Analysis of distruptive winds to overwintering monarch butterflies

Kyle Nessen

2025-09-05

Table of contents

Introduction

This analysis investigates the first hypothesis of my master's thesis: that wind acts as a disruptive force to overwintering monarch butterflies. If true, we predict that monarch abundance at roosts will decrease when exposed to disruptive winds. I use labeled photos from my 2023-2024 dataset to test this hypothesis. I employed GAM (Generalized Additive Models) because they allow for non-linear relationships in fixed effects while maintaining the necessary random effect structure to account for temporal autocorrelation and nested sampling design.

Setup

Load libraries and data:

```
library(tidyverse)
library(mgcv)
library(lubridate)
library(plotly)
library(knitr)
library(DT)
library(here)
# Load the monarch analysis data
monarch_data <- read_csv(here("data", "monarch_analysis_lag30min.csv"))
```

Exploratory Data Analysis

The response variable is the difference in monarch counts between time t and t-1 at 30-minute intervals. I applied a cube root transformation to achieve a more normal distribution. Because the lagged comparisons create overlapping pairs of observations, I include an AR1 autocorrelation structure to account for temporal dependence.

```
knitr::include_graphics("images/clipboard-1435734413.png")
```

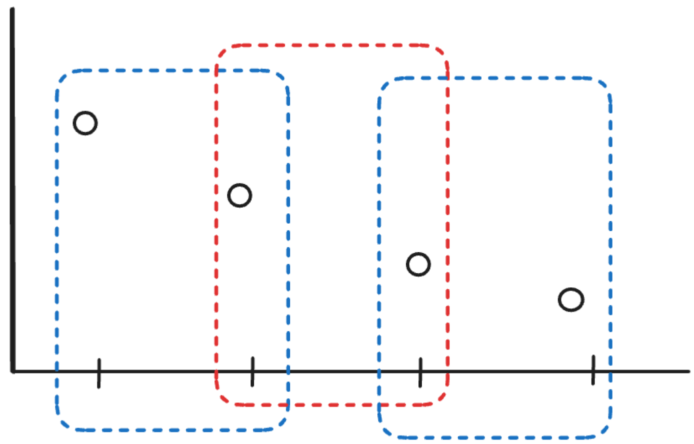
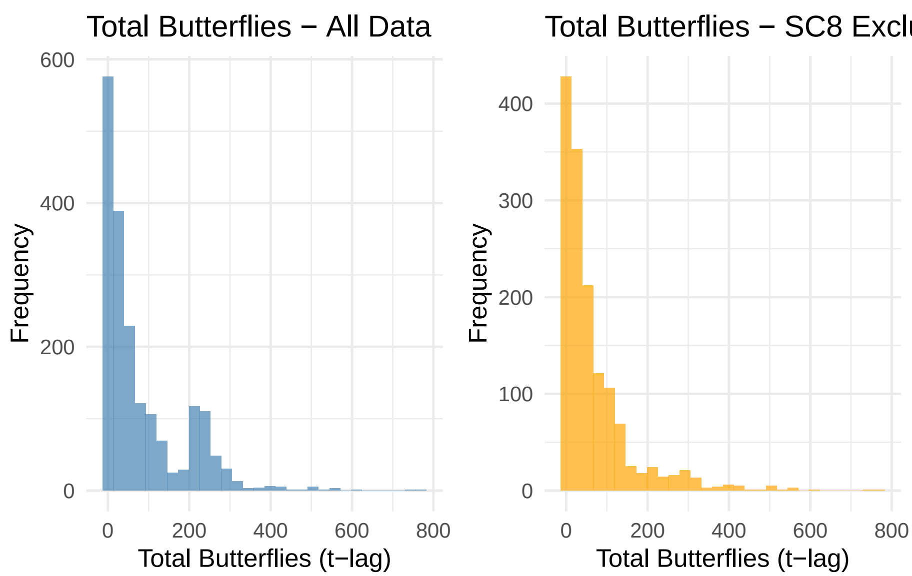


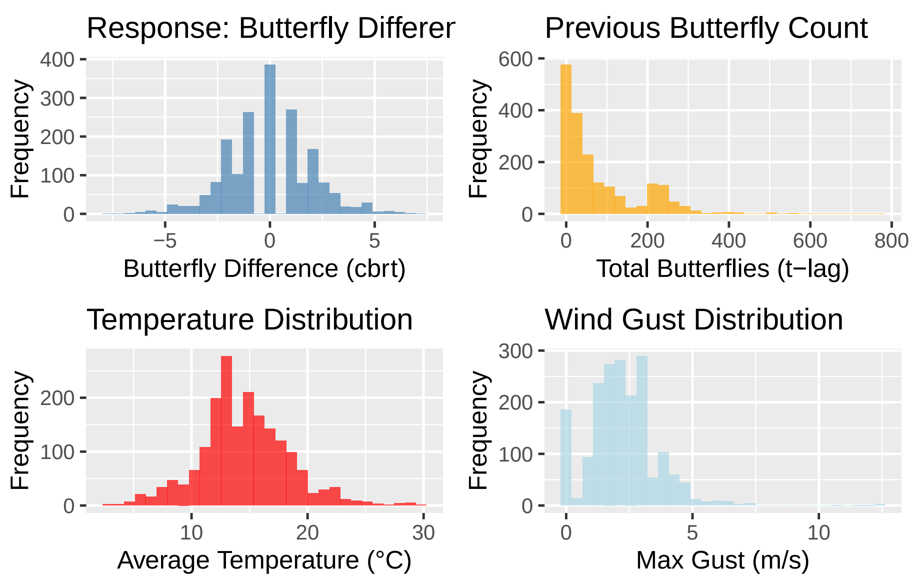
Figure 1: Illustration of temporal dependency in observation pairs. Points represent photos with labeled count data at 30-minute intervals. Blue boxes show non-overlapping pairs of observations. The red box shows an overlapping comparison where one observation is shared between adjacent pairs, creating temporal autocorrelation that is controlled by the AR1 structure.

```
library(gridExtra)

# Compare total butterfly counts with and without SC8
p_all <- ggplot(monarch_data, aes(x = total_butterflies_t_lag)) +
    geom_histogram(bins = 30, fill = "steelblue", alpha = 0.7) +
    labs(
        title = "Total Butterflies - All Data",
        x = "Total Butterflies (t-lag)", y = "Frequency"
    ) +
    theme_minimal()

p_no_sc8 <- ggplot(monarch_data %>% filter(deployment_id != "SC8"), aes(x = total_butterflie
    geom_histogram(bins = 30, fill = "orange", alpha = 0.7) +
    labs(
        title = "Total Butterflies - SC8 Excluded",
        x = "Total Butterflies (t-lag)", y = "Frequency"
    ) +
    theme_minimal()

grid.arrange(p_all, p_no_sc8, ncol = 2)
```



