# Monarch Butterflies and Microclimate

**STAT 414 - Final Report**

Kyle Nessen, Justin Mai, Aiden Kelly, Arneh Begi

2024-12-11

## Introduction

Every fall, monarch butterflies (*Danaus plexippus*) undergo an annual mass migration from their summer breeding grounds across western North America to overwintering groves along the Pacific Coast (Tuskes and Brower 1978; Urquhart and Urquhart 1978). During winter months, these butterflies "cluster" by the thousands on trees within the groves, often in the same location year after year, despite being those butterflies being generations apart. Scientists hypothesize that monarchs seek out these locations within the groves because they provide ideal weather conditions, including mild temperatures, high humidity, and variable light (Leong et al. 1991; Chaplin and Wells 1982; Weiss et al. 1991). They predict that monarchs will select habitat that meet all of these climatic conditions and will not be found in areas where those conditions are not met, both at the landscape and site levels.

This study investigates how microclimate conditions (temperature, humidity, and light) differ between monarch butterfly aggregation sites and non-aggregation areas within the same grove, seeking to identify whether specific environmental characteristics distinguish the locations where monarchs choose to cluster.

## Materials and Methods

Our dataset comes from research conducted by Saniee and Dr. Villablanca (Saniee and Villablanca 2022). They directly investigated the "microclimate hypothesis" by taking temperature, humidity, and light measurements at various locations within monarch overwintering habitat. Within each grove, they placed a weather station (array) at a known aggregation site, where monarch butterflies are thought to be selecting favorable weather conditions. From that aggregation site, they placed four additional arrays in a repeated manner to capture variation within the grove (see Figure 1). This arrangement of weather arrays was repeated across nine

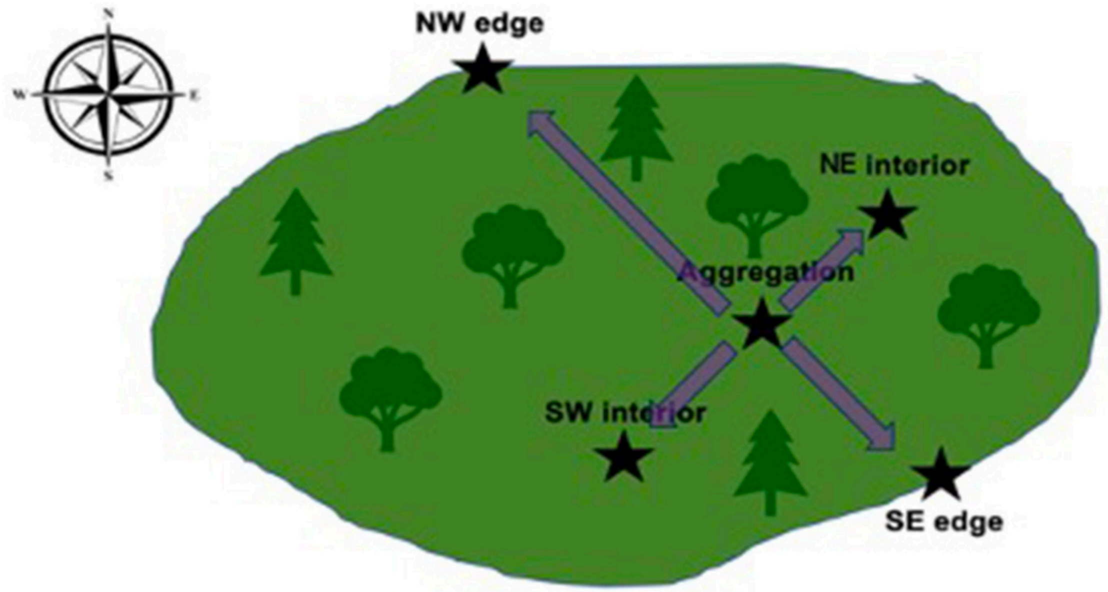



Figure 1: Sampling design relative to the aggregation's location in the groves. The first sample location was placed in the location of an aggregation (Aggregation). Two more sample locations were placed on the SE and NW edges of each grove relative to the aggregation's location to capture morning light and prevailing wind (SE edge and NW edge, respectively). Two interior sample locations were placed halfway between the aggregation's location and the grove's edge in the NE and SW directions (NE interior and SW interior, respectively).

groves in San Luis Obispo, Santa Barbara, and Ventura County (see Figure 2). Climate measurements were taken every 5 minutes, and derivative summary measures were generated for each day (e.g., minimum, maximum, and average temperature).

Point Conception, the westernmost point of California, represents a significant biogeographical boundary that influences numerous ecological systems along the Pacific coast. To examine potential latitudinal patterns in monarch butterfly habitat selection, we classified our study sites into two distinct regions, those located north and those located south of Point Conception (see Figure 2). This classification enabled us to investigate whether the geographical divide corresponds to meaningful differences in monarch butterfly aggregation patterns and associated microclimate conditions.


Figure 2: Groves where weather station arrays were placed in this study. Sites north of yellow dashed line were categorized as north, others were categorized as south.

