Draft: Carbon respiration within the blue oak savanna ecosystem

Nate Dailey Shivam Pawar

2024-12-19

Abstract

This study investigates the relationship between the explanatory variables of soil water content (SWC) and Temperature with the response variable Ecosystem Respiration (RECO) within the blue oak savanna ecosystem of Tonzi Ranch, California, spanning 2000-2021. Data were sourced from an AmeriFlux tower, capturing a variety of weather variables. After aggregating the observations to monthly means, and shifting RECO back 2 months temporally, we found a linear relationship between RECO and soil water content with a bivariate model, and a statistically significant relationship between RECO and a combination of SWC and temperature.

Keywords: Precipitation, soil moisture, ecosystem respiration, blue oak savanna, AmeriFlux, drought, soil

Introduction

Understanding the relationship between rainfall and ecosystem processes is critical in the face of rising climatic variability, particularly in drought-prone regions. Precipitation serves as a primary driver of water availability, which directly influences plant growth and the overall carbon balance within ecosystems. In Mediterranean climates, such as our study area, prolonged droughts and shifting precipitation patterns pose significant challenges to ecosystem resilience. These changes have far-reaching implications for carbon cycling and the sustainability of ecosystems **(Young et al, 2020)**.

This study investigates the blue oak savanna ecosystem at Tonzi Ranch, southeast of Sacramento (coordinates: 38.4309, -120.9660), with a focus on the influence of soil water content (SWC) on ecosystem respiration (RECO). RECO is an important ecological variable, as it reflects the exchange of carbon between the ecosystem and the atmosphere, providing insights into plant and soil interactions.

Our research utilizes 21 years of data from the AmeriFlux tower (2000 to 2021) at Tonzi Ranch, California. Using this data, this study aims to uncover the relationship between SWC and RECO, with an emphasis on the seasonal variability and lagged effects of precipitation. This could inform strategies for drought resilience in similar ecosystems, such as Chico's own blue oak savanna ecosystem.

Figure 1: Google Street View of Tonzi Ranch

Background

California has undergone an intense drought over the last 20 years–particularly from 2012 to 2016–which has been called one of the most consequential droughts in the last century. This has caused deep soil drying and extreme moisture stress in trees throughout California, resulting in frequent tree mortality **(Dwomoh et al. 2021)**. Understanding the impacts of drought on ecosystem processes is critical, particularly in regions like Tonzi Ranch, where water availability plays a central role in shaping ecological dynamics.

Climate models predict that ecosystems will experience increased precipitation variability in the coming decades, characterized by more extreme precipitation events separated by longer dry periods **(Wolters et al. 2000, Kharin et al. 2007)**. These changes in precipitation patterns can significantly impact ecosystem processes, particularly in water-limited environments where soil moisture dynamics strongly influence biological activities **(Seastedt et al., 2008)**.

Studies have shown that grasslands (such as the blue oak savanna ecosystem of Tonzi Ranch) are particularly sensitive to precipitation variability, with changes in soil moisture affecting various components of the carbon cycle, including photosynthesis and soil respiration **(Knapp and Smith, 2001; Huxman et al. 2004)**. These responses can vary significantly between mesic and arid systems, highlighting the importance of site-specific studies in understanding ecosystem responses to water availability.

One key metric for studying ecosystem productivity under drought conditions is Ecosystem respiration (RECO), which measures the amount of carbon released by an ecosystem **(Luo and Zhou 2006)**. Precipitation plays a pivotal role in modulating RECO **(Baldocchi et al, 2018)**. However, precipitation alone does not fully capture the complexities of water availability. Due to factors such as soil evaporation, plant transpiration, runoff, and subsurface drainage, soil moisture provides important additional information (Berg and Sheffield, 2018).

Given these insights, we decided to incorporate soil water content as our primary explanatory variable in our analysis of the Tonzi Ranch ecosystem. By considering soil water content, we aim to better understand the interactions between water availability and carbon flux in this semi-arid environment, contributing to a more comprehensive perspective on ecosystem responses to climatic variability.

Study Design

Goal

The primary objective of this research is to evaluate the influence of soil water content (SWC) and temperature on ecosystem respiration (RECO) at Tonzi Ranch, California. Specifically, this study seeks to determine the extent to which RECO is correlated with SWC, both independently and in combination with temperature.

Hypothesis

We hypothesize that RECO exhibits a positive correlation with SWC and that a combination of SWC and temperature provides an enhanced explanatory framework for variations in RECO.

Data Collection Methods

Data utilized in this study was obtained from the AmeriFlux tower located at Tonzi Ranch, California (latitude: 38.4309, longitude: -120.9660). The AmeriFlux network comprises 109 towers across North and South America, and it provides open-source datasets for scientific research. The Tonzi Ranch tower is equipped with advanced sensors capable of measuring the carbon exchange between the ecosystem and the atmosphere. Through established mathematical modeling and processing, these measurements yield estimates of RECO.

The dataset used in this analysis comprises daily observations derived from the aggregation of higher-frequency measurements captured throughout the day. To reduce noise and highlight seasonal patterns, the daily observations were further aggregated into monthly averages. Data points containing missing values for RECO, SWC, or temperature were excluded from the analysis to ensure the integrity of the statistical models. The final dataset consisted of n = 252 monthly observations.

Variables

Our variables of interest include monthly averages of soil water content (%) and temperature (°C) as explanatory variables, and ecosystem respiration (grams of CO2 respired per square meter) as the response variable. All of these variables are quantitative. We also considered using precipitation as a variable in our analysis, but we chose SWC in its place due to its more direct relationship with plant water access **(Berg and Sheffield, 2018)** and the abundance of 0 mm observations in precipitation. We also tested year as a quantitative explanatory variable in our multivariate model. However, it was excluded due to its statistical insignificance in our model.

Variable Name	Type	Mean	Standard Deviation	Units
Ecosystem Respiration (RECO) - Adjusted	Q	2.506	1.332	Average grams of carbon per square meter per day $gC/m^2/day$).
Soil Water Content (SWC)	Q	21.438	10.300	Avg. % per day
Temperature	Q	17.692	6.004	Avg. Temp in °C

Table 1: Description of variables used in the dataset, including type, mean, standard deviation, and units. Q = Quantitative.

Data Preparation

Time Series Analysis of Variables

At the beginning of our analysis, we hypothesized a temporal difference between the peak of precipitation and RECO. Therefore, we used a time series analysis to investigate our variables (see figure 1 in results section).

Upon analyzing the monthly mean graphs for both precipitation and RECO, it became evident that the trends only align slightly–RECO appeared to lag behind SWC by about 2 months. This lag makes sense contextually, due to the time it takes for plants to respond to atmospheric conditions **(Berg and Sheffield, 2018)**. Therefore, we decided to shift RECO back by two months.

Upon analyzing the monthly mean graphs for both precipitation and RECO, it becomes evident that the trends only align slightly. Precipitation peaks in December and remains high through March, while RECO peaks in April, following the precipitation peak. This lag in RECO is logical, as spring offers optimal moisture and sunlight for plant growth and reproduction, leading to higher respiration rates. This suggests RECO may lag behind precipitation rather than exhibit a direct correlation. Therefore, we decided to shift RECO back by two months and create a new variable: RECO adjusted.

Statistical Analysis Methods

We used "Pearson's Product-Moment Correlation Test" to determine whether there is a statistically significant positive correlation between the time-adjusted version of RECO (RECO adjusted) and SWC. We also performed a multivariate model of RECO adjusted vs. a combination of SWC and Temperature.