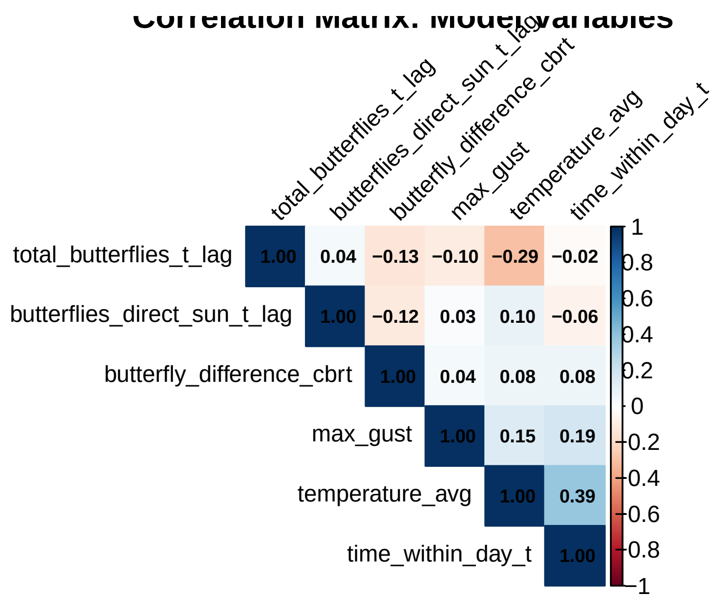

```r
library(corrplot)
library(gridExtra)

# Select variables used in the models
model_vars <- monarch_data %>%
    select(
        butterfly_difference_cbrt, total_butterflies_t_lag, max_gust,
        temperature_avg, butterflies_direct_sun_t_lag, time_within_day_t
    )

# Histograms of key variables
p1 <- ggplot(monarch_data, aes(x = butterfly_difference_cbrt)) +
    geom_histogram(bins = 30, fill = "steelblue", alpha = 0.7) +
    labs(
        title = "Response: Butterfly Difference (Cube Root)",
        x = "Butterfly Difference (cbrt)", y = "Frequency"
    )

p2 <- ggplot(monarch_data, aes(x = total_butterflies_t_lag)) +
    geom_histogram(bins = 30, fill = "orange", alpha = 0.7) +
    labs(
        title = "Previous Butterfly Count",
        x = "Total Butterflies (t-lag)", y = "Frequency"
    )

p3 <- ggplot(monarch_data, aes(x = temperature_avg)) +
```

```r
    geom_histogram(bins = 30, fill = "red", alpha = 0.7) +
    labs(
        title = "Temperature Distribution",
        x = "Average Temperature (°C)", y = "Frequency"
    )

p4 <- ggplot(monarch_data, aes(x = max_gust)) +
    geom_histogram(bins = 30, fill = "lightblue", alpha = 0.7) +
    labs(
        title = "Wind Gust Distribution",
        x = "Max Gust (m/s)", y = "Frequency"
    )

grid.arrange(p1, p2, p3, p4, ncol = 2)
```




```r
# Correlation matrix for model variables
cor_matrix <- cor(model_vars, use = "complete.obs")

# Create correlation plot
corrplot(cor_matrix,
    method = "color",
    type = "upper",
    order = "hclust",
    tl.cex = 0.8,
    tl.col = "black",
    tl.srt = 45,
```

```
    addCoef.col = "black",
    number.cex = 0.7,
    title = "Correlation Matrix: Model Variables"
)
```



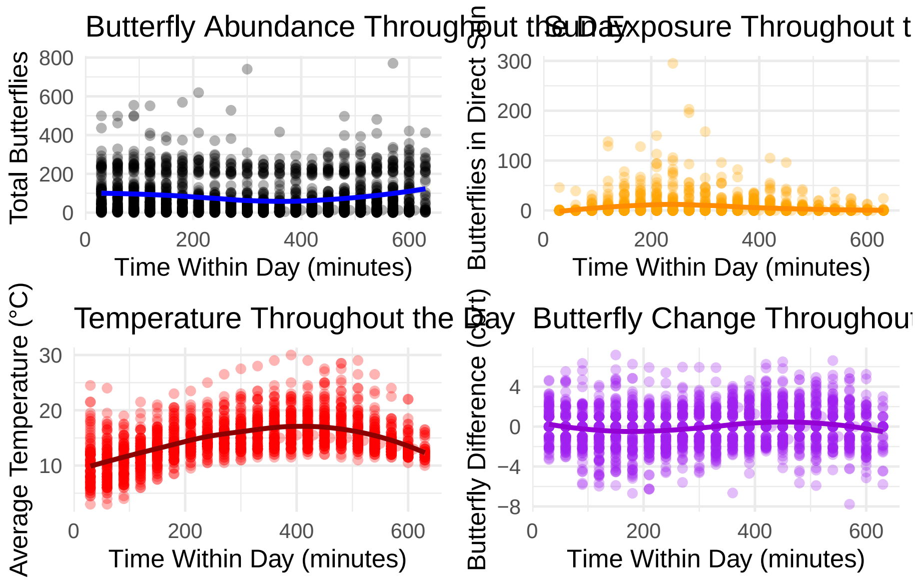

```
# Print correlation table
kable(round(cor_matrix, 3),
    caption = "Correlation Matrix for Model Variables"
)
```

Table 1: Correlation Matrix for Model Variables

	butterfly_difference_cbrt	total_butterflies_t_lag	max_gust	temperature_avg	butterflies_direct_sun_t_lag	time_within_day_t
butterfly_difference_cbrt	1.000	-0.131	0.040	0.079	-0.116	0.077
total_butterflies_t_lag	-0.131	1.000	-0.105	-0.291	0.041	-0.023
max_gust	0.040	-0.105	1.000	0.145	0.027	0.185
temperature_avg	0.079	-0.291	0.145	1.000	0.099	0.386
butterflies_direct_sun_t_lag	-0.116	0.041	0.027	0.099	1.000	-0.064
time_within_day_t	0.077	-0.023	0.185	0.386	-0.064	1.000

```
# Butterfly activity by time of day
p1 <- ggplot(monarch_data, aes(x = time_within_day_t, y = total_butterflies_t_lag)) +
    geom_point(alpha = 0.3) +
```

```r
    geom_smooth(method = "loess", se = TRUE, color = "blue") +
    labs(
        title = "Butterfly Abundance Throughout the Day",
        x = "Time Within Day (minutes)", y = "Total Butterflies"
    ) +
    theme_minimal()

# Sun exposure patterns by time
p2 <- ggplot(monarch_data, aes(x = time_within_day_t, y = butterflies_direct_sun_t_lag)) +
    geom_point(alpha = 0.3, color = "orange") +
    geom_smooth(method = "loess", se = TRUE, color = "darkorange") +
    labs(
        title = "Sun Exposure Throughout the Day",
        x = "Time Within Day (minutes)", y = "Butterflies in Direct Sun"
    ) +
    theme_minimal()

# Temperature patterns by time
p3 <- ggplot(monarch_data, aes(x = time_within_day_t, y = temperature_avg)) +
    geom_point(alpha = 0.3, color = "red") +
    geom_smooth(method = "loess", se = TRUE, color = "darkred") +
    labs(
        title = "Temperature Throughout the Day",
        x = "Time Within Day (minutes)", y = "Average Temperature (°C)"
    ) +
    theme_minimal()

# Response variable by time
p4 <- ggplot(monarch_data, aes(x = time_within_day_t, y = butterfly_difference_cbrt)) +
    geom_point(alpha = 0.3, color = "purple") +
    geom_smooth(method = "loess", se = TRUE, color = "darkviolet") +
    labs(
        title = "Butterfly Change Throughout the Day",
        x = "Time Within Day (minutes)", y = "Butterfly Difference (cbrt)"
    ) +
    theme_minimal()

grid.arrange(p1, p2, p3, p4, ncol = 2)
```



Modeling Strategy

Our modeling approach used a comprehensive AIC-based comparison to evaluate all possible combinations of three key environmental predictors: wind speed (max_gust), temperature (temperature_avg), and solar exposure (butterflies_direct_sun_t_lag). We tested two fundamental modeling frameworks: models that include total_butterflies_t_lag as a control variable (testing effects on relative/proportional change) and models that exclude it (testing effects on absolute change). Within each framework, we systematically evaluated linear main effects, two-way and three-way interactions, and non-linear relationships using smooth terms. We also incorporated time-of-day effects to capture diurnal patterns. This resulted in 47 candidate models that comprehensively explore the parameter space while maintaining proper mixed-effects structure with random effects for deployment, observer, and day, plus AR1 correlation for within-day autocorrelation.