Our dataset has two levels: arrays (weather stations) are Level 1, and grove is Level 2. Below is summary table of the variables used in our study:

| Variable | Level | Description |
|---|---|---|
| Average Temperature (C) | 1 | Average temperature for the sampled day |
| Minimum temperature (C) | 1 | Minimum temperature for the sampled day |
| Maximum temperature (C) | 1 | Maximum temperature for the sampled day |
| Average humidity (%) | 1 | Average humidity for the sampled day |
| Average light (lux) | 1 | Average brightness at weather array for the sampled day |
| Standard Deviation of light (lux) | 1 | The variability of light for the sampled day |
| Monarch Presence (binary) | 1 | Whether the weather array was positioned at a known monarch aggregation site. 1 = monarch site, 0 = all others |
| Day of Season | 1 | Count of days since the beginning of the monarch overwintering season |
| North/South (binary) | 2 | Grove level categorization of latitudinal position. 0 = southern groves, 1 = northern groves. |

can also outline statistical methods, soft

## Results

To more directly investigate the microclimate hypothesis, we reclassified the array stations (aggregation, NE, NW, SE, SW) into a binary variable representing monarch butterfly presence or absence (1, 0). We attempted several logistic regression models using this presence/absence classification as our response variable, with various microclimate measurements as predictors.

However, these models proved statistically inappropriate due to complete separation in our dataset — monarch presence was perfectly consistent at aggregation sites and perfectly absent at all other array locations throughout the study period. This perfect separation prevented meaningful model convergence and parameter estimation, as there was no variation in presence/absence at any given location. Consequently, none of our attempted models provided better explanatory power than the null model, suggesting that alternative analytical approaches were needed to understand the relationship between microclimate conditions and monarch site selection.

right idea, thou

In our alternative approach, we developed a series of linear mixed-effects models to predict each microclimate variable (temperature, humidity, and light) independently. Our base models incorporated `grove` as a random effect to account for site-specific variation and included `seasonDay` (days since the start of the season) as a fixed effect to control for temporal trends. We then compared these base models to expanded versions that included monarch presence as an additional predictor. This approach allowed us to evaluate whether monarch presence significantly improved our ability to predict microclimate conditions, while accounting for both the temporal nature of our data and the nested structure of our sampling design within different groves.

but then need to worry about linearity with tl

### Exploratory Data Analysis

Our analysis focused on environmental variables with direct biological significance for monarch butterfly survival and habitat selection. For temperature, we examined three metrics: minimum, maximum, and mean daily temperatures. These represent critical biological thresholds - minimum temperatures that could prevent flight or cause freezing, maximum temperatures that could drive excessive metabolic costs, and mean temperatures to characterize overall thermal conditions. Daily humidity averages were analyzed as higher humidity generally supports butterfly survival, with extreme values being less biologically relevant. For light conditions, we examined both average daily light levels and their standard deviation, with the latter serving as a proxy for "dappled light" - the variable light conditions created by forest canopy that may help butterflies regulate their body temperature (see appendix for video). so these are o

Initial exploratory analysis revealed substantial variation in these microclimate variables both within and across groves. Temperature patterns showed clear site-specific differences, with mean daily temperatures ranging from approximately 9°C to 14°C across groves. The boxplots indicate that some groves (e.g., HR) consistently maintained higher temperatures, while others (e.g., BL) showed consistently lower temperatures throughout the study period. include some/

Light measurements demonstrated particularly interesting patterns. While average light levels varied considerably between groves, the standard deviation of light measurements - our measure of light variability showed even more pronounced differences between butterfly aggregation sites and control locations. This variation in light conditions could be especially important given the hypothesized role of dappled light in butterfly thermoregulation.

by whom?

Humidity patterns revealed generally high variance within groves but also suggested some systematic differences between northern and southern sites, warranting further investigation of geographical effects in our subsequent modeling.

Please see Appendix for all plots related to temperature, humidity, and light.

### Level 2 Grouping

Using our temperature as a test response variable, we ran an anova with and without north to see if there were differences.

1st ANOVA request

| | Variable | AIC_with_north | AIC_without_north | p_value |
|---|---|---|---|---|
| 1 | temp.min_centered | 11426.83 | 11430.94 | 0.41741372 |
| 2 | temp.max_centered | 15132.34 | 15141.69 | 0.51111874 |
| 3 | hum.avg_centered | 15255.60 | 15261.82 | 0.01197886 |
| 4 | light.avg_centered | 47400.77 | 47415.71 | 0.90341318 |
| 5 | light.std_centered | 50787.98 | 50807.99 | 0.86721917 |

In the first ANOVA results, we studied the importance of the Level 2 grouping variable, North, across different small climate variables. North did not show a big effect on temperature minimum (temp.min_centered) and temperature maximum (temp.max_centered). The p-values were 0.417 and 0.511, meaning North/South grouping probably does not affect these temperature measures much. For humidity average (hum.avg_centered), the p-value of 0.012 showed that North really impacted humidity, suggesting differences in humidity between North and South. For light average (light.avg_centered) and light standardized (light.std_centered), North did not play a big role, as seen from p-values of 0.903 and 0.867. These results suggest that North/South differences matter for humidity but not much for temperature or light. Humidity differences are very important.

### The Null Model

Since with our unique model structure we could not construct a null model without a set response variable. So what we decided to do was create a general null model; we needed to account for time with seasonDay, and have seasonDay vary across the groves, to account for groves that had their studies start earlier than the others. Since we only cared about if an environmental condition was associated with a monarch presence our null model would not include it as a predictor. Our response in the null model would be the environmental condition that we wanted to study. We would test the association with monarch presence by including it as a predictor and comparing it to the null model, so we reached this null model:

need a graph

```
null_model <- lmer(enviornmental_condtion ~ seasonDay + (1 + seasonDay |
grove), data = df2)
```

### The Final Model

We started developing our model by checking basic models for each centered response variable. These variables were temp.min_centered, temp.max_centered, hum.avg_centered, light.avg_centered and light.std_centered. The null models included seasonDay as a fixed effect and allowed both random intercepts and random slopes for seasonDay across groves to account for variability in patterns over time within each grove

Then, we added butterfly_present as another factor to the basic models and checked how it interacted with north, our Level 2 grouping variable. We used likelihood ratio tests via ANOVA to compare these updated models to their null versions.

