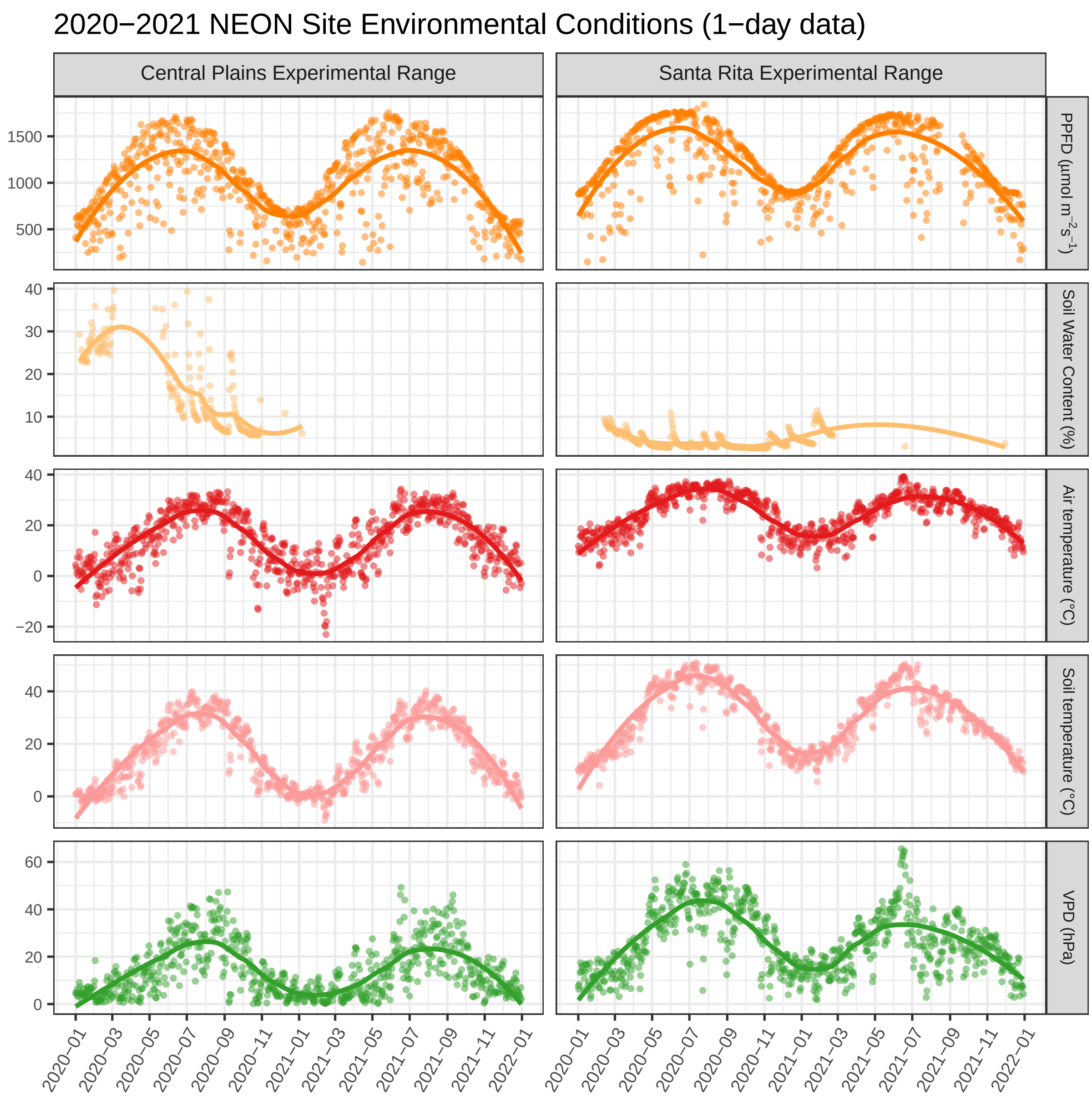
Plots of air temperature, soil temperature, soil moisture and PAR/PPFD over time

```
env_vars <- c("swc", "ta", "ts", "ppfd_in", "vpd")

# filter data
    dta_1day %>%
      dplyr::filter(
        year %in% c(min_yr:max_yr)
      ) %>%
      dplyr::select(
        neon_site_name
        , date_id
```

```
    , tidyselect::all_of(env_vars)
  ) %>%
# pivot dependent vars
  tidyr::pivot_longer(
    cols = tidyselect::all_of(env_vars)
    , names_to = "var_name"
    , values_to = "var_value"
    , values_drop_na = FALSE
  ) %>%
  dplyr::mutate(
    var_name_long = factor(var_name, labels = c(
      expression(paste("PPFD (", mu, "mol ", m^{-2}, s^{-1}, ")"))
      , expression(paste("Soil Water Content (%)"))
      , expression(paste("Air temperature (\u00B0C)"))
      , expression(paste("Soil temperature (\u00B0C)"))
      , expression(paste("VPD (hPa)"))
      )
    )
    , site_lab = factor(neon_site_name, labels = c(
        expression(paste("Central Plains Experimental Range"))
        , expression(paste("Santa Rita Experimental Range"))
      )
    )
  ) %>%
  dplyr::arrange(
    neon_site_name
    , date_id
    , var_name
  ) %>%
# plot
  ggplot(
    data = . # "." means the data that is passed to ggplot via the pipe "%>%"
    , mapping = aes(
      x = date_id
      , y = var_value
      , color = var_name_long
      , group = neon_site_name
    )
  ) +
  geom_point(
    alpha = 0.5
    , size = 0.8
  ) +
  geom_smooth(
    method = "loess"
    , se = FALSE
    , span = 0.5
    , lwd = 1
    , alpha = 0.8
  ) +
  facet_grid(var_name_long~site_lab, scales = "free_y", labeller = label_parsed) +
  scale_color_manual(values = RColorBrewer::brewer.pal(n = 8, "Paired")[8:4]) +
  scale_x_date(date_breaks = "2 month", date_labels = "%Y-%m") +
```

```
  xlab("") +
  ylab("") +
  labs(
    title = paste0(yr_lab, " NEON Site Environmental Conditions (1-day data)")
    # , subtitle = "Smoothed Data"
  ) +
  theme_bw() +
  theme(
    legend.position = "none"
    , axis.text.x = element_text(angle = 60, hjust=1, size = 8)
    , axis.text.y = element_text(size = 7)
    , strip.text.y = element_text(size = 7)
  )
```