Results

Time series analysis

The following plot displays our aforementioned time series exploratory analysis of RECO and SWC, which inspired us to to shift RECO back by two months and use this derived value in our regression models.

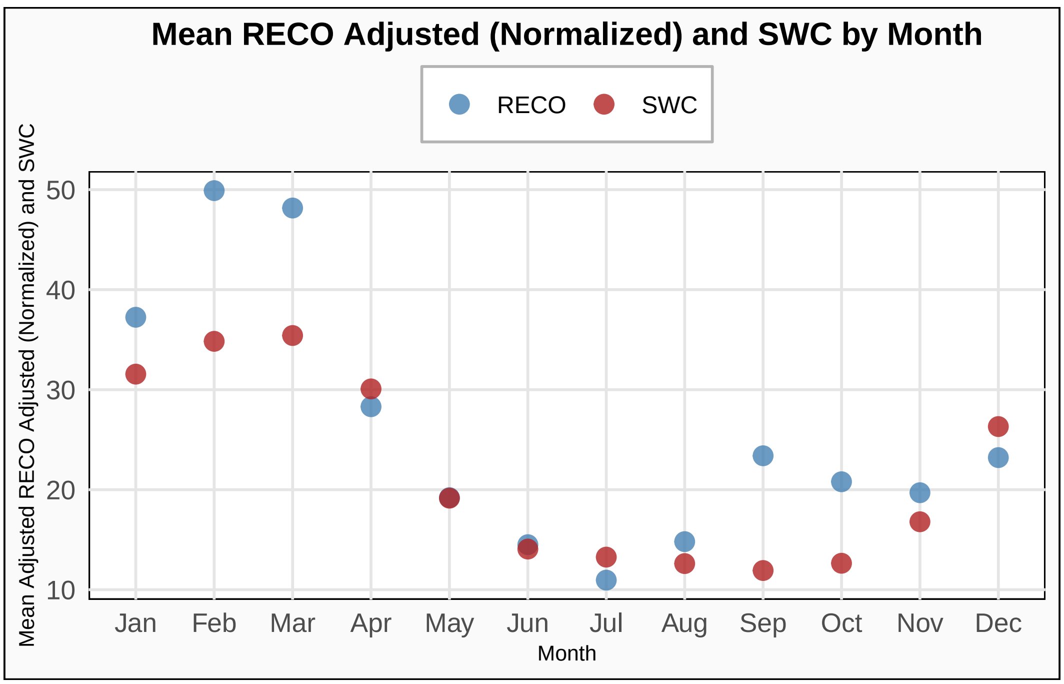


Figure 2: Monthly mean ecosystem respiration (RECO, blue) and precipitation (red) from 2000 to 2021. RECO is expressed in grams of carbon respired per square meter, and precipitation in millimeters. The figure shows seasonal trends, with higher RECO in fall, winter and spring, and increased SWC during winter and early spring months.

Ecosystem Respiration vs. Soil Water Content

Figure 3 utilizes our temporally shifted RECO variable. In this bivariate model, we found both explanatory variables to be statistically significant.

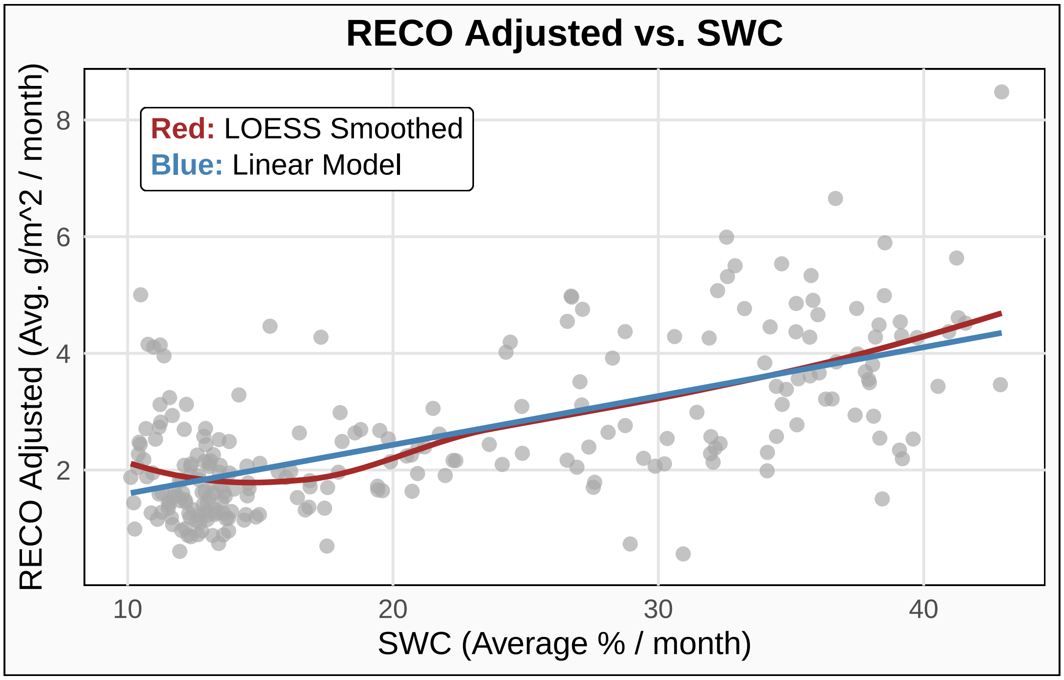


Figure 3: Scatter plot of RECO adjusted and SWC with a linear model and a LOESS Smoothed line overlaid. The variables show a strong positive correlation visually.

Measure	Value	95% Confidence Interval	P-Value
SWC	0.083	[0.071, 0.096]	$p < 0.001$
Correlation Coefficient	0.650	[0.570, 0.717]	$p < 0.001$

Table 2: Table of measures related to our RECO Adjusted vs. SWC linear regression model.

The table above displays our results for our RECO Adjusted vs. SWC linear regression model. SWC's slope coefficient represents the amount that RECO Adjusted is predicted to increase with a 1% increase in SWC.

Based on our model, the correlation coefficient is 0.650 (95% CI: 0.570, 0.717). Our Pearson's Correlation Test showed us that this correlation is statistically significant ($p < 0.001$). Without the temporal adjustment to RECO, the correlation coefficient is less, at 0.390 ($p < 0.001$). Overall, these results show greater correlation between SWC and RECO when a 2 month backward time-shift is applied.

Multivariate model of Ecosystem Respiration vs. Precipitation and Temperature

We developed a multivariable model to assess the combined impact of **temperature** and **SWC** on **RECO Adjusted**. Although temperature and SWC exhibit a statistically significant negative correlation ($p < 0.001$), both variables were included in the final model because they provide unique and valuable information about the environment.

Precipitation was also considered as a predictor; however, it was excluded because we found it to be statistically insignificant in the model. Suspecting that this might have been due to its collinearity with SWC, we also tried swapping out SWC for precipitation in the model, but this also performed worse. Ultimately, we landed on SWC and temperature as the best combination of predictors for RECO adjusted.