These comparisons showed that butterfly_present had a strong effect only on light.avg_centered and light.std_centered. Since we saw that north was significant with one of our previous predictors, we wanted to interact it with butterfly_present to see if butterflies were more clarify interacti apparent in the north or south. Further examination showed that, when including how butterfly_present and north interacted, the effect of butterfly_present on light.avg_centered stopped being statistically significant, but for light.std_centered, butterfly_present was still statistically significant, so that was our final model.

The final model for light.std_centered:

```
Linear mixed model fit by REML ['lmerMod']
Formula: light.std_centered ~ seasonDay + butterfly_present * north +
    (1 + seasonDay | grove)
   Data: df2

REML criterion at convergence: 50755.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.8358 -0.6378 -0.2744  0.3839  3.5816

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 grove    (Intercept)  78045647  8834.34
          seasonDay        1374    37.07 -1.00
 Residual             240198680 15498.34
Number of obs: 2295, groups:  grove, 9

Fixed effects:
                        Estimate Std. Error t value
(Intercept)             -955.264   3866.301  -0.247
seasonDay                  7.852     22.259   0.353
butterfly_present       2862.969   1238.385   2.312
north                   -443.008   4282.139  -0.103
butterfly_present:north  497.309   1634.651   0.304

Correlation of Fixed Effects:
            (Intr) sesnDy bttrf_ north
seasonDay   -0.548
```

```
bttrfly_prs -0.058  0.019
north       -0.618 -0.025  0.042
bttrfly_pr:  0.034  0.003 -0.757 -0.055
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

don't include

```
Linear mixed model fit by REML ['lmerMod']
Formula: light.std_centered ~ seasonDay + (1 + seasonDay | grove)
   Data: df2

REML criterion at convergence: 50820.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.6598 -0.6475 -0.2747  0.3700  3.5385

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 grove    (Intercept) 5.283e+07  7268.12
          seasonDay   8.666e+02    29.44 -1.00
 Residual             2.417e+08 15547.89
Number of obs: 2295, groups:  grove, 9

Fixed effects:
            Estimate Std. Error t value
(Intercept) -467.129   2530.864  -0.185
seasonDay      4.718     20.988   0.225

Correlation of Fixed Effects:
          (Intr)
seasonDay -0.674
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

null model?

As we could not fit a null model with our structure, we had to find that light.std was associated first

Fixed Effects: (Intercept): The estimated average value of light.std_centered when seasonDay is at the start, butterflies are not present, and in the south. The value of -955.264 suggests that, the average centered light variability is negative, but the high standard error 3866.302 indicates this estimate is not precise.

ah, so each gro

seasonDay: The fixed effect of time (days into the season) on light.std_centered. The coefficient (7.852) suggests a slight increasing trend over time, though it is not statistically significant (t = 0.353).

butterfly_present: The effect of butterfly presence on light.std_centered. The estimate (2862.969) suggests that when butterflies are present, the light variability is high associated compared to when they are absent. This effect is statistically significant, as the absolute ? value of the t-statistic (2.312) is greater than the rule-of-thumb threshold (~2). agrees with

north: The baseline difference in light.std_centered between northern and southern groves. The estimate (-443.008) indicates slightly lower light variability in northern groves, but this effect is not statistically significant (t = -0.103).

butterfly_present:north: The interaction term quantifies how the effect of butterfly presence differs between northern and southern groves. The estimate (497.309) suggests a slightly larger increase in light variability in northern groves when butterflies are present. However, this effect is not statistically significant (t = 0.304). This is also supported in (Figure 1) where we can see that even when north is 0 and 1, there is no visual difference in butterfly presence values. Figure 1?

Random Effects: grove Intercept Variance (78045647): There is considerable variability in baseline light variability across groves. The standard deviation (8834.34) represents the average deviation in intercepts across groves.

grove seasonDay Slope Variance (1374): There is minimal variability in how light variability trends over time across groves. The correlation (-1.00) suggests a perfect negative relationship between the random intercepts and slopes. meaning?

Residual Variance (240198680): This represents the variability in light.std_centered not explained by the fixed or random effects, with a standard deviation of 15498.34.