Evaluation

There is a strong seasonal trend at both sites in the envirnomental variables for which data is available over the full time-period. Air, soil temperature, Photosynthetic Photon Flux Density (PPFD), and vapor pressure deficit (VPD) all have maximums in the summer (i.e. June-August) and minimums in the winter (i.e. December-March). Unfortunatley, data relating to water-availability in these water-limited ecosystems is not avaialbe over the full time-period. Air temperature and soil temperature exhibit much less within-season variability compared to PPFD and VPD.

Plots of C fluxes vs. Environmental variables

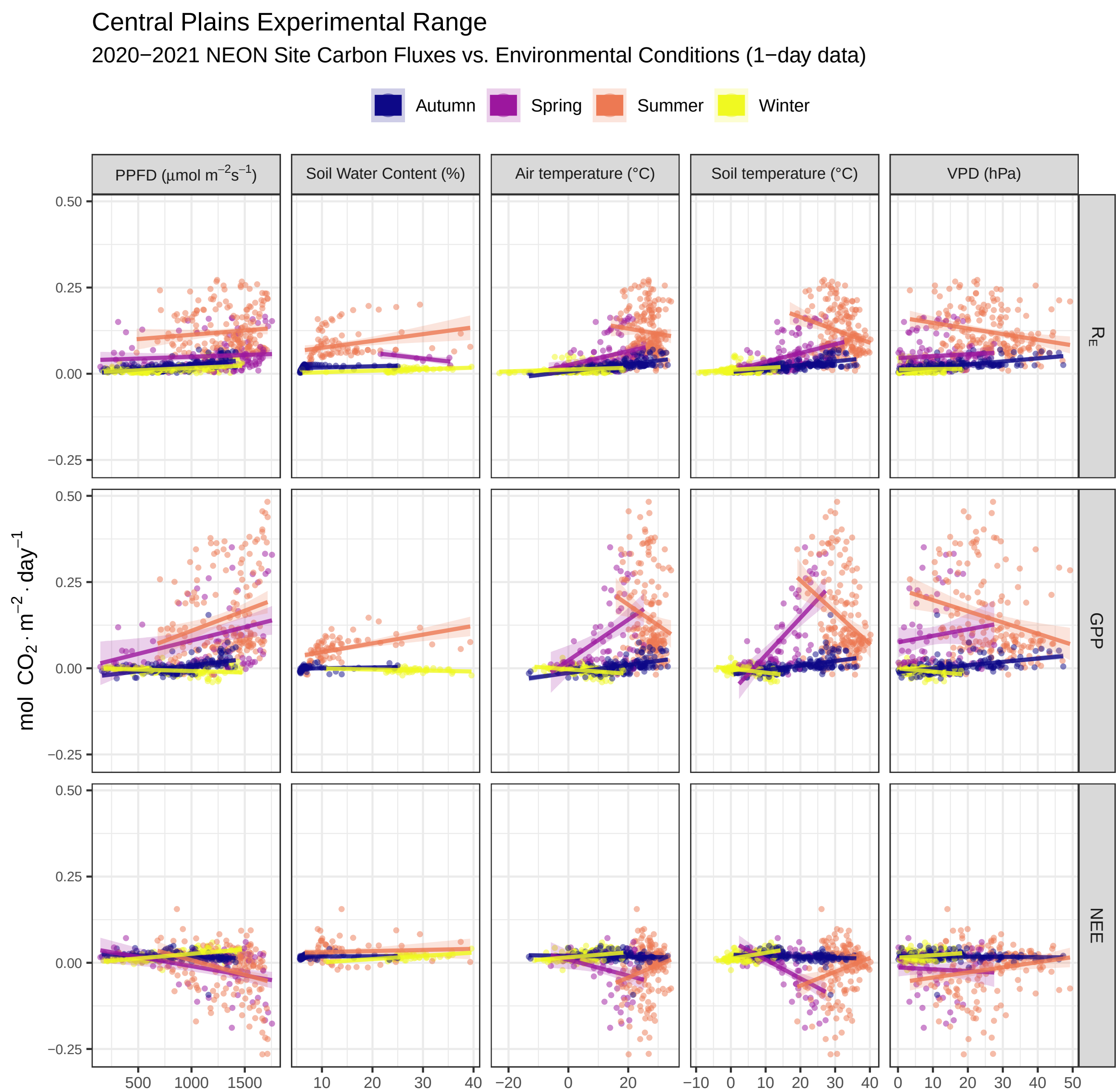
Plots of GPP, Re, NEE vs. Environmental variables and color by ~~day of year~~ **season**