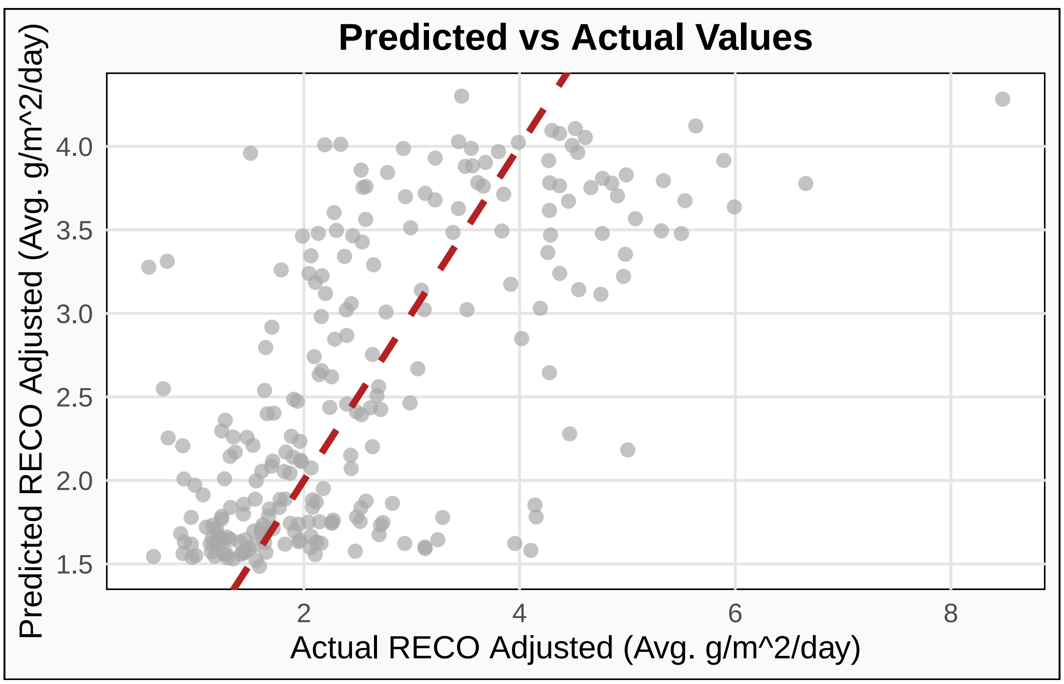


Figure 4: Actual vs. Predicted plot of multivariate model RECO Adjusted vs. Precipitation + Temperature.

The Figure 4 displays the actual vs. predicted values for RECO adjusted, serving as a reference for model performance. The dotted red line represents the scenario where the predicted RECO Adjusted equals the actual RECO Adjusted (i.e., y = x).

Measure	**Value**	**95% Confidence Interval**	**P-Value**
SWC	0.0642	[0.0463, 0.0822]	$p < 0.001$
Temperature	-0.0458	[-0.0766, -0.0150]	$p < 0.001$

Table 3: Measures related to our RECO Adjusted vs. SWC + Temperature multivariate linear regression model.

The table above shows the coefficient for SWC which indicates that for each unit increase in SWC, the predicted value of RECO Adjusted increases by 0.0642, holding temperature constant. This effect was statistically significant ($p < 0.001$), highlighting the strong positive relationship between SWC and RECO Adjusted. Conversely, the coefficient for temperature suggests that for each unit increase in temperature, the predicted value of RECO Adjusted decreases by 0.0458, holding SWC constant. This negative relationship was also statistically significant ($p = 0.0037$), demonstrating that temperature has a meaningful and inverse effect on RECO Adjusted.

Comparison of Models

Model	**AIC**	**BIC**	**R^2 or Multiple R^2**
Bivariate (RECO ~ SWC)	683.738	694.155	0.423
Multivariate (RECO ~ SWC + Temperature)	677.206	691.095	0.443

Table 4: A comparison of our simple linear regression and multivariate models, by the measures AIC, BIC, and R Squared.

We can extract several key insights by comparing the simple linear regression and multivariate models. The multivariate model has a lower AIC and BIC compared to the bivariate model. This indicates improved model fit and justifies the inclusion of additional predictors, despite the penalty for increased complexity. Additionally, the multiple R-squared value for the multivariate model (0.443) is slightly higher than the R-squared value of the simple model (0.423), suggesting a modest improvement in the proportion of variance explained by the model. Overall, these metrics indicate that the multivariate model provides a better balance between model fit and complexity, offering improved explanatory power for RECO.

Limitations

The study also has several limitations that should be acknowledged to better understand the findings and their implications. First, the results are generalizable only to ecosystems or conditions with characteristics similar to the study site, such as the blue oak woodland ecosystem in Chico. Caution is advised when applying these findings to ecosystems with different climatic, biotic, or abiotic conditions. Second, while the models used in this study were robust, some assumptions were not fully upheld. The explanatory variables showed slight deviations from normality, subtle inconsistencies in variance, and mild collinearity, which, although not strong enough to invalidate the models, should be considered when interpreting the results. Third, because this study is observational in nature, it describes associations rather than causation. For example, the positive relationship observed between SWC and RECO does not imply that increasing SWC directly causes an increase in RECO. Establishing causation would require experimental approaches, which are challenging in ecological studies due to the complexity of natural systems. Additionally, many factors not included in the models could influence RECO. These include factors such as vegetation type, soil type, nutrient availability, wind speed, humidity, and extreme weather events. Finally, the scope of the study is limited by its reliance on data collected over 21 years from a single site (Tonzi Ranch). While the dataset is extensive, it may not capture broader variability across other regions or ecosystems.

Discussion

Adjusting RECO for its temporal lag was a key component to examine the true impact of soil water content and temperature on carbon exchange in this ecosystem. This suggests that ecosystem respiration is not an immediate function of soil moisture but is also influenced by a time component.

This study offers valuable insights into how ecosystems like the blue oak savanna adapt to drought and increasing climatic variability. The blue oak savanna serves as a model for understanding the resilience of semi-arid ecosystems under the pressures of prolonged droughts and shifting precipitation patterns. Our findings underscore the critical role of timing and rainfall distribution in sustaining ecosystem functionality, as evidenced by the delayed response of RECO to SWC.

To build on this understanding, future research should focus on additional factors that may influence ecosystem respiration, such as the effects of extreme drought years. Namely, incorporating the year variable through more advanced time series analysis could reveal temporal trends and further enhance the model's ability to capture the effects of climatic variability. These efforts could play a significant role in enhancing our capacity to predict and manage the resilience of ecosystems in an era of increasing environmental uncertainty.

Conclusion

This study provides evidence of a significant relationship between ecosystem respiration (RECO) and SWC–and a significant relationship between RECO and a combination of SWC and temperature–within the blue oak savanna ecosystem at Tonzi Ranch, California. By utilizing data from an AmeriFlux tower, we demonstrated that RECO exhibits a lagged response to SWC, peaking approximately two months after rainfall events. This temporal delay highlights the complex interplay between ecosystem processes, where moisture availability and seasonal conditions align to optimize carbon flux dynamics.

The adjusted RECO variable, which accounts for this lag, reveals a stronger correlation with precipitation compared to the unadjusted RECO. This suggests that ecosystem respiration is not an immediate function of rainfall but is influenced by subsequent environmental factors. These findings contribute to a deeper understanding of how ecosystems like the blue oak savanna respond to climatic variability, particularly under the stress of prolonged drought conditions.

Bibliography

AmeriFlux. "About AmeriFlux Data." AmeriFlux Management Project, Lawrence Berkeley National Laboratory, https://ameriflux.lbl.gov/data/aboutdata/. Accessed 14 Dec. 2024.

Berg, Alexis, and Justin Sheffield. “Climate Change and Drought: The Soil Moisture Perspective.” Current Climate Change Reports, vol. 4, 2018, pp. 180–191. Springer, https://doi.org/10.1007/s40641-018-0095-0.