#### Confidence Interval

```
[ 24.2124 , 49.9276 ]
```

The 95% confidence interval for the random slope of seasonDay is [24.2124, 49.9276]. This suggests that, on average, the effect of seasonDay on the centered light.std varies across groves. We are 95% confident that the random slope of seasonDay lies within this range, so there is extreme variance on how seasonDay impacts light.std across different groves. Since 0 is not within the 95% confidence interval for the random slope of seasonDay ([24.2124, 49.9276]), we can conclude that there is significant variability in how seasonDay affects light.std across groves, indicating that the association of butterfly presense is not consistent across all groves. CI for coeffici

Reduction in Level 2 (Intercept) Variance:

$$Reduction = \frac{53,000,000 - 7,800,000}{53,000,000} = 0.851$$

Reduction in Level 1 (Season Day) Variance:

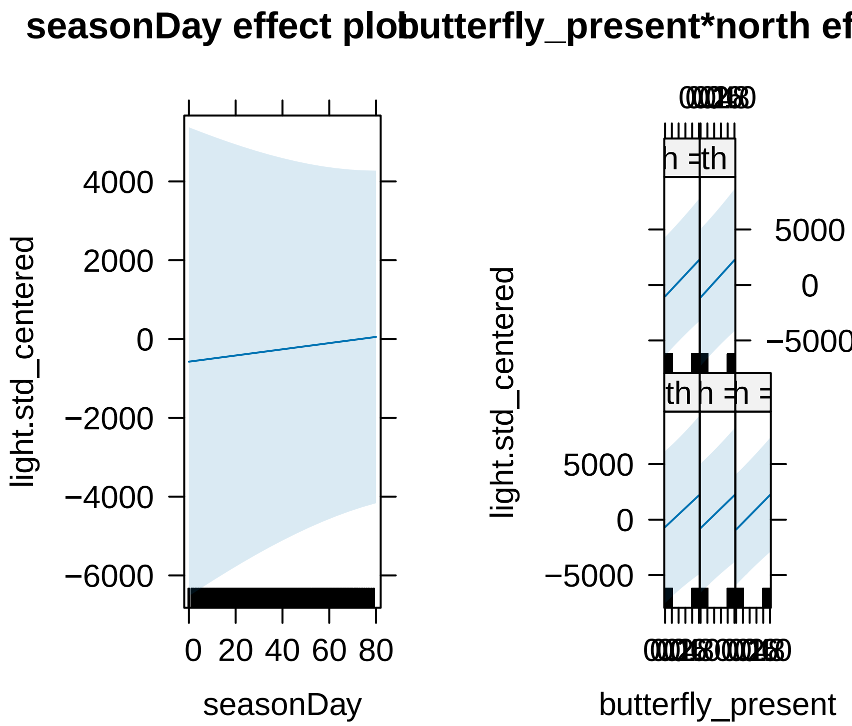
$$Reduction = \frac{866 - 1,374}{866} = -0.587$$

In the final model, we observed a substantial reduction in Level 2 (grove-level) variance, accounting for 85.1% of the unexplained variation. For the fixed effects, the interaction between butterfly presence and north-south grouping explained much of the site-specific variation. slopes? The Level 1 (season-day) variance saw a slight increase due to the added complexity of the interaction terms, indicating that these terms captured additional sources of variation at the daily temporal scale. The residual variance unchanged.

Looking at the residual plots for the final model (Figure 2) suggest that the assumptions of the linear mixed model are generally met. The residuals vs. fitted values plot shows a random scatter of points around the horizontal line, indicating that the residuals are evenly distributed where? reall across the range of fitted values. The normal Q-Q plot shows that the residuals follow a roughly normal distribution, with most points falling close to the diagonal line. However, there are a few points deviating from the line in the tails, suggesting some from normality.

The random effects plot for grove level shows a relatively even distribution of variance across normality? the groves, indicating that the random effects assumption is reasonable.

While there are some minor deviations from the ideal assumptions, the overall model fit appears to be satisfactory. However, it is important to consider the potential impact of the identified issues on the interpretation of the results. Future analyses could explore alternative model specifications or data transformations to improve the model fit.

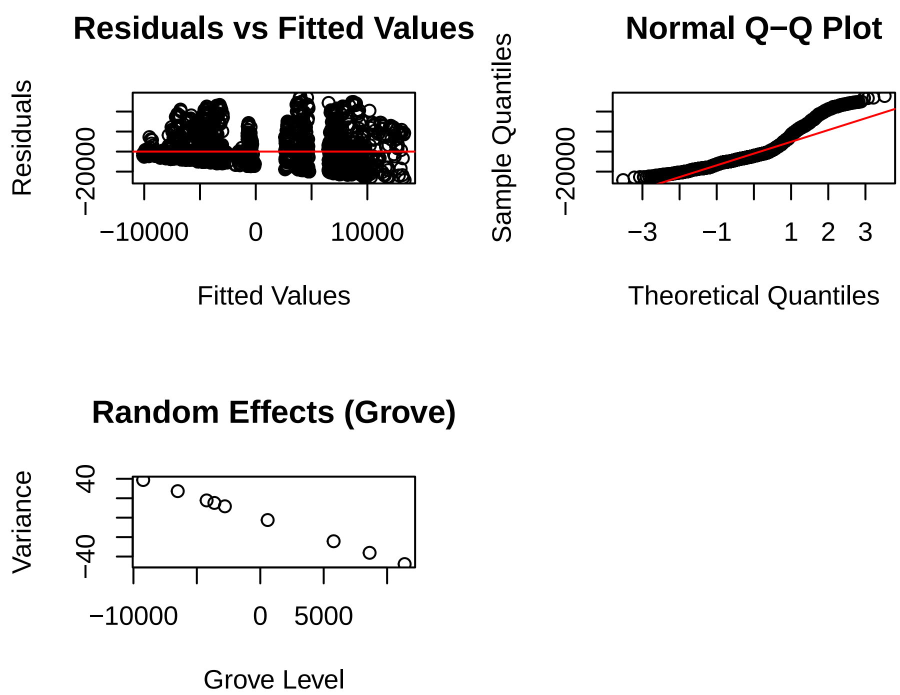



make sure is readable

modelled

This figure shows the relationship between seasonDay, butterfly_present*north, and light.std_centered. The plots indicate a weak or non-existent correlation between the variables, suggesting limited influence on light.std_centered.

why of interest? interpret interaction?