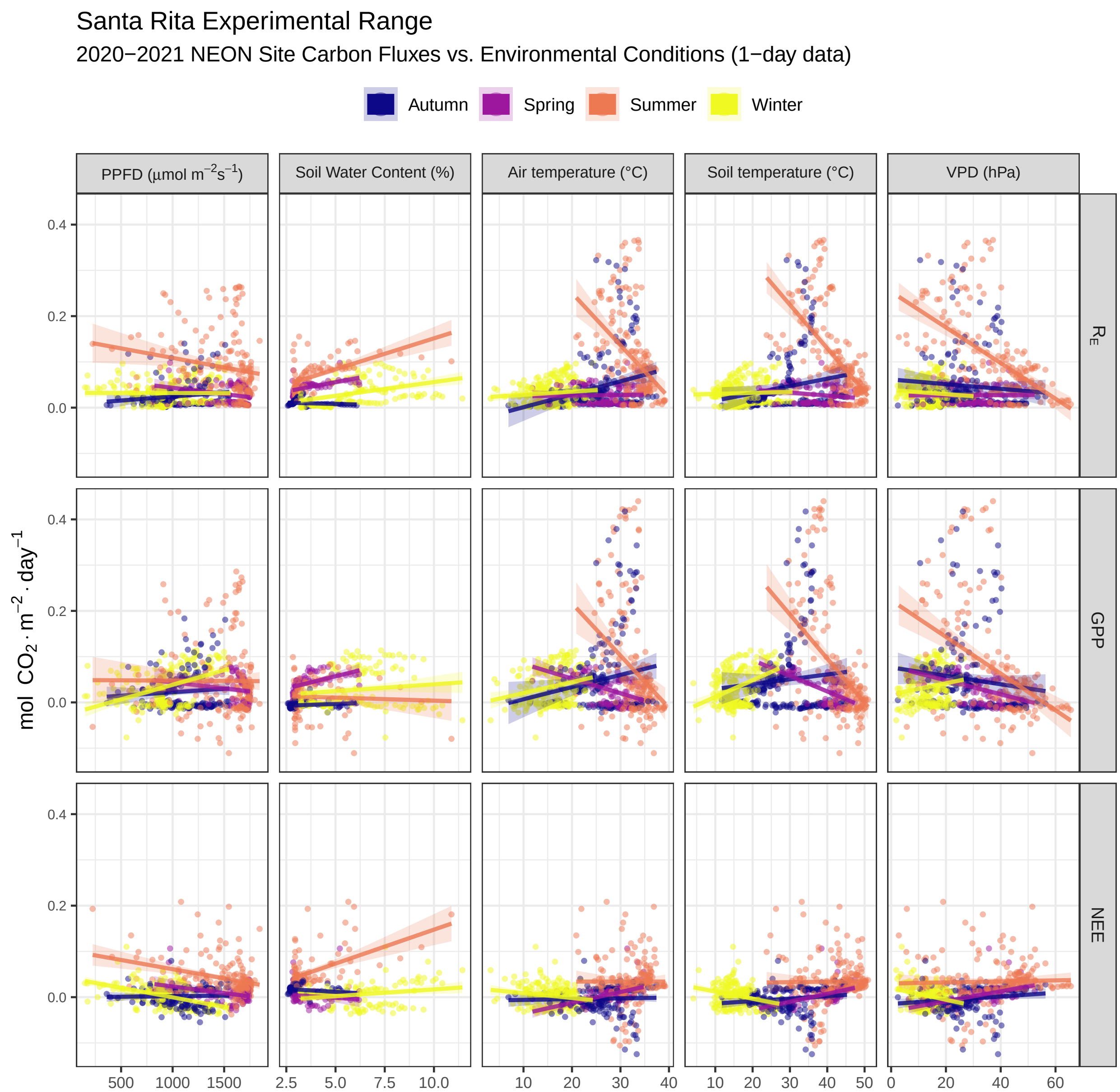
```
env_vars <- c("swc", "ta", "ts", "ppfd_in", "vpd")
dep_vars <- c("nee", "re", "gpp")
plt_flx_env <- function(my_site){(
# filter data
    dta_1day %>%
      dplyr::filter(
        year %in% c(min_yr:max_yr)
        , neon_site_name %in% my_site
      ) %>%
      dplyr::select(
        neon_site_name
        , date_id
        , season
        , tidyselect::all_of(dep_vars)
        , tidyselect::all_of(env_vars)
      ) %>%
  # pivot dependent vars
    tidyr::pivot_longer(
      cols = tidyselect::all_of(dep_vars)
      , names_to = "dep_var_name"
      , values_to = "dep_var_value"
      , values_drop_na = FALSE
    ) %>%
  # pivot independent vars
    tidyr::pivot_longer(
      cols = tidyselect::all_of(env_vars)
      , names_to = "var_name"
      , values_to = "var_value"
      , values_drop_na = FALSE
    ) %>%
    dplyr::mutate(
      var_name_long = factor(var_name, labels = c(
        expression(paste("PPFD (", mu, "mol ", m^{-2}, s^{-1}, ")"))
        , expression(paste("Soil Water Content (%)"))
        , expression(paste("Air temperature (\u00B0C)"))
        , expression(paste("Soil temperature (\u00B0C)"))
        , expression(paste("VPD (hPa)"))
        )
      )
      , dep_var_lab = ordered(dep_var_name, levels = c("re", "gpp", "nee"), labels = c(
```

```r
          expression(paste(R[E]))
          , expression(paste("GPP"))
          , expression(paste("NEE"))
        )
      )
    ) %>%
    dplyr::arrange(