Model Building and Selection

Please expand the code block to see the full list of models tested.

```
library(nlme)

# Define the random effects structure and correlation
random_structure <- list(deployment_id = ~1, Observer = ~1, deployment_day = ~1)
```

```r
correlation_structure <- corAR1(form = ~ observation_order_within_day_t | deployment_day)

# Model specifications for AIC comparison
model_specs <- list(
    # Null model
    "M0_null" = "butterfly_difference_cbrt ~ total_butterflies_t_lag",

    # Single variable models
    "M1_gust" = "butterfly_difference_cbrt ~ total_butterflies_t_lag + max_gust",
    "M2_temp" = "butterfly_difference_cbrt ~ total_butterflies_t_lag + temperature_avg",
    "M3_sun" = "butterfly_difference_cbrt ~ total_butterflies_t_lag + butterflies_direct_sun

    # Two-variable combinations
    "M4_gust_temp" = "butterfly_difference_cbrt ~ total_butterflies_t_lag + max_gust + tempe
    "M5_gust_sun" = "butterfly_difference_cbrt ~ total_butterflies_t_lag + max_gust + butter
    "M6_temp_sun" = "butterfly_difference_cbrt ~ total_butterflies_t_lag + temperature_avg +

    # Three-variable model (main effects only)
    "M7_all_main" = "butterfly_difference_cbrt ~ total_butterflies_t_lag + max_gust + temper

    # Two-way interactions
    "M8_gust_temp_int" = "butterfly_difference_cbrt ~ total_butterflies_t_lag + max_gust * t
    "M9_gust_sun_int" = "butterfly_difference_cbrt ~ total_butterflies_t_lag + max_gust * bu
    "M10_temp_sun_int" = "butterfly_difference_cbrt ~ total_butterflies_t_lag + temperature_

    # Two-way interactions with third variable as main effect
    "M11_gust_temp_int_plus_sun" = "butterfly_difference_cbrt ~ total_butterflies_t_lag + ma
    "M12_gust_sun_int_plus_temp" = "butterfly_difference_cbrt ~ total_butterflies_t_lag + ma
    "M13_temp_sun_int_plus_gust" = "butterfly_difference_cbrt ~ total_butterflies_t_lag + te

    # All two-way interactions
    "M14_all_two_way" = "butterfly_difference_cbrt ~ total_butterflies_t_lag + max_gust * te

    # Three-way interaction
    "M15_three_way" = "butterfly_difference_cbrt ~ total_butterflies_t_lag + max_gust * temp

    # Smooth terms models (with lag term)
    "M16_smooth_temp" = "butterfly_difference_cbrt ~ s(total_butterflies_t_lag) + s(temperat
    "M17_smooth_sun" = "butterfly_difference_cbrt ~ s(total_butterflies_t_lag) + temperature
    "M18_smooth_gust" = "butterfly_difference_cbrt ~ s(total_butterflies_t_lag) + s(max_gust
    "M19_smooth_temp_sun" = "butterfly_difference_cbrt ~ s(total_butterflies_t_lag) + s(temp
    "M20_smooth_all_main" = "butterfly_difference_cbrt ~ s(total_butterflies_t_lag) + s(max_
    "M21_time_of_day" = "butterfly_difference_cbrt ~ s(total_butterflies_t_lag) + temperatur
    "M22_temp_time" = "butterfly_difference_cbrt ~ s(total_butterflies_t_lag) + s(temperatur
    "M23_all_smooth_time" = "butterfly_difference_cbrt ~ s(total_butterflies_t_lag) + s(max_
```

```
    # Models WITHOUT lag term - testing environmental effects on absolute change
    "M24_no_lag_null" = "butterfly_difference_cbrt ~ 1",
    "M25_no_lag_gust" = "butterfly_difference_cbrt ~ max_gust",
    "M26_no_lag_temp" = "butterfly_difference_cbrt ~ temperature_avg",
    "M27_no_lag_sun" = "butterfly_difference_cbrt ~ butterflies_direct_sun_t_lag",
    "M28_no_lag_gust_temp" = "butterfly_difference_cbrt ~ max_gust + temperature_avg",
    "M29_no_lag_gust_sun" = "butterfly_difference_cbrt ~ max_gust + butterflies_direct_sun_t
    "M30_no_lag_temp_sun" = "butterfly_difference_cbrt ~ temperature_avg + butterflies_direc
    "M31_no_lag_all_main" = "butterfly_difference_cbrt ~ max_gust + temperature_avg + butter
    "M32_no_lag_gust_temp_int" = "butterfly_difference_cbrt ~ max_gust * temperature_avg",
    "M33_no_lag_gust_sun_int" = "butterfly_difference_cbrt ~ max_gust * butterflies_direct_s
    "M34_no_lag_temp_sun_int" = "butterfly_difference_cbrt ~ temperature_avg * butterflies_d
    "M35_no_lag_gust_temp_int_plus_sun" = "butterfly_difference_cbrt ~ max_gust * temperatur
    "M36_no_lag_gust_sun_int_plus_temp" = "butterfly_difference_cbrt ~ max_gust * butterflie
    "M37_no_lag_temp_sun_int_plus_gust" = "butterfly_difference_cbrt ~ temperature_avg * but
    "M38_no_lag_all_two_way" = "butterfly_difference_cbrt ~ max_gust * temperature_avg + max
    "M39_no_lag_three_way" = "butterfly_difference_cbrt ~ max_gust * temperature_avg * butte

    # Smooth terms models WITHOUT lag term
    "M40_no_lag_smooth_temp" = "butterfly_difference_cbrt ~ s(temperature_avg) + s(butterfli
    "M41_no_lag_smooth_sun" = "butterfly_difference_cbrt ~ temperature_avg + s(butterflies_d
    "M42_no_lag_smooth_gust" = "butterfly_difference_cbrt ~ s(max_gust) + temperature_avg +
    "M43_no_lag_smooth_temp_sun" = "butterfly_difference_cbrt ~ s(temperature_avg) + s(butte
    "M44_no_lag_smooth_all_main" = "butterfly_difference_cbrt ~ s(max_gust) + s(temperature_
    "M45_no_lag_time_of_day" = "butterfly_difference_cbrt ~ temperature_avg + s(butterflies_
    "M46_no_lag_temp_time" = "butterfly_difference_cbrt ~ s(temperature_avg) + s(butterflies
    "M47_no_lag_all_smooth_time" = "butterfly_difference_cbrt ~ s(max_gust) + s(temperature_
)

cat("Total models to fit:", length(model_specs), "\n")
```

```
Total models to fit: 48
```

Model Fitting

```
# Function to safely fit models
fit_model_safely <- function(formula_str, data) {
    tryCatch(
        {
            formula_obj <- as.formula(formula_str)
            gamm(formula_obj,
                data = data,
```

```
                random = random_structure,
                correlation = correlation_structure,
                method = "REML"
            )
        },
        error = function(e) {
            message("Failed to fit model: ", formula_str)
            message("Error: ", e$message)
            return(NULL)
        }
    )
}

# Fit all models
cat("Fitting models...\n")
```

```
Fitting models...
```

```
fitted_models <- map(model_specs, ~ fit_model_safely(.x, model_data))

# Remove failed models
successful_models <- fitted_models[!map_lgl(fitted_models, is.null)]
cat("Successfully fitted", length(successful_models), "out of", length(model_specs), "models
```

```
Successfully fitted 48 out of 48 models
```