Dwomoh, Francis K., et al. “Hotter Drought Escalates Tree Cover Declines in Blue Oak Woodlands of California.” Frontiers in Climate, vol. 3, 2021, https://doi.org/10.3389/fclim.2021.689945.

Google. Google Street View of 38.4309, -120.9660. Google Maps, https://www.google.com/maps/place/38.4309,-120.9660,15z. Accessed 14 Dec. 2024.

Aiken, R. M., Nielsen, D. C., Vigil, M. F., & Ahuja, L. R. (2018). Field evaluation of energy balance simulation of surface soil temperatures. Agricultural and Forest Meteorology, 250–251, 318–329. https://doi.org/10.1016/j.agrformet.2017.11.004

Baldocchi, D., Chu, H., & Reichstein, M. (2018). Inter-annual variability of net and gross ecosystem carbon fluxes: A review. Agricultural and Forest Meteorology, 249, 520–533. https://doi.org/10.1016/j.agrformet.2017.05.015

Barbosa, C. C. R., Atkinson, P. M., & Dearing, J. A. (2015). Remote sensing of ecosystem services: A systematic review. Ecological Indicators, 52, 430–443. https://doi.org/10.1016/j.ecolind.2015.01.007

López, J., D.A. Way, and W. Sadok. “Systemic Effects of Rising Atmospheric Vapor Pressure Deficit on Plant Physiology and Productivity.” Global Change Biology, vol. 27, 2021, pp. 1704-1720, https://doi.org/10.1111/gcb.15548.

Luo, Yiqi, and Xuhui Zhou. “Introduction and Overview.” Soil Respiration and the Environment, edited by Yiqi Luo and Xuhui Zhou, Academic Press, 2006, pp. 3–15. ScienceDirect, https://doi.org/10.1016/B978-012088782-8/50001-2. Accessed 18 Dec. 2024.

Wolters V, Silver WL, Bignell DE et al. (2000) Effects of global changes on above- and belowground biodiversity in terrestrial ecosystems: implications for ecosystem functioning. BioScience, 50, 1089–1098. https://doi.org/10.1641/0006-3568(2000)050[1089:EOGCOA]2.0.CO;2

Kharin VV, Zwiers FW, Zhang X, Hegerl GC (2007) Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. Journal of Climate, 20, 1419–1443. https://doi.org/10.1175/JCLI4066.1.

Seastedt TR, Hobbs RJ, Suding KN (2008) Management of novel ecosystems: are novel approaches required? Frontiers in Ecology and the Environment, 6, 547–553. https://doi.org/10.1890/070046

Alan K. Knapp, Melinda D. Smith Variation Among Biomes in Temporal Dynamics of Aboveground Primary Production.Science 291,481-484(2001).DOI:10.1126/science.291.5503.481

Huxman, T.E., Cable, J.M., Ignace, D.D. et al. Response of net ecosystem gas exchange to a simulated precipitation pulse in a semi-arid grassland: the role of native versus non-native grasses and soil texture. Oecologia 141, 295–305 (2004). https://doi.org/10.1007/s00442-003-1389-y

Appendix

This following code describes our methodology, from cleaning our raw dataset to generating plots.

```
# Read in the data
raw <- read.delim(
  "../data/AMF_US-Ton_FLUXNET_SUBSET_DD_2001-2021_3-5.csv",
  sep = ',',
  na.strings = "-9999" # convert -9999 values to NA
  )

# Select desired variables
data_clean  <- raw %>% select(
  "TIMESTAMP",
  "TS_F_MDS_1",
  "SWC_F_MDS_2",
  "RECO_DT_VUT_REF",
  "P_F",
  )
```

Group by month / year

```
data_clean <- data_clean %>%
  # Convert TIMESTAMP to a date format and extract year and month in one step
  mutate(
    DATE = as.Date(as.character(TIMESTAMP), format = "%Y%m%d"),
```

```
    year = format(DATE, "%Y"),
    month = as.integer(format(DATE, "%m"))
  ) %>%
  # Group by both year and month
  group_by(year, month) %>%
  # Summarize all numerical columns by mean
  summarise(across(everything(), mean, na.rm = TRUE))
```

ADJUST RECO

`RECO_adjusted` is a mutated version of our original ecological respiration variable (`RECO_DT_VUT_REF`). For a given row, this value represents the RECO value 2 months before the month of the given row.

- We created this variable because we noticed that RECO tends to "lag" behind precipitation, that is, it tends to peak about 2 months after presentation (we postulate that RECO's peak is a result of precipitation's peak). Thus, we created this new variable in order to test whether we will see a greater correlation than if we used the plain RECO variable.

```
data_clean <- data_clean %>%
  mutate(
    # Extract the year and month
    year = as.integer(format(as.Date(DATE), "%Y")),
    month = as.integer(month),

    # Adjust the month backward by 2 for all rows
    month_adjusted = ifelse(month <= 10, month + 2, month - 10),
    year_adjusted = ifelse(month <= 10, year, year + 1)
  )

data_clean <- data_clean %>%
  # Convert year_adjusted to character to match the year column
  mutate(year_adjusted = as.integer(year_adjusted)) %>%
  # Perform a self-join to match month_adjusted/year_adjusted with month/year
  left_join(
```

```
    data_clean,
    by = c(
      "month_adjusted" = "month",
      "year_adjusted" = "year"),
    suffix = c("", "_matched")) %>%
  # Create the RECO_adjusted column by using the RECO_DT_VUT_REF
  # from the matched row
  mutate(RECO_adjusted = RECO_DT_VUT_REF_matched) %>%
  # Drop the extra columns from the join if not needed (this is optional)
  select(-contains("_matched"))
```

Create YEARMONTH column

This will show the year and month in YYYYMM format.

```
data_clean <- data_clean %>%
  mutate(YEARMONTH = paste0(year, sprintf("%02d", month))) %>%
  group_by(YEARMONTH)
```

YEAR

Pull `year` out of `YEARMONTH`.

```
data_clean$year <- substr(data_clean$YEARMONTH, 1, 4)
```

Time Series

Normalize `RECO_adjusted` so that its scale will line up with SWC.

```
RECO_min <- min(data_clean$RECO_adjusted, na.rm = TRUE)
RECO_max <- max(data_clean$RECO_adjusted, na.rm = TRUE)

data_clean$RECO_adjusted_normalized <-
  (data_clean$RECO_adjusted - RECO_min) / (RECO_max - RECO_min) * 100
```

Prepare the data by calculating monthly means of SWC and RECO Adjusted.