      neon_site_name
      , date_id
      , dep_var_name
      , var_name
    ) %>%
  # plot
    ggplot(
      data = . # "." means the data that is passed to ggplot via the pipe "%>%"
      , mapping = aes(
        x = var_value
        , y = dep_var_value
        , color = season
        , fill = season
        # , group = season
      )
    ) +
    geom_point(
      alpha = 0.5
      , size = 0.8
    ) +
    geom_smooth(
      method = "lm"
      , se = TRUE
      # , span = 3
      , lwd = 1
      , alpha = 0.2
    ) +
    facet_grid(dep_var_lab~var_name_long, scales = "free_x", labeller = label_parsed) +
    scale_color_viridis_d(option = "plasma", alpha = 0.8) +
    scale_fill_viridis_d(option = "plasma") +
    # scale_x_date(date_breaks = "2 month", date_labels = "%Y-%m") +
    xlab("") +
    ylab(latex2exp::TeX("$mol \\; CO_{2} \\cdot m^{-2} \\cdot day^{-1}$")) +
    labs(
      title = my_site
      , subtitle = paste0(yr_lab, " NEON Site Carbon Fluxes vs. Environmental Conditions (1-day data)")
    ) +
    theme_bw() +
    theme(
      legend.position = "top"
      , legend.direction = "horizontal"
      , legend.title = element_blank()
      , axis.text.x = element_text(size = 8)
      , axis.text.y = element_text(size = 7)
      , strip.text.x = element_text(size = 8)
      , strip.text.y = element_text(size = 9)
```

```
    ) +
    guides(color = guide_legend(override.aes = list(size = 5)))
)}

unique(dta_1day$neon_site_name) %>%
  purrr::map(plt_flx_env)
```