Model Comparison

```
# Extract AIC values
aic_results <- map_dfr(names(successful_models), function(model_name) {
    model <- successful_models[[model_name]]
    data.frame(
        Model = model_name,
        Formula = model_specs[[model_name]],
        AIC = AIC(model$lme),
        LogLik = logLik(model$lme)[1],
        df = attr(logLik(model$lme), "df")
    )
}) %>%
    arrange(AIC) %>%
    mutate(
        Delta_AIC = AIC - min(AIC),
```

```r
        AIC_weight = exp(-0.5 * Delta_AIC) / sum(exp(-0.5 * Delta_AIC))
    )

# Display results
aic_results %>%
    select(Model, AIC, Delta_AIC, AIC_weight, df) %>%
    kable(digits = 3, caption = "Model comparison by AIC")
```

Table 2: Model comparison by AIC

Model	AIC	Delta_AIC	AIC_weight	df
M22_temp_time	8081.848	0.000	0.88	14
M21_time_of_day	8086.644	4.796	0.08	13
M23_all_smooth_time	8088.049	6.200	0.04	16
M46_no_lag_temp_time	8101.296	19.448	0.00	12
M16_smooth_temp	8105.876	24.028	0.00	12
M19_smooth_temp_sun	8105.876	24.028	0.00	12
M47_no_lag_all_smooth_time	8107.724	25.876	0.00	14
M45_no_lag_time_of_day	8108.295	26.447	0.00	11
M20_smooth_all_main	8109.249	27.401	0.00	14
M17_smooth_sun	8114.431	32.583	0.00	11
M18_smooth_gust	8119.075	37.227	0.00	13
M40_no_lag_smooth_temp	8126.061	44.212	0.00	10
M43_no_lag_smooth_temp_sun	8126.061	44.212	0.00	10
M44_no_lag_smooth_all_main	8127.871	46.023	0.00	12
M6_temp_sun	8130.775	48.927	0.00	9
M3_sun	8131.696	49.848	0.00	8
M15_three_way	8132.647	50.799	0.00	14
M5_gust_sun	8134.945	53.097	0.00	9
M11_gust_temp_int_plus_sun	8135.392	53.544	0.00	11
M7_all_main	8136.217	54.369	0.00	10
M39_no_lag_three_way	8137.407	55.559	0.00	13
M41_no_lag_smooth_sun	8139.237	57.389	0.00	9
M9_gust_sun_int	8139.410	57.562	0.00	10
M12_gust_sun_int_plus_temp	8140.795	58.946	0.00	11
M35_no_lag_gust_temp_int_plus_sun	8141.931	60.082	0.00	10
M42_no_lag_smooth_gust	8142.038	60.190	0.00	11
M30_no_lag_temp_sun	8142.927	61.079	0.00	8
M10_temp_sun_int	8144.554	62.705	0.00	10
M31_no_lag_all_main	8146.374	64.526	0.00	9
M36_no_lag_gust_sun_int_plus_temp	8148.813	66.964	0.00	10
M13_temp_sun_int_plus_gust	8150.004	68.156	0.00	11
M0_null	8153.582	71.734	0.00	7
M29_no_lag_gust_sun	8154.129	72.281	0.00	8

Model	AIC	Delta_AIC	AIC_weight	df
M27_no_lag_sun	8155.073	73.225	0.00	7
M14_all_two_way	8155.075	73.227	0.00	13
M34_no_lag_temp_sun_int	8156.678	74.830	0.00	9
M33_no_lag_gust_sun_int	8156.943	75.095	0.00	9
M2_temp	8157.623	75.775	0.00	8
M1_gust	8157.885	76.037	0.00	8
M38_no_lag_all_two_way	8160.095	78.247	0.00	12
M37_no_lag_temp_sun_int_plus_gust	8160.174	78.326	0.00	10
M8_gust_temp_int	8162.939	81.091	0.00	10
M4_gust_temp	8163.059	81.210	0.00	9
M26_no_lag_temp	8170.575	88.727	0.00	7
M32_no_lag_gust_temp_int	8171.945	90.096	0.00	9
M28_no_lag_gust_temp	8175.113	93.264	0.00	8
M24_no_lag_null	8177.191	95.342	0.00	6
M25_no_lag_gust	8178.495	96.647	0.00	7

```
# Show top 5 models
cat("\nTop 5 models by AIC:\n")
```

```
Top 5 models by AIC:
```

```
head(aic_results, 5) %>%
    select(Model, Formula, AIC, Delta_AIC) %>%
    kable(digits = 3)
```

Model	Formula	AIC	Delta_AIC
M22_temp_time	butterfly_difference_cbrt ~ s(total_butterflies_t_lag) + s(temperature_avg) + s(butterflies_direct_sun_t_lag) + s(time_within_day_t)	8081.848	0.000
M21_time_of_day	butterfly_difference_cbrt ~ s(total_butterflies_t_lag) + temperature_avg + s(butterflies_direct_sun_t_lag) + s(time_within_day_t)	8086.644	4.796
M23_all_smooth_time	butterfly_difference_cbrt ~ s(total_butterflies_t_lag) + s(max_gust) + s(temperature_avg) + s(butterflies_direct_sun_t_lag) + s(time_within_day_t)	8088.069	6.220

Model	Formula	AIC	Delta_AIC
M46_no_lag_temp_time	butterfly_difference_cbrt ~ s(temperature_avg) + s(butterflies_direct_sun_t_lag) + s(time_within_day_t)	8101.296	16.448
M16_smooth_temp	butterfly_difference_cbrt ~ s(total_butterflies_t_lag) + s(temperature_avg) + s(butterflies_direct_sun_t_lag)	8105.876	21.028

Best Model Analysis

```r
# Get the best model
best_model_name <- aic_results$Model[1]
best_model <- successful_models[[best_model_name]]

cat("Best model:", best_model_name, "\n")
```

```
Best model: M22_temp_time
```

```r
cat("Formula:", aic_results$Formula[1], "\n\n")
```

```
Formula: butterfly_difference_cbrt ~ s(total_butterflies_t_lag) + s(temperature_avg) + s(but
```

```r
# Model summary
summary(best_model$gam)
```

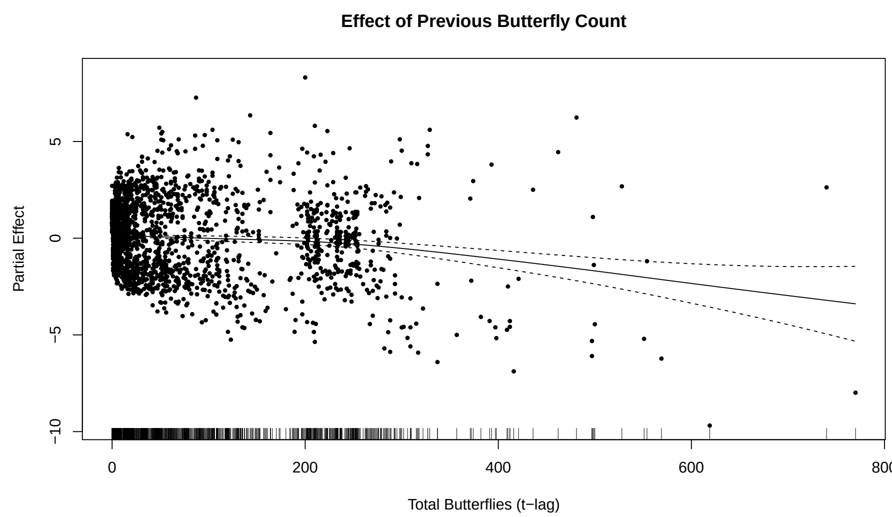
```
Family: gaussian
Link function: identity

Formula:
butterfly_difference_cbrt ~ s(total_butterflies_t_lag) + s(temperature_avg) +
    s(butterflies_direct_sun_t_lag) + s(time_within_day_t)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.1765     0.4453   0.396    0.692

Approximate significance of smooth terms:
                                  edf Ref.df      F  p-value
s(total_butterflies_t_lag)      2.621  2.621 12.020 8.26e-07 ***
s(temperature_avg)              3.930  3.930  3.230   0.0283 *
```

```
s(butterflies_direct_sun_t_lag) 1.534  1.534 19.356 1.22e-05 ***
s(time_within_day_t)            4.898  4.898  8.901  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.0568
  Scale est. = 4.0332    n = 1894
```