## Discussion

The most significant results that we found were that butterfly presence at an array location was associated with significantly higher deviations of light. As well as, the other weather conditions, such as average humidity and average temperature, not being significantly predicted by butterfly presents. These results indicated to us that the microclimate hypothesis is not broadly supported by our data. In addition to these results, we also found that the weather conditions at each aggregation site were different from each other and there was no difference between each array location within a grove. Again, this is in contrast to the microclimate hypothesis, which would predict that all the aggregation sites would have the weather conditions and the other arrays within the grove would be different from the aggregation site. In the previous study by Kiana Saniee and Francis Villablanca, they found the same results, using a different statistical model, which further supports our findings.

For light to be higher on average at locations where monarch butterflies are present, this means that throughout the day, there is more direct sunlight on the arrays with butterflies. While, a higher standard deviation of light means that throughout the day, the amount of light fluctuates a lot more on sites with butterflies compared to one without. This finding could potentially be in support of another idea discussed in monarch butterfly research. Dappled light refers to varied light that shines through gaps in tree canopies. It is hypothesized that butterflies prefer dappled light since they are cold blooded insects and rely on light in order to warm and cool their bodies off. Since here was not evidence in support of the microclimate hypothesis, potentially this phenomenon is part of the reason why the monarchs are selecting the specify trees they go to. It is also interesting to think of the other factors that could be influencing the butterflies behavior.

A major limitation of our study is that there was no data collected on the count of butterflies at each array location. This would have been useful to have in order to see if the number of butterflies at each location was correlated with the weather conditions. This would have allowed us to run models with butterfly count as a response rather than the conditions of the weather as responses. Enabling us to see how the presence of butterflies changed over time and how potential fluctuations of weather conditions effect monarch presences. Another limitation is the use of the multiple "reverse models". This doesn't allow us to have one model where we would get to include all the variables and see how they interact with each other. This limits the amount of variability that we can account for in any single model.

Despite these limitations the study dataset has several strengths that give credibility to our results. The sampling design was rigorous and well designed, with consistent placement of weather arrays throughout nine distint groves. These groves were far apart, allowing us to control for grove-to-grove differences. Additionally, the high frequency of data capture gives us a rich picture of microclimate variability within the grove across a range of biologically relevant factors. We used a multilevel modeling approach toa ccount for the nested structure of our data. This approach allows us to investigate patterns that may have been missed with simpler approaches.

Future research should incorporate butterfly abundance as a direct response variable to environmental conditions. Using automated cameras for remote monitoring would allow researchers to collect detailed data on butterfly behavior and movement patterns in relation to changing light conditions throughout the day. This technology could help quantify how monarchs utilize different light environments and validate our findings about the importance of variable light conditions.

Given that traditional microclimate variables like temperature and humidity fluctuate significantly and are challenging to predict, investigating light patterns as an indicator of habitat suitability offers a promising alternative approach. The physical structure of tree canopies that creates these light patterns remains relatively stable over time, making it a potentially reliable predictor that monarchs could assess when selecting overwintering sites. Monarchs' well-developed visual system suggests they could be particularly attuned to these structural cues. This approach has practical advantages: habitat assessment based on canopy structure and resulting light patterns could be conducted outside of the overwintering season and without requiring butterfly presence. If structural features prove to be reliable predictors of monarch habitat use, researchers could conduct manipulation experiments - such as selective branch removal or installation of reflective surfaces - to test how monarchs respond to altered light conditions.

Understanding how monarchs select overwintering sites has become increasingly crucial since their recent designation as a federally threatened species ("Endangered and Threatened Wildlife and Plants; Threatened Species Status With Section 4(d) Rule for Monarch Butterfly and Designation of Critical Habitat" 2024). Translating this research into practical guidance for land managers could significantly impact conservation efforts by helping identify, protect, and potentially enhance suitable overwintering habitats.

