Monarch Butterfly Abundance Analysis: Wind and Environmental Factors

Analysis Report

2025-08-12

Monarch Butterfly Abundance Analysis: LLM Implementation

Part 1: Setup and Data Loading

```
library(tidyverse)
library(lubridate)
library(mgcv)  # for GAM models
library(corrplot)
library(knitr)
```

```
counts <- read_csv("../data/butterfly_abundance_index.csv")
```

```
Rows: 11885 Columns: 4
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (2): deployment_id, image_filename
dbl (2): total_butterflies, butterflies_direct_sun

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
deployments <- read_csv("../data/deployments.csv")
```

```
Rows: 19 Columns: 23
-- Column specification --------------------------------------------------------
```

```
Delimiter: ","
chr (14): camera_name, wind_meter_name, Deployed_time, Recovered_time, notes...
dbl  (8): height_m, horizontal_dist_to_cluster_m, view_direction, cluster_co...
lgl  (1): monarchs_present

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
temp <- read_csv("../data/temperature_data_2023.csv")
```

```
Rows: 56066 Columns: 6
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (3): filename, deployment_id, extraction_status
dbl (3): timestamp, temperature, confidence

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
wind <- read_csv("../data/wind_all.csv")
```

```
Rows: 681013 Columns: 5
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (2): db_path, wind_meter_name
dbl  (2): speed, gust
dttm (1): time

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
cat("Data loaded successfully\n")
```

```
Data loaded successfully
```

```r
cat("Counts rows:", nrow(counts), "\n")
```

```
Counts rows: 11885
```

```r
cat("Deployments rows:", nrow(deployments), "\n")
```

```
Deployments rows: 19
```

```r
cat("Temperature rows:", nrow(temp), "\n")
```

```
Temperature rows: 56066
```

```r
cat("Wind rows:", nrow