Effect Plots

Effect of Previous Butterfly Count

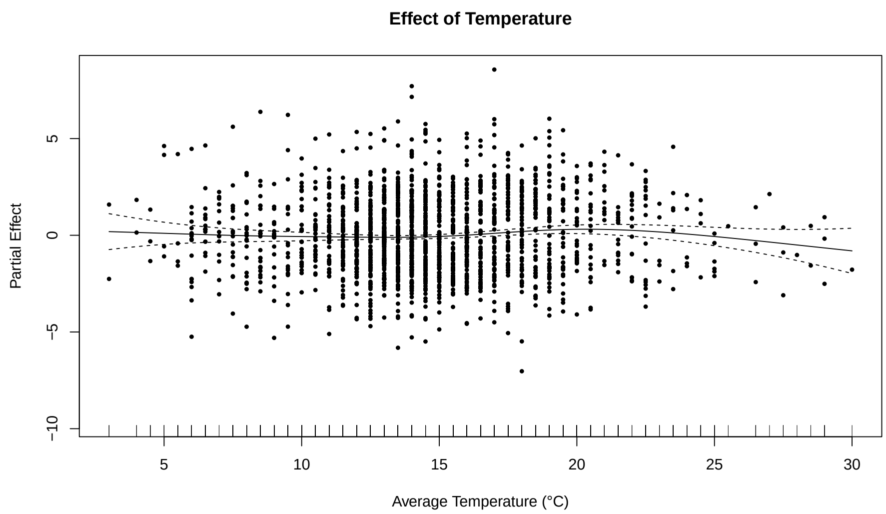
```r
plot(best_model$gam,
    select = 1, main = "Effect of Previous Butterfly Count",
    xlab = "Total Butterflies (t-lag)", ylab = "Partial Effect",
    residuals = TRUE, pch = 19, cex = 0.5
)
```



Effect of Temperature

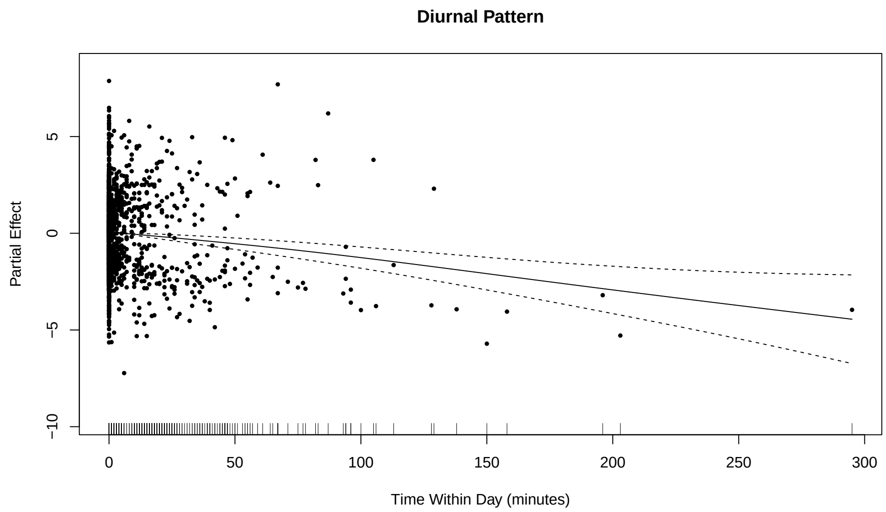
```r
plot(best_model$gam,
    select = 2, main = "Effect of Temperature",
    xlab = "Average Temperature (°C)", ylab = "Partial Effect",
```

```
    residuals = TRUE, pch = 19, cex = 0.5
)
```



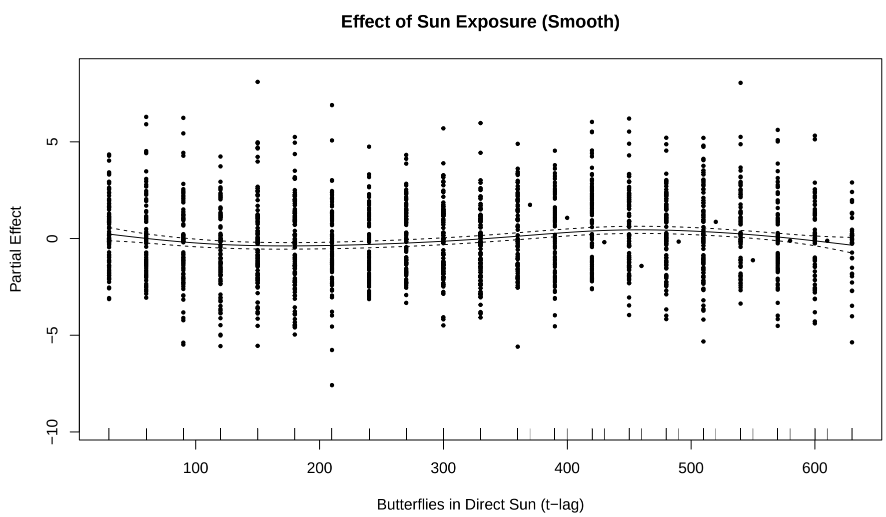
Diurnal Pattern

```
plot(best_model$gam,
    select = 3, main = "Diurnal Pattern",
    xlab = "Time Within Day (minutes)", ylab = "Partial Effect",
    residuals = TRUE, pch = 19, cex = 0.5
)
```



Effect of Sun Exposure (Smooth)

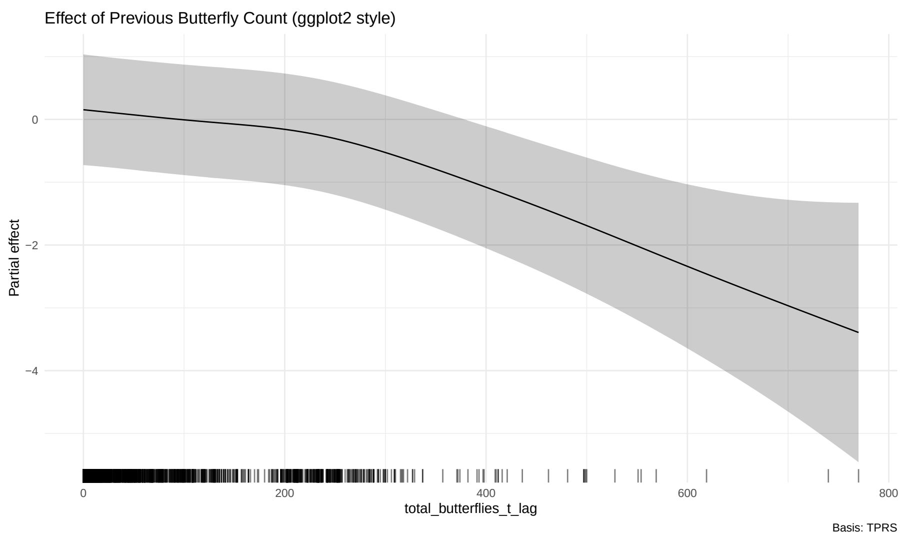
```
# Smooth effect of sun exposure
plot(best_model$gam,
    select = 4, main = "Effect of Sun Exposure (Smooth)",
    xlab = "Butterflies in Direct Sun (t-lag)", ylab = "Partial Effect",
    residuals = TRUE, pch = 19, cex = 0.5
)
```