Evaluation

I personally found the coloring of the points by day of year to be difficult to interpret and instead colored the points by season so that high and low days of the year (i.e. winter) are colored the same.

At CPER, the strongest NEE response to variability in the environmental factors considered was with PPFD in the summer season. There was limited correlation between NEE and PPFD in the autumn and winter but in the summer and spring, there was a significant negative relationship between PPFD and NEE. During these warmer seasons, as PPFD increased, NEE decreased due to increasing GPP and relatively flat R_E. There was also a strong correlation between NEE and soil temperature at CPER during the spring and summer seasons. During the spring, there was a negative relationship between NEE and soil temperature but during the summer this relationship turned positive. That is, as soil temperature increased during the spring, NEE decreased (i.e. increased CO_2 input to the ecosystem) but in the summer, increasing soil temperature was associated with increased NEE (a threshold response pattern).

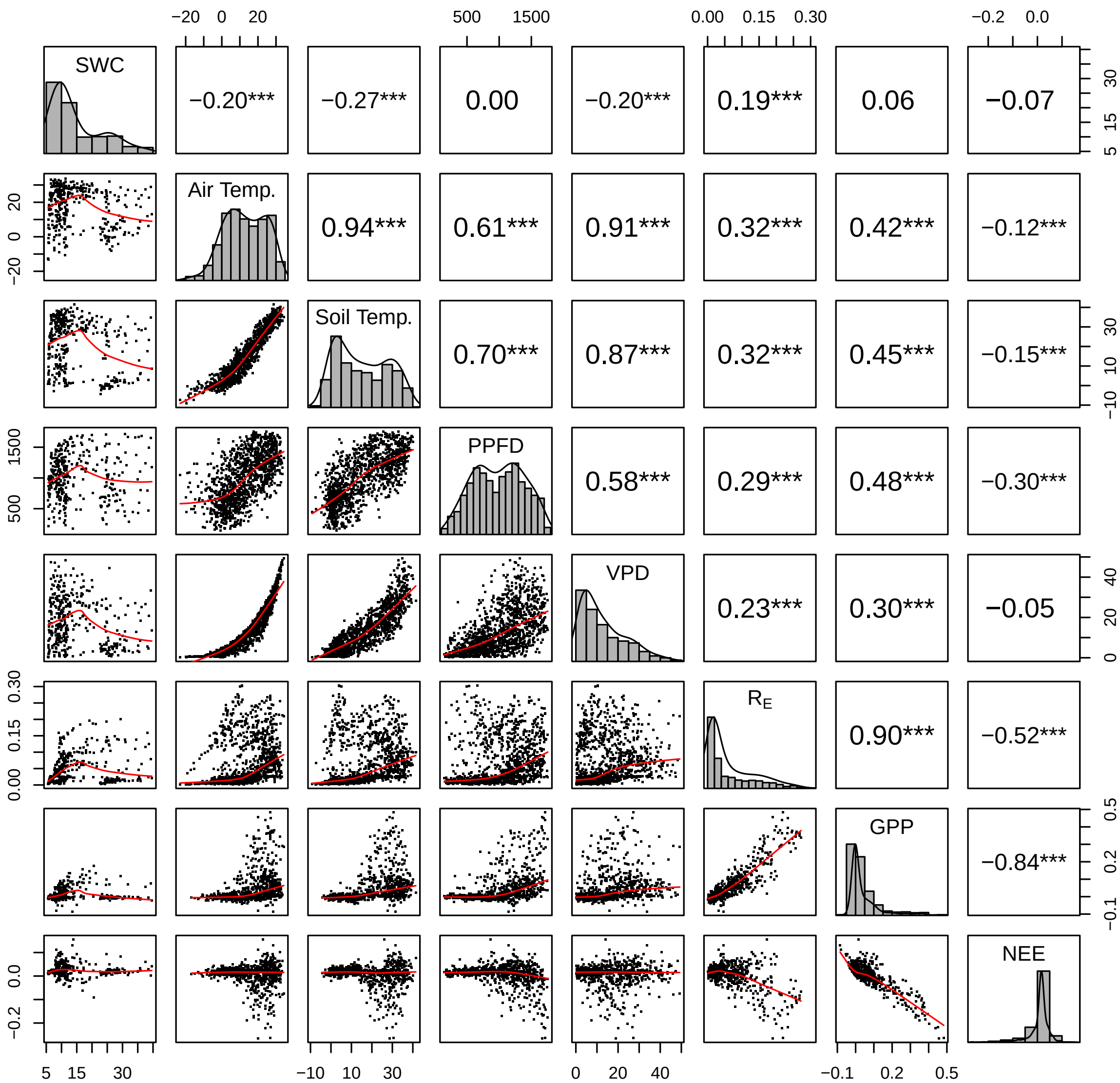
At SRER, the strongest NEE response to variability in the environmental factors considered was with soil water content in the summer season. This relationship was positive, meaning that as soil water content increased in the summer, NEE also increased (i.e. decreased CO_2 input to the ecosystem). This is a non-intuitive result and may be driven by a few outlier points with high soil moisture content readings during the summer season. There were significant negative relationships between NEE and the environmental