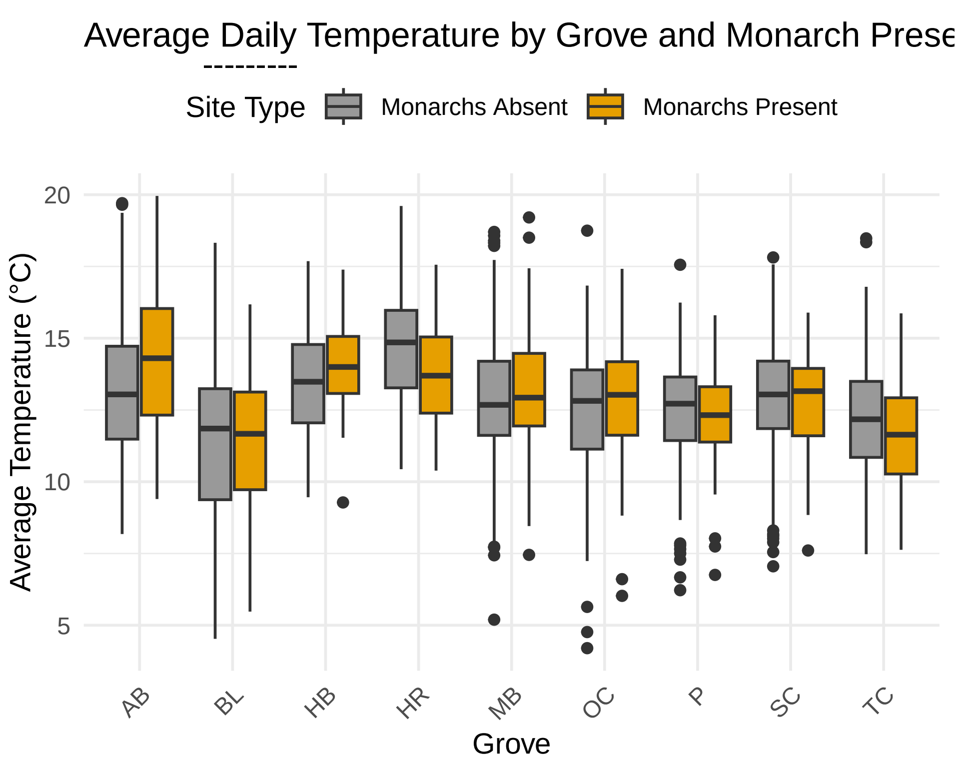
## Appendix

### Dappled Light Timelapse

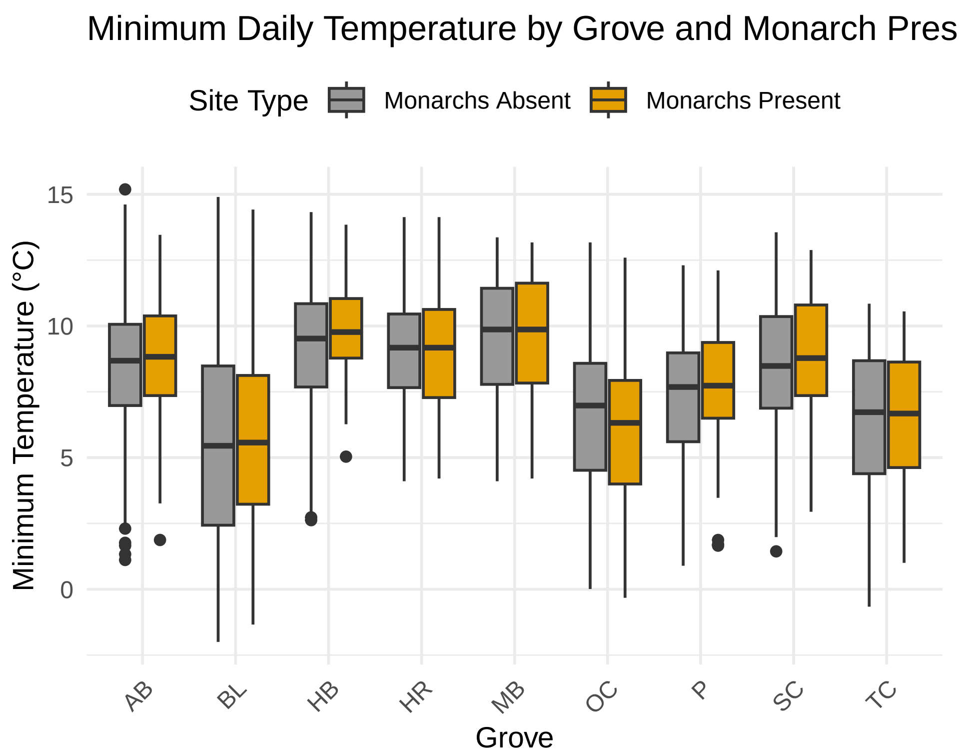
You can see a brief video at the link below of what monarch biologists refer to "dappled light." This is our proxy for the standard deviation of light. SD of light is a proxy for dappled light?

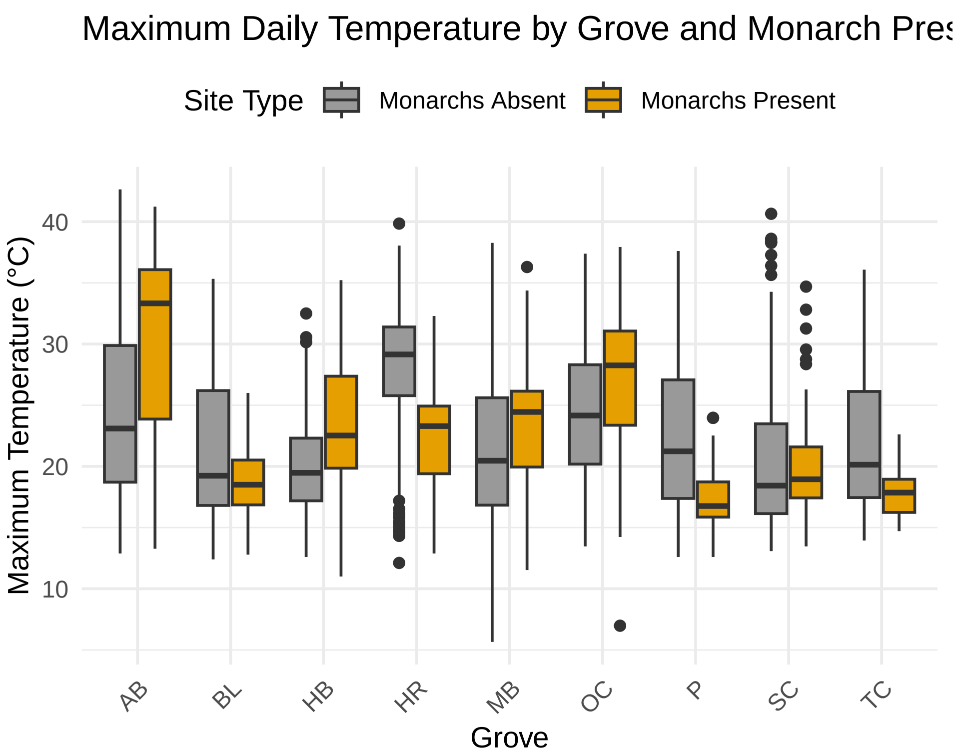
[Dappled Light Timelapse]