Smooth Effects (ggplot2 style)

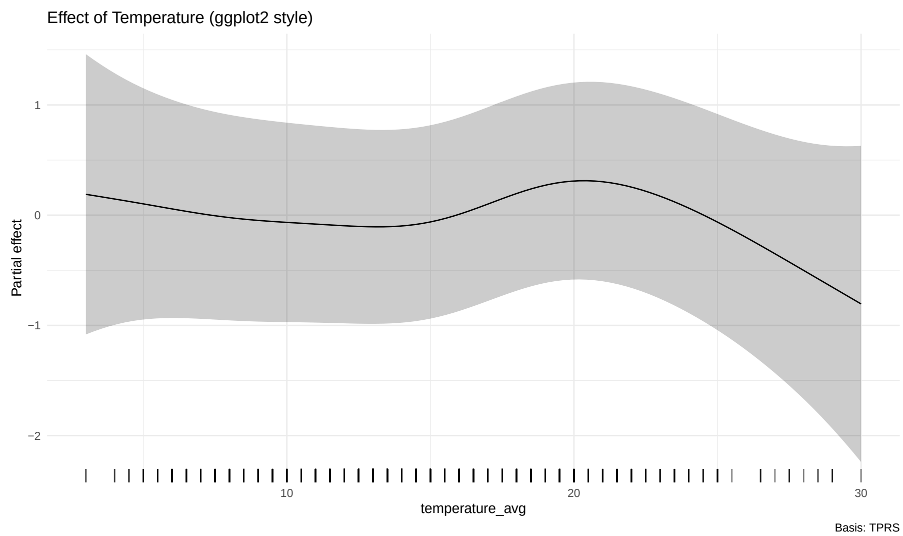
Previous Butterfly Count Effect

```
library(gratia)
draw(best_model$gam, select = "s(total_butterflies_t_lag)") +
    theme_minimal() +
    labs(title = "Effect of Previous Butterfly Count (ggplot2 style)")
```



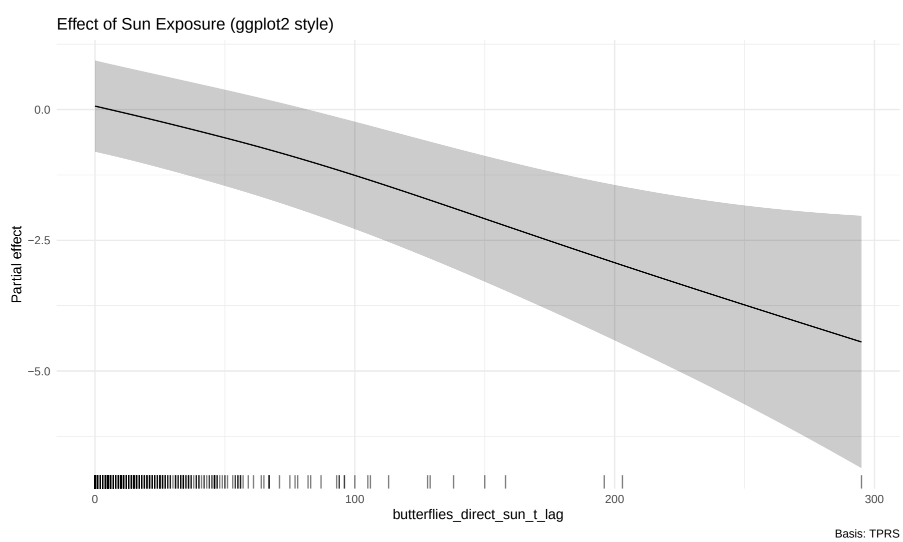
Temperature Effect

```
draw(best_model$gam, select = "s(temperature_avg)") +
    theme_minimal() +
    labs(title = "Effect of Temperature (ggplot2 style)")
```



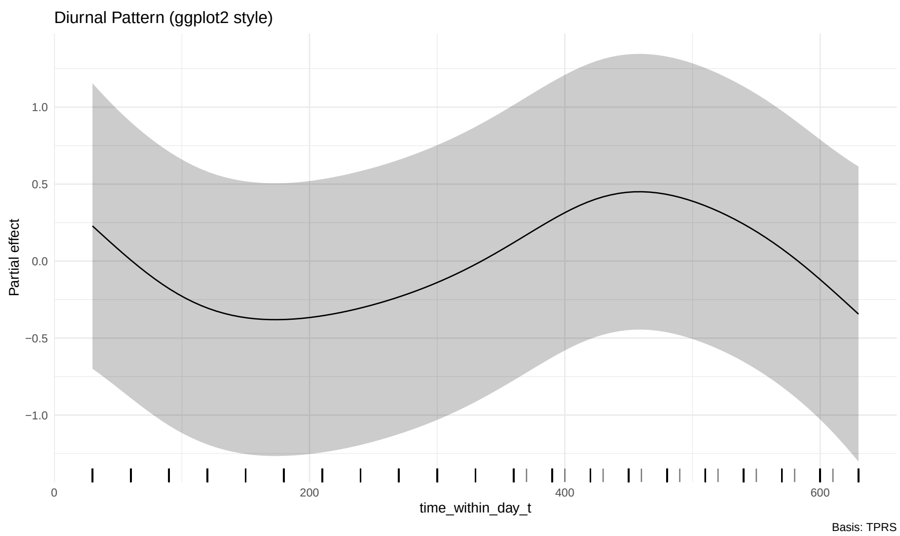
Sun Exposure Effect

```
draw(best_model$gam, select = "s(butterflies_direct_sun_t_lag)") +
    theme_minimal() +
    labs(title = "Effect of Sun Exposure (ggplot2 style)")
```



Diurnal Pattern

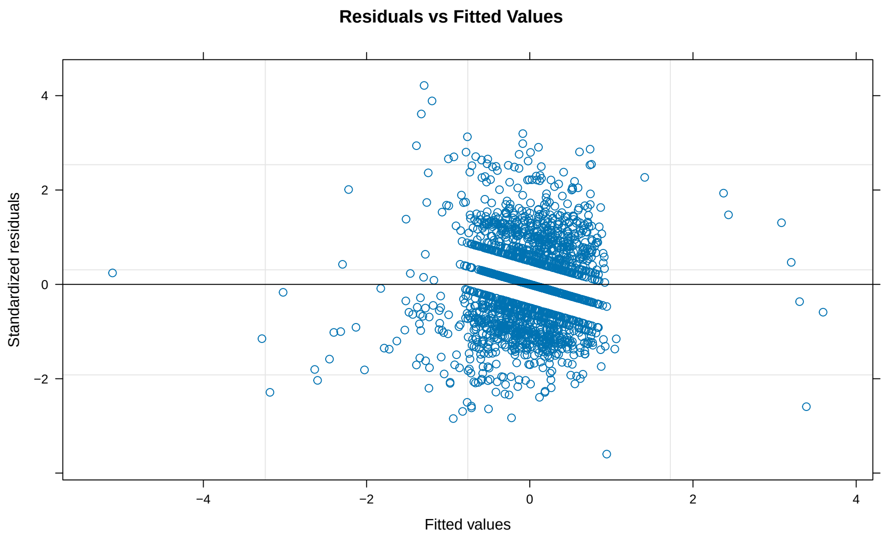
```r
draw(best_model$gam, select = "s(time_within_day_t)") +
    theme_minimal() +
    labs(title = "Diurnal Pattern (ggplot2 style)")
```