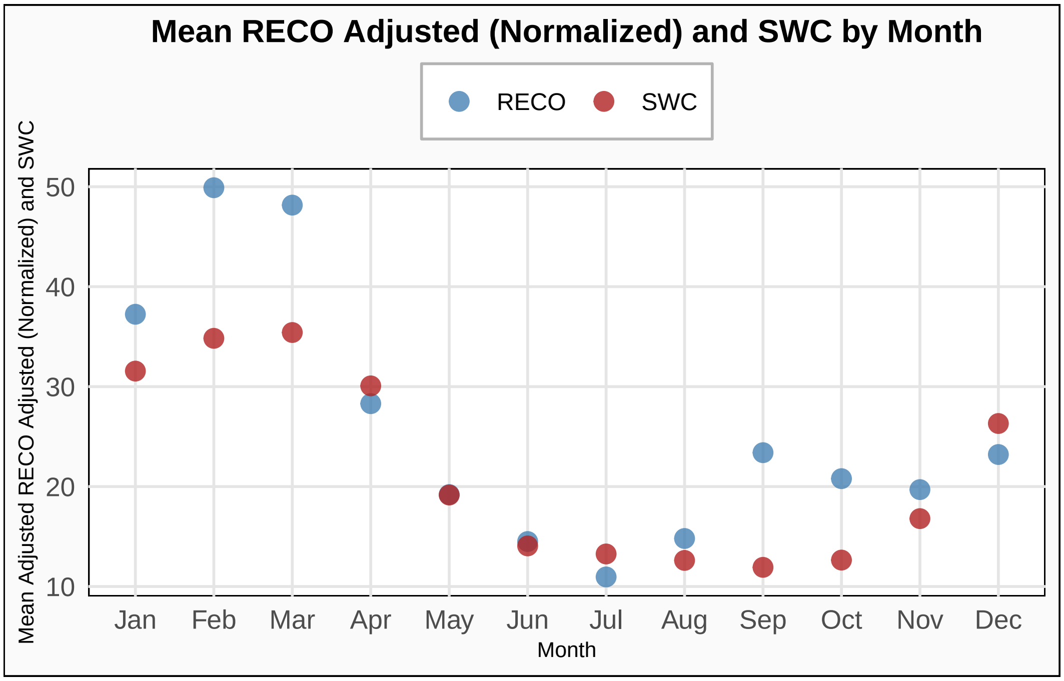
```
month_names <- c("Jan", "Feb", "Mar", "Apr", "May", "Jun",
                 "Jul", "Aug", "Sep", "Oct", "Nov", "Dec")

monthly_summary_RECO_adjusted_normalized <- data_clean %>%
  group_by(month) %>%
  summarise(
    mean_RECO = mean(RECO_adjusted_normalized, na.rm = TRUE),
    std_RECO = sd(RECO_adjusted_normalized, na.rm = TRUE)
  ) %>%
  mutate(month = factor(month, levels = 1:12, labels = month_names))
```

```
monthly_summary_SWC <- data_clean %>%
  group_by(month) %>%
  summarise(
    mean_SWC = mean(SWC_F_MDS_2, na.rm = TRUE),
    std_SWC = sd(SWC_F_MDS_2, na.rm = TRUE)
  ) %>%
  mutate(month = factor(month, levels = 1:12, labels = month_names))
```

Create time series plot:

```
plot2 <- ggplot() +
  # RECO adjusted points
  geom_point(data = monthly_summary_RECO_adjusted_normalized,
             aes(x = month, y = mean_RECO, color = "RECO"),
             size = 3, alpha = 0.8) +    # Points for RECO with transparency

  # Precipitation points
  geom_point(data = monthly_summary_SWC,
             aes(x = month, y = mean_SWC, color = "SWC"),
             size = 3, alpha = 0.8) +

  # Define colors manually for the legend
  scale_color_manual(
    name = NULL,  # Remove the legend title
```

```
    values = c("RECO" = "steelblue", "SWC" = "firebrick")
  ) +

  # Labels and title
  labs(
    x = "Month",
    y = "Mean Adjusted RECO Adjusted (Normalized) and SWC",
    title = "Mean RECO Adjusted (Normalized) and SWC by Month"
  ) +

  # Minimal theme and formatting
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold", size = 12),
    axis.title.x = element_text(size = 8),  # X-axis title size
    axis.title.y = element_text(size = 8),  # Smaller Y-axis title size
    axis.text = element_text(size = 10),  # Larger axis text
    panel.grid.major = element_line(color = "gray90"),
    panel.grid.minor = element_blank(),  # Hide minor gridlines
    panel.background = element_rect(fill = "white"),  # White panel background
    plot.background = element_rect(fill = "gray98"),
    legend.background = element_rect(fill = "white", color = "gray70"),
    legend.position = "top"  # Legend on top for better layout
  )

plot2
```



Simple linear regression model

RECO Adjusted ~ Precipitation

First, we tried a simple linear regression model that predicts RECO adjusted based on precipitation. We opted not to use this model because RECO adjusted ~ SWC was more accurate.

```
model_reco_p_f <- lm(RECO_adjusted ~ P_F, data = data_clean)
summary(model_reco_p_f)
```

```
Call:
lm(formula = RECO_adjusted ~ P_F, data = data_clean)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.1934 -0.7750 -0.3070  0.4993  4.3733 

Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.04369    0.09522  21.463  < 2e-16 ***
P_F          0.29796    0.03749   7.948 6.68e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.192 on 248 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.203, Adjusted R-squared:  0.1998
F-statistic: 63.17 on 1 and 248 DF,  p-value: 6.682e-14
```

```r
# Perform the correlation test
result_reco_p_f <- cor.test(
  data_clean$P_F,
  data_clean$RECO_adjusted,
  method = "pearson"
  )
print(result_reco_p_f)
```

```
    Pearson's product-moment correlation

data:  data_clean$P_F and data_clean$RECO_adjusted
t = 7.9477, df = 248, p-value = 6.682e-14
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3458165 0.5441997
sample estimates:
      cor
0.4505528
```

RECO Adjusted ~ SWC

The following model predicts RECO adjusted based on SWC. We used this model on our poster.

```r
model_reco_swc <- lm(RECO_adjusted ~ SWC_F_MDS_2, data = data_clean)
summary(model_reco_swc)
```

```

Call:
lm(formula = RECO_adjusted ~ SWC_F_MDS_2, data = data_clean)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.7851 -0.6510 -0.2028  0.4962  4.1297 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.759056   0.151305   5.017 1.03e-06 ***
SWC_F_MDS_2 0.083645   0.006364  13.143  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.009 on 236 degrees of freedom
  (14 observations deleted due to missingness)
Multiple R-squared:  0.4226,	Adjusted R-squared:  0.4202 
F-statistic: 172.7 on 1 and 236 DF,  p-value: < 2.2e-16
```

```r
confint(model_reco_swc)
```

```
                 2.5 %     97.5 %
(Intercept) 0.46097528 1.05713768
SWC_F_MDS_2 0.07110696 0.09618242
```

```r
# Perform the correlation test
result_RECO <- cor.test(
  data_clean$SWC_F_MDS_2,
  data_clean$RECO_adjusted,
  method = "pearson"
  )
print(result_RECO)
```

```
	Pearson's product-moment correlation

data:  data_clean$SWC_F_MDS_2 and data_clean$RECO_adjusted
t = 13.143, df = 236, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5700553 0.7179070
sample estimates:
      cor 
0.6500924 
```

```r
# Print the confidence interval
print(result_RECO$conf.int)
```

```
[1] 0.5700553 0.7179070
attr(,"conf.level")
[1] 0.95
```

AIC and BIC

```r
aic_value <- AIC(model_reco_swc)

bic_value <- BIC(model_reco_swc)

# Print values
print("AIC and BIC for linear Regression Model")
```

```
[1] "AIC and BIC for linear Regression Model"
```

```r
print(paste("AIC:", aic_value))
```

```
[1] "AIC: 683.737786654956"
```