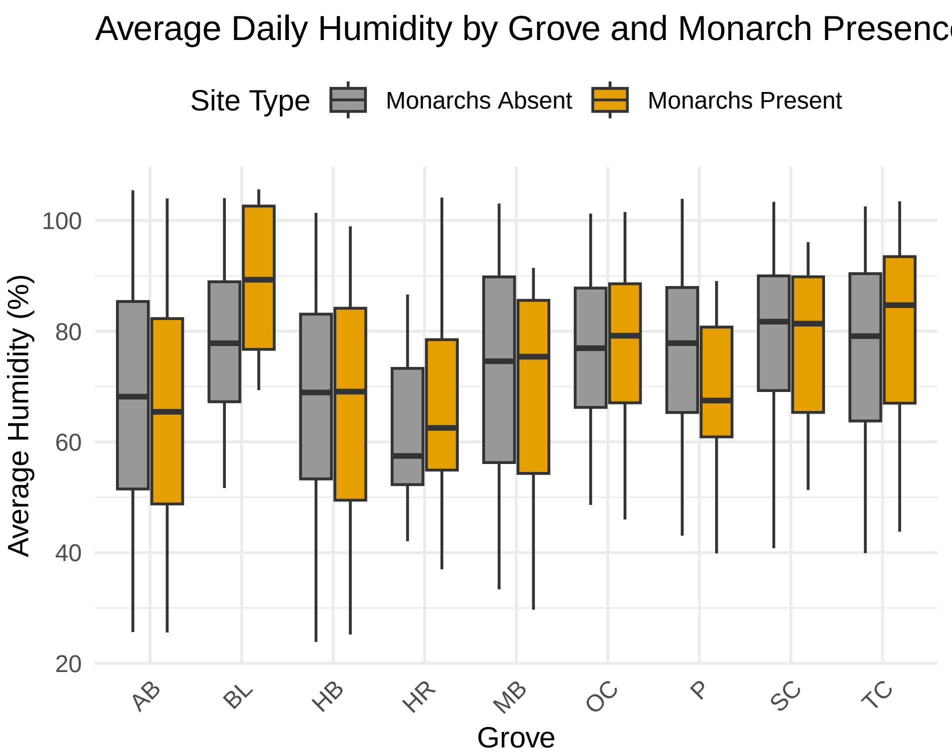
### Exploratory Analysis Plots

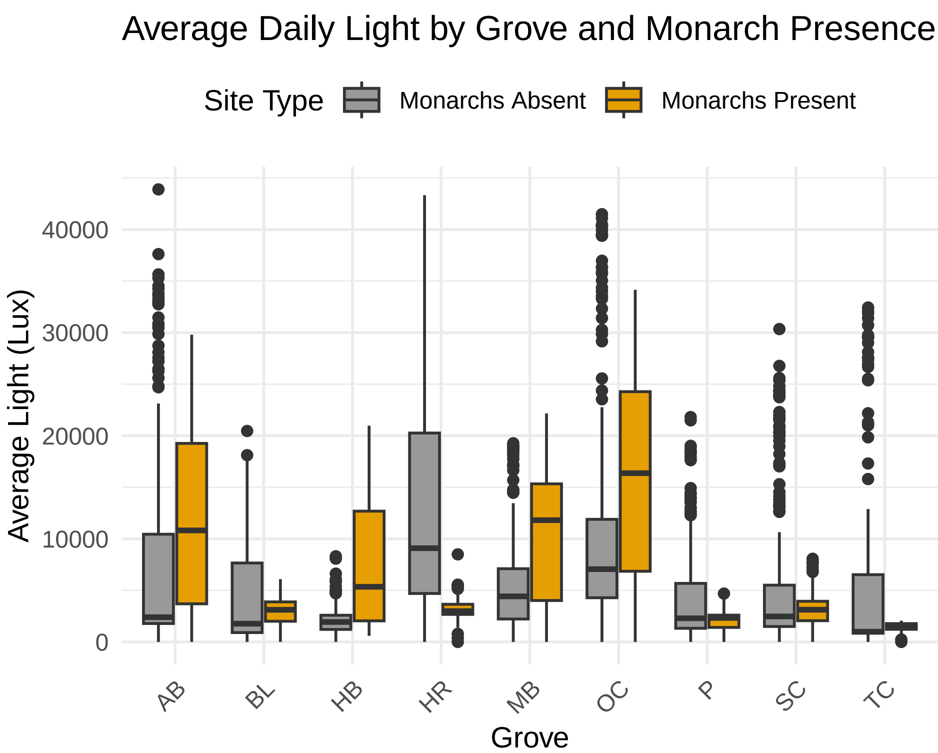


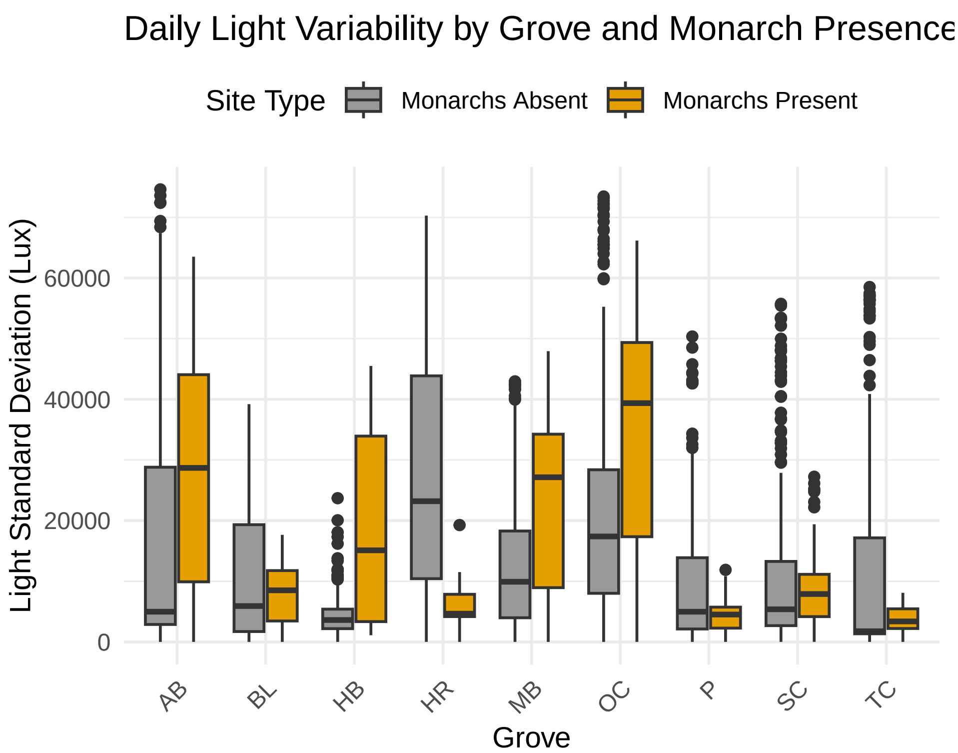

is it clear what the "units" a


Minimum Daily Temperature by Grove and Monarch Pres
Site Type
Monarchs Absent
Monarchs Present
Minimum Temperature (°C)
15
10
5
0
AB
BL
HB
HR
MB
OC
P
SC
TC
Grove


Maximum Daily Temperature by Grove and Monarch Pres
Site Type
Monarchs Absent
Monarchs Present
Maximum Temperature (°C)
40
30
20
10
AB
BL
HB
HR
MB
OC
P
SC
TC
Grove


Average Daily Humidity by Grove and Monarch Presenc
Site Type
Monarchs Absent
Monarchs Present
Average Humidity (%)
100
80
60
40
20
AB
BL
HB
HR
MB
OC
P
SC
TC
Grove


Average Daily Light by Grove and Monarch Presence
Site Type
Monarchs Absent
Monarchs Present
Average Light (Lux)
40000
30000
20000
10000
0
AB
BL
HB
HR
MB
OC
P
SC
TC
Grove




## References

Chaplin, Susan, and Patrick Wells. 1982. "Energy Reserves and Metabolic Expenditures of Monarch Butterflies Overwintering in Southern California." *Ecological Entomology* 7 (3): 249256.

"Endangered and Threatened Wildlife and Plants; Threatened Species Status With Section 4(d) Rule for Monarch Butterfly and Designation of Critical Habitat." 2024. https://www.federalregister.gov/documents/2024/12/12/2024-28855/endangered-and-threatened-wildlife-and-plants-threatened-species-status-with-section-4d-rule-for.