Model Diagnostics

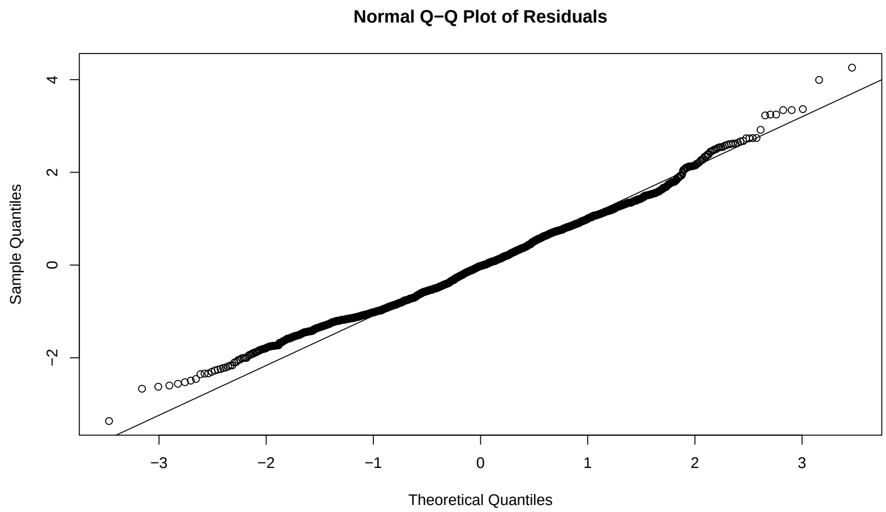
Residuals vs Fitted Values

```r
plot(best_model$lme, main = "Residuals vs Fitted Values")
```



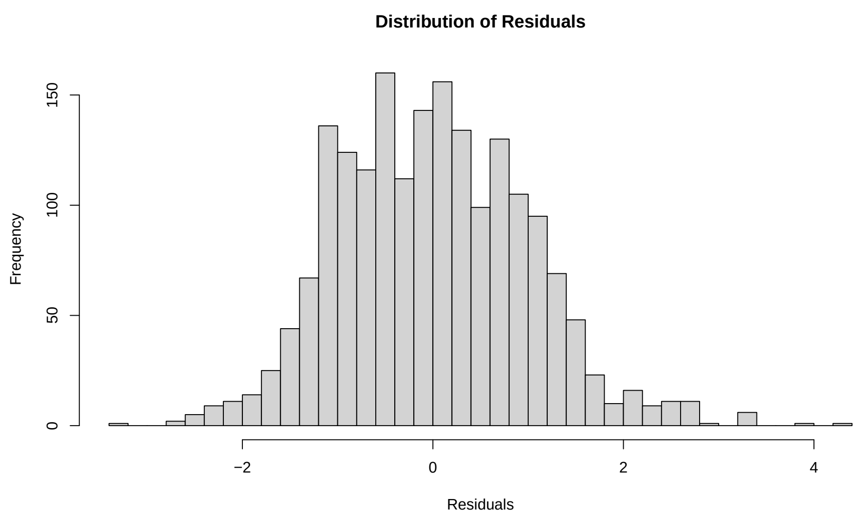
Q-Q Plot of Residuals

```
residuals_df <- data.frame(
    fitted = fitted(best_model$lme),
    residuals = residuals(best_model$lme, type = "normalized")
)

qqnorm(residuals_df$residuals, main = "Normal Q-Q Plot of Residuals")
qqline(residuals_df$residuals)
```



Distribution of Residuals

```
hist(residuals_df$residuals, main = "Distribution of Residuals", xlab = "Residuals", breaks
```



Second Best Model Analysis (Wind)

```
# Get the second best model
second_best_model_name <- aic_results$Model[2]
second_best_model <- successful_models[[second_best_model_name]]

cat("Second best model:", second_best_model_name, "\n")
```

```
Second best model: M21_time_of_day
```

```
cat("Formula:", aic_results$Formula[2], "\n\n")
```

```
Formula: butterfly_difference_cbrt ~ s(total_butterflies_t_lag) + temperature_avg + s(butter
```

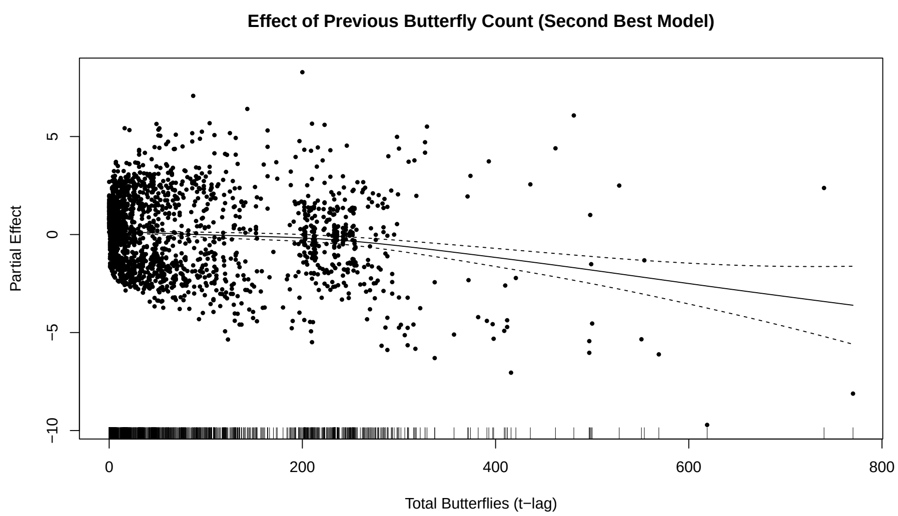
```
# Model summary
summary(second_best_model$gam)
```

```
Family: gaussian
Link function: identity

Formula:
butterfly_difference_cbrt ~ s(total_butterflies_t_lag) + temperature_avg +
    s(butterflies_direct_sun_t_lag) + s(time_within_day_t)

Parametric coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     -0.09618    0.52432  -0.183    0.854
temperature_avg  0.01903    0.01595   1.193    0.233

Approximate significance of smooth terms:
                                  edf Ref.df      F p-value
s(total_butterflies_t_lag)      2.698  2.698 13.127 2.0e-07 ***
s(butterflies_direct_sun_t_lag) 1.637  1.637 18.684 1.5e-05 ***
s(time_within_day_t)            5.023  5.023  9.559 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.0525
  Scale est. = 4.0316    n = 1894
```

Effect Plots - Second Best Model

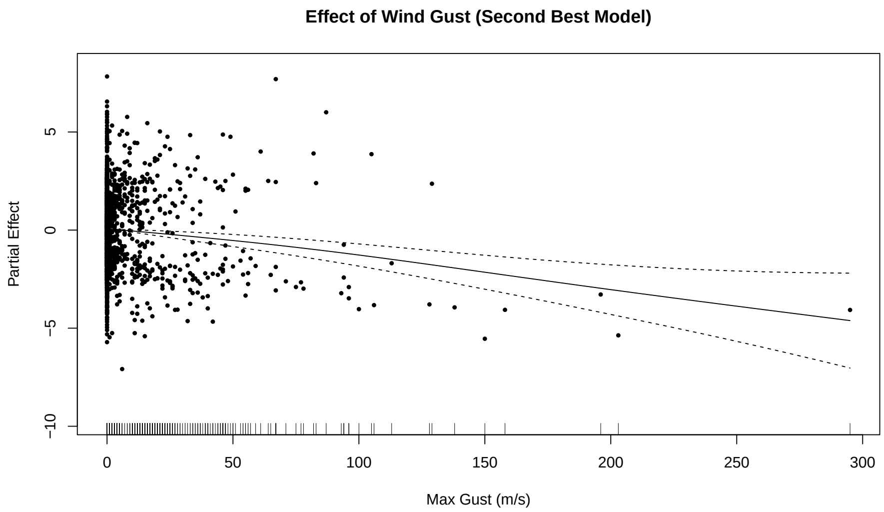
Effect of Previous Butterfly Count

```
plot(second_best_model$gam,
    select = 1, main = "Effect of Previous Butterfly Count (Second Best Model)",
    xlab = "Total Butterflies (t-lag)", ylab = "Partial Effect",
    residuals = TRUE, pch = 19, cex = 0.5
)
```