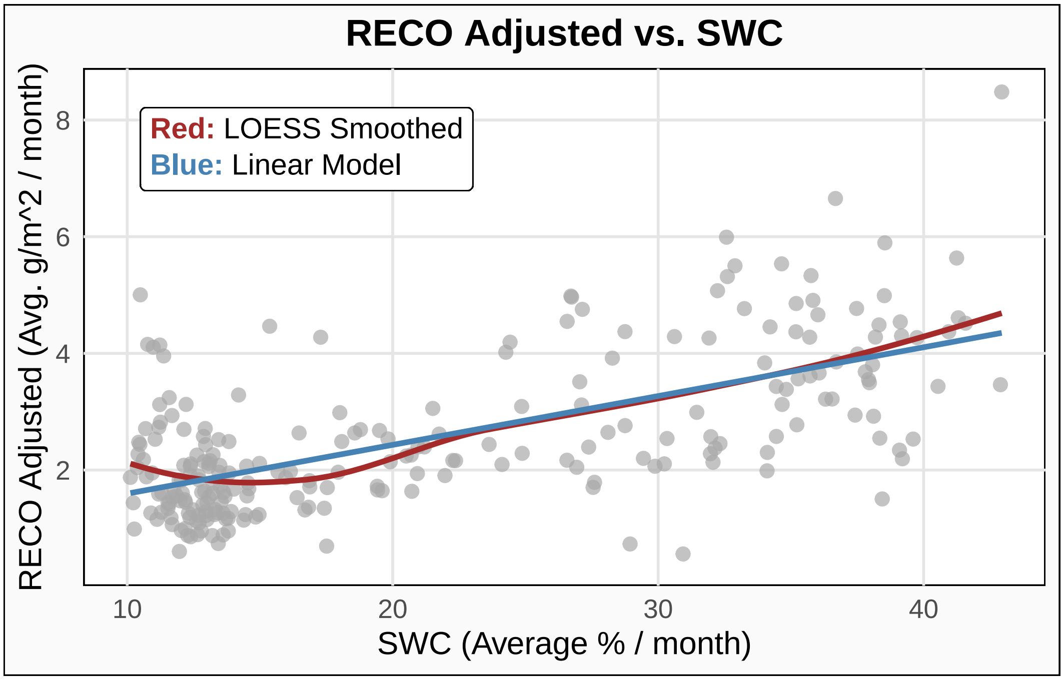
```
print(paste("BIC:", bic_value))
```

```
[1] "BIC: 694.15459867597"
```

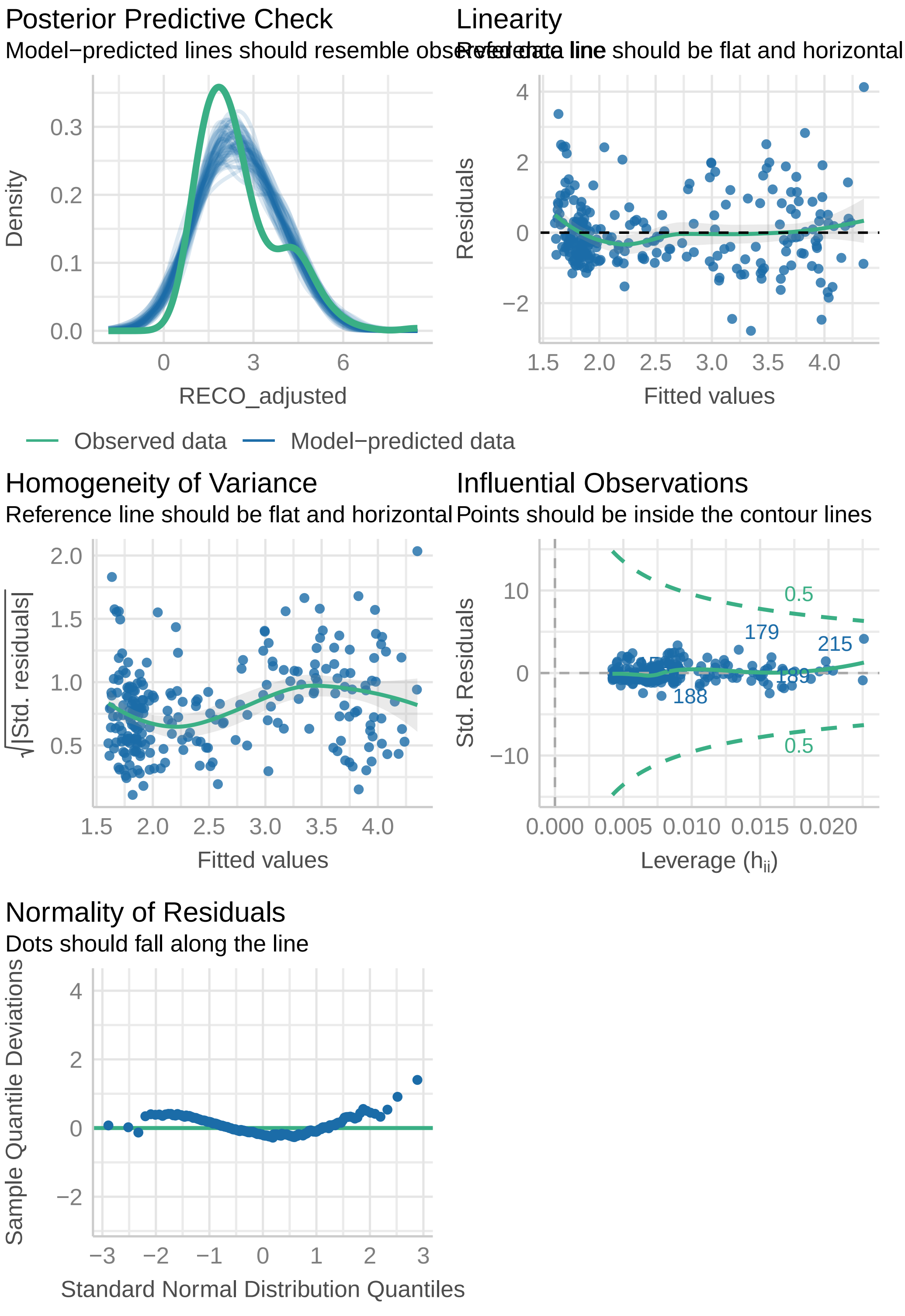
Plot that we used on our poster:

```
plot4 <- ggplot(data_clean, aes(x = SWC_F_MDS_2, y = RECO_adjusted)) +
  geom_point(color = "darkgray", alpha = 0.7, size = 2) +
  geom_smooth(se = FALSE, col = "brown", method = "loess") +
  geom_smooth(se = FALSE, method = "lm", col = "steelblue") +
  theme_bw() +  # Apply theme_bw()
  labs(
    title = "RECO Adjusted vs. SWC",
    x = "SWC (Average % / month)",
    y = "RECO Adjusted (Avg. g/m^2 / month)"
  ) +
  geom_richtext(
    aes(x = 10.5, y = 7.5),
    label = paste0(
      "<b style='color:brown;'>Red:</b> LOESS Smoothed<br>",
      "<b style='color:steelblue;'>Blue:</b> Linear Model"
    ),
    fill = "white",
    label.color = "black",  # Border color
    size = 4,
    hjust = 0,
    inherit.aes = FALSE
  ) +
  # Minimal theme and formatting
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
    axis.title = element_text(size = 12),  # Larger axis titles
    axis.text = element_text(size = 10),  # Larger axis text
    panel.grid.major = element_line(color = "gray90"),
    panel.grid.minor = element_blank(),  # Hide minor gridlines
```

```
    panel.background = element_rect(fill = "white"),  # White panel background
    plot.background = element_rect(fill = "gray98")
  ) +
  xlim(10, NA)

plot4
```




```
check_model(model_reco_swc)
```