Leong, Kingston L. H., D. Frey, G. Brenner, S. Baker, and D. Fox. 1991. "Use of Multivariate Analyses to Characterize the Monarch Butterfly (Lepidoptera: Danaidae) Winter Habitat." *Annals of the Entomological Society of America* 84 (3): 263–67. https://doi.org/10.1093/aesa/84.3.263.

Saniee, Kiana, and Francis Villablanca. 2022. "Hierarchy and Scale Influence the Western Monarch Butterfly Overwintering Microclimate." *Frontiers in Conservation Science* 3. https://www.frontiersin.org/articles/10.3389/fcosc.2022.844299.

Tuskes, P. M., and L. P. Brower. 1978. "Overwintering Ecology of the Monarch Butterfly, Danaus Plexippus L., in California." *Ecological Entomology* 3 (2): 141–53. https://doi.org/10.1111/j.1365-2311.1978.tb00912.x.

Urquhart, F. A., and N. R. Urquhart. 1978. "Autumnal Migration Routes of the Eastern Population of the Monarch Butterfly (*Danaus p. Plexippus* l.; Danaidae; Lepidoptera) in North America to the Overwintering Site in the Neovolcanic Plateau of Mexico." *Canadian Journal of Zoology* 56 (8): 17591764.

Weiss, Stuart B., Paul M. Rich, Dennis D. Murphy, William H. Calvert, and Paul R. Ehrlich. 1991. "Forest Canopy Structure at Overwintering Monarch Butterfly Sites: Measurements with Hemispherical Photography." *Conservation Biology* 5 (2): 165–75. https://doi.org/10.1111/j.1523-1739.1991.tb00121.x.