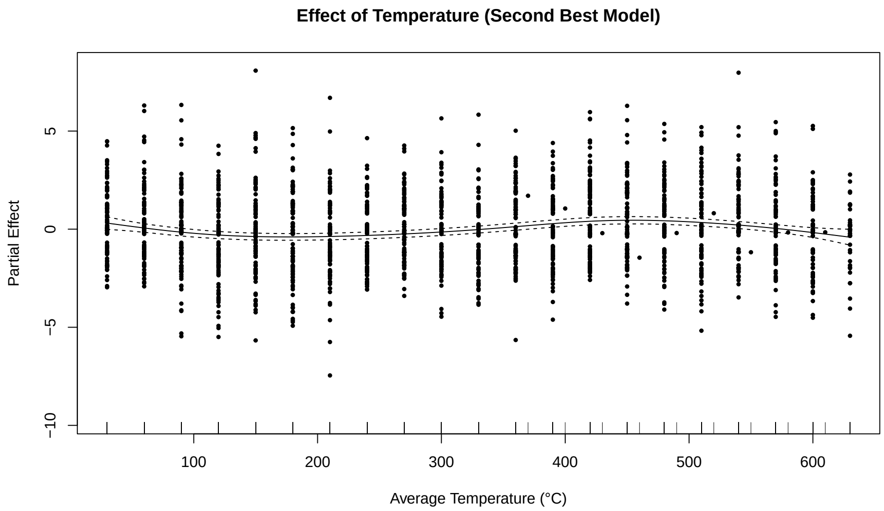
Effect of Wind Gust

```
plot(second_best_model$gam,
    select = 2, main = "Effect of Wind Gust (Second Best Model)",
    xlab = "Max Gust (m/s)", ylab = "Partial Effect",
    residuals = TRUE, pch = 19, cex = 0.5
)
```



Effect of Temperature

```r
plot(second_best_model$gam,
    select = 3, main = "Effect of Temperature (Second Best Model)",
    xlab = "Average Temperature (°C)", ylab = "Partial Effect",
    residuals = TRUE, pch = 19, cex = 0.5
)
```



Effect of Sun Exposure

```
plot(second_best_model$gam,
    select = 4, main = "Effect of Sun Exposure (Second Best Model)",
    xlab = "Butterflies in Direct Sun (t-lag)", ylab = "Partial Effect",
    residuals = TRUE, pch = 19, cex = 0.5
)
```

Diurnal Pattern

```
plot(second_best_model$gam,
    select = 5, main = "Diurnal Pattern (Second Best Model)",
    xlab = "Time Within Day (minutes)", ylab = "Partial Effect",
    residuals = TRUE, pch = 19, cex = 0.5
)
```

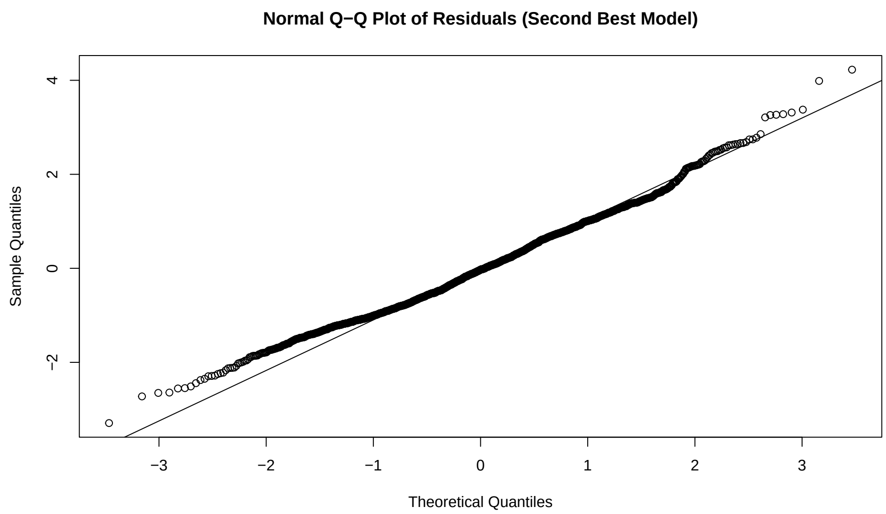
Model Diagnostics - Second Best Model

Residuals vs Fitted Values

```
plot(second_best_model$lme, main = "Residuals vs Fitted Values (Second Best Model)")
```

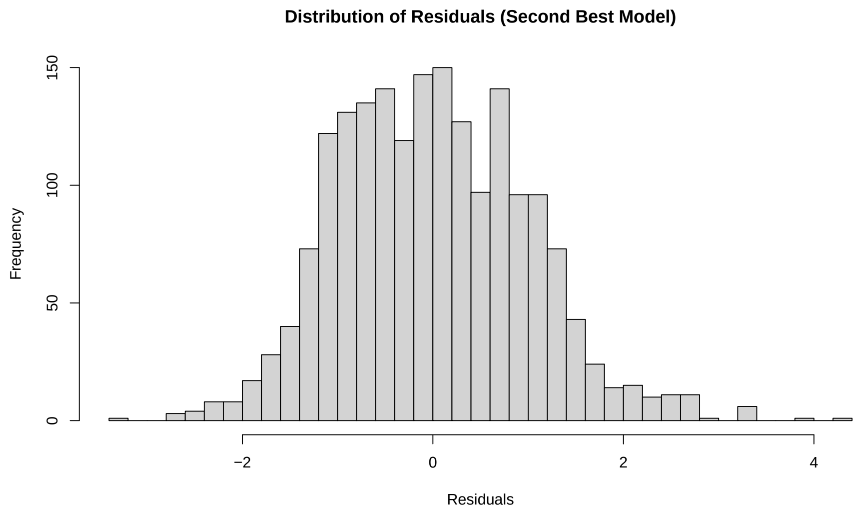
Q-Q Plot of Residuals

```
second_residuals_df <- data.frame(
    fitted = fitted(second_best_model$lme),
    residuals = residuals(second_best_model$lme, type = "normalized")
)

qqnorm(second_residuals_df$residuals, main = "Normal Q-Q Plot of Residuals (Second Best Mode
qqline(second_residuals_df$residuals)
```



Distribution of Residuals

```
hist(second_residuals_df$residuals,
    main = "Distribution of Residuals (Second Best Model)",
    xlab = "Residuals", breaks = 30
)
```



Results Summary

This analysis provides robust evidence regarding wind effects on overwintering monarch butterfly movement through comprehensive model comparison across 47 candidate models. The results reveal several key findings:

Wind Effects: Wind was not selected in the best-performing model and only appeared once in the top 5 models (plotted above) with a non-significant effect ($p = 0.218$). This suggests that wind is not a primary driver of short-term monarch movement patterns at the temporal and spatial scales examined.

Primary Drivers: Temperature and diurnal patterns emerged as the strongest predictors of monarch movement. The best model revealed non-linear temperature responses with apparent thermal optima, and strong diurnal cycles consistent with monarch thermoregulatory behavior.

Model Performance: Including smooth terms substantially improved model fit (R^2 increased from 2.74% to 5.61%), highlighting the importance of capturing non-linear relationships in ecological modeling.

Hypothesis Evaluation: These results do not support the hypothesis that wind acts as a disruptive force to overwintering monarchs at the 30-minute temporal scale examined.