Posterior Predictive Check
Model−predicted lines should resemble observed data line
Density
RECO_adjusted
Linearity
Reference line should be flat and horizontal
Residuals
Fitted values
Observed data
Model−predicted data
Homogeneity of Variance
Reference line should be flat and horizontal
√|Std. residuals|
Fitted values
Influential Observations
Points should be inside the contour lines
Std. Residuals
Leverage (h_{ii})
179
215
188
189
0.5
0.5
Normality of Residuals
Dots should fall along the line
Sample Quantile Deviations
Standard Normal Distribution Quantiles

Multiple Linear Regression Model

RECO Adjusted ~ SWC + temperature + year

Attempting multiple linear regression model that predicts RECO adjusted based on SWC, temperature, and year. We decided not to use this model because `year` is not statistically significant.

```r
data_clean$year <- as.numeric(data_clean$year)

model_multiple_year <- lm(
  RECO_adjusted ~ SWC_F_MDS_2 + TS_F_MDS_1 + year,
  data=data_clean
  )
summary(model_multiple_year)
```

```
Call:
lm(formula = RECO_adjusted ~ SWC_F_MDS_2 + TS_F_MDS_1 + year,
    data = data_clean)

Residuals:
    Min      1Q  Median      3Q     Max
-2.7673 -0.6076 -0.1607  0.4818  4.1161

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -17.001507  22.714666  -0.748   0.4549
SWC_F_MDS_2   0.064723   0.009141   7.080 1.67e-11 ***
TS_F_MDS_1   -0.044805   0.015696  -2.855   0.0047 **
year          0.009429   0.011278   0.836   0.4040
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.994 on 234 degrees of freedom
  (14 observations deleted due to missingness)
Multiple R-squared:  0.4446,    Adjusted R-squared:  0.4375
```

```
F-statistic: 62.44 on 3 and 234 DF,  p-value: < 2.2e-16
```

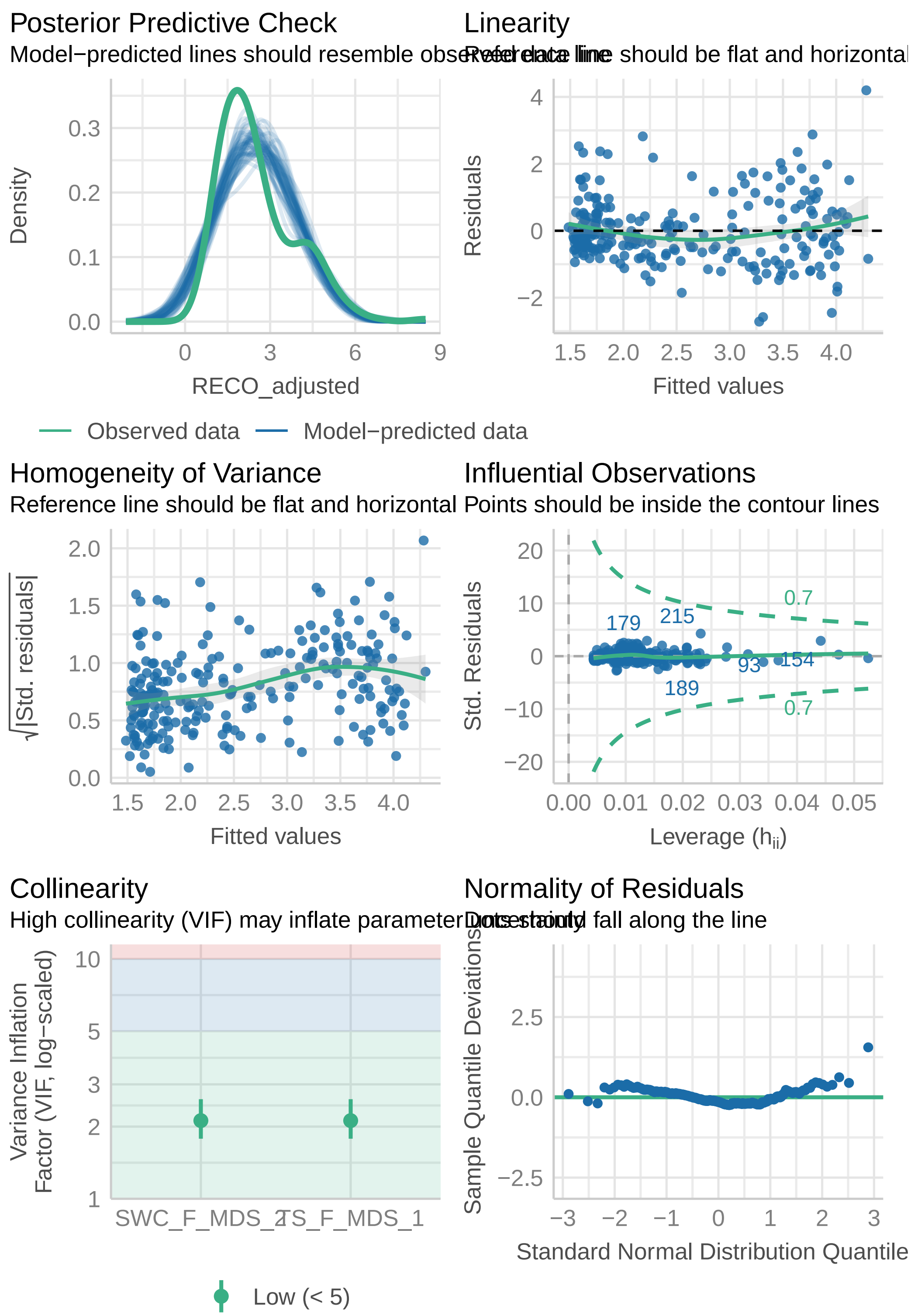
RECO Adjusted ~ SWC + temperature

Following is our multiple linear regression model that predicts RECO adjusted based on SWC and temperature. We used this model in our poster.

```r
model_lm <- lm(RECO_adjusted ~ TS_F_MDS_1 + SWC_F_MDS_2, data = data_clean)
# Summarize the model to see results
summary(model_lm)
```

```
Call:
lm(formula = RECO_adjusted ~ TS_F_MDS_1 + SWC_F_MDS_2, data = data_clean)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.7149 -0.5906 -0.1436  0.4891  4.1974 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.985355   0.444426   4.467 1.23e-05 ***
TS_F_MDS_1  -0.045804   0.015640  -2.929  0.00374 ** 
SWC_F_MDS_2  0.064244   0.009117   7.046 2.02e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9933 on 235 degrees of freedom
  (14 observations deleted due to missingness)
Multiple R-squared:  0.443, Adjusted R-squared:  0.4382 
F-statistic: 93.43 on 2 and 235 DF,  p-value: < 2.2e-16
```

```r
check_model(model_lm)
```