variables PPFD, air temperature, soil temperature, and VPD in the winter at SRER. Increases in PPFD, air temperature, soil temperature, and VPD during the winter resulted in decreased NEE (i.e. increased CO_2 input to the ecosystem). In the warmer months of summer and spring, this relationship was either reversed or did not exhibit any pattern. In the summer at SRER, there was high sensitivity of plant productivity (GPP) to increasing temperatures. As temperatures increased at SRER during the summer, plant productivity (GPP) decreased steeply. However, this decreasing production was offset by comparable decreasing respiration (R_E) so that the overall impact on NEE was negligible.

Scatter Plot Matrices (SPLOMS)

```
my_corr_plot_fn <- function(my_site) {
  (
    psych::pairs.panels(
        dta_1day %>%
        dplyr::filter(neon_site_name == my_site) %>%
        dplyr::select(
          swc, ta, ts, ppfd_in
          , vpd
          , re
          , gpp
          , nee
        )
      , pch = "."  # 21 for color by group in line above # "." for no color
      , labels = c(
          "SWC"
          , "Air Temp."
          , "Soil Temp."
          , "PPFD"
          , "VPD"
          , latex2exp::TeX("$R_E$")
          , "GPP"
          , "NEE"
      )
      , method = "pearson" # correlation method
      , hist.col = "gray70"
      , density = TRUE  # show density plots
      , ellipses = FALSE # show correlation ellipses
      , rug = FALSE
      , stars = TRUE
      , main = my_site
    )
  )
}

corr_plts <- unique(dta_1day$neon_site_name) %>%
  purrr::map(my_corr_plot_fn)
```

Central Plains Experimental Range



Santa Rita Experimental Range