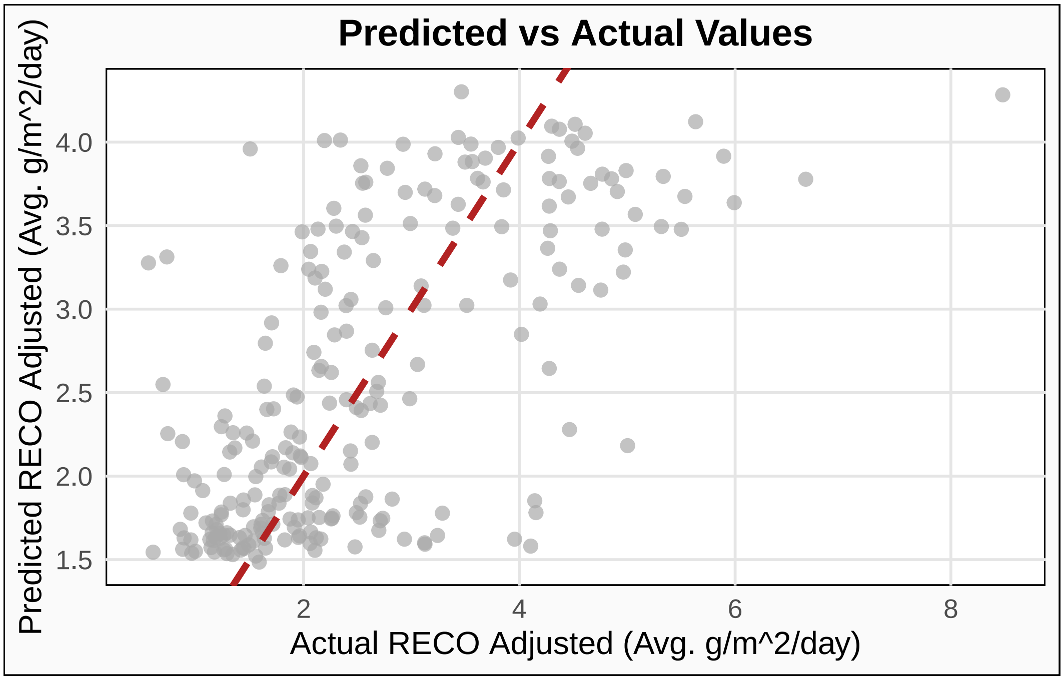

Posterior Predictive Check
Model−predicted lines should resemble observed data line
Density
RECO_adjusted
Observed data
Model−predicted data
Linearity
Reference line should be flat and horizontal
Residuals
Fitted values
Homogeneity of Variance
Reference line should be flat and horizontal
√|Std. residuals|
Fitted values
Influential Observations
Points should be inside the contour lines
Std. Residuals
Leverage (h_{ii})
179
215
93
154
189
0.7
Collinearity
High collinearity (VIF) may inflate parameter uncertainty
Variance Inflation Factor (VIF, log−scaled)
SWC_F_MDS_2
TS_F_MDS_1
Low (< 5)
Normality of Residuals
Dots should fall along the line
Sample Quantile Deviations
Standard Normal Distribution Quantile

```r
confint(model_lm)
```

```
                  2.5 %      97.5 %
(Intercept)  1.10978611  2.86092308
TS_F_MDS_1  -0.07661702 -0.01499132
SWC_F_MDS_2  0.04628156  0.08220611
```

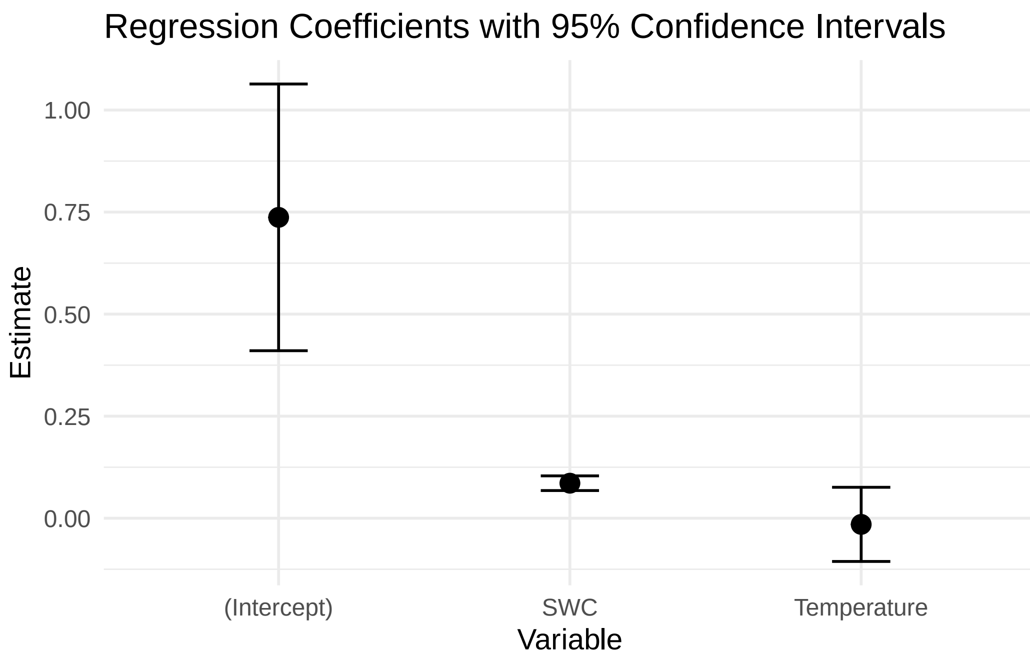
```r
# Extract rows used in the model
data_clean_used <- model_lm$model

# Add predictions to the filtered dataset
data_clean_used$predicted <- predict(model_lm)

# Create the ggplot with enhanced styling
plot3 <- ggplot(data_clean_used, aes(x = RECO_adjusted, y = predicted)) +
  # Scatter points with adjusted color and transparency
  geom_point(color = "darkgray", alpha = 0.7, size = 2) +

  # Reference line (y = x) with a red dashed line
  geom_abline(intercept = 0, slope = 1, color = "firebrick",
              linetype = "dashed", size = 1.2) +

  # Labels and title
  labs(
    title = "Predicted vs Actual Values",
    x = "Actual RECO Adjusted (Avg. g/m^2/day)",
    y = "Predicted RECO Adjusted (Avg. g/m^2/day)"
  ) +

  # Minimal theme and formatting
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
    axis.title = element_text(size = 12),  # Larger axis titles
    axis.text = element_text(size = 10),  # Larger axis text
    panel.grid.major = element_line(color = "gray90"),
```

```
    panel.grid.minor = element_blank(),  # Hide minor gridlines
    panel.background = element_rect(fill = "white"),  # White panel background
    plot.background = element_rect(fill = "gray98")
  )

plot3
```




```
# Create a data frame for coefficients
coeff_df <- data.frame(
  Variable = c("(Intercept)", "Temperature", "SWC"),
  Estimate = c(0.7371, -0.0151, 0.0858),
  CI_Lower = c(0.4104, -0.1059, 0.0678),
  CI_Upper = c(1.0638, 0.0758, 0.1038)
)

# Coefficient plot
ggplot(coeff_df, aes(x = Variable, y = Estimate)) +
  geom_point(size = 3) +  # Points for estimates
  geom_errorbar(aes(ymin = CI_Lower, ymax = CI_Upper), width = 0.2) +
  labs(title = "Regression Coefficients with 95% Confidence Intervals",
```

```
       x = "Variable",
       y = "Estimate") +
  theme_minimal()
```



AIC and BIC

```
aic_value <- AIC(model_lm)
bic_value <- BIC(model_lm)

# Print values
print("AIC and BIC for *Multiple* Linear Regression Model")
```

```
[1] "AIC and BIC for *Multiple* Linear Regression Model"
```

```
print(paste("AIC:", aic_value))
```

```
[1] "AIC: 677.206216727659"
```

```r
print(paste("BIC:", bic_value))
```

```
[1] "BIC: 691.095299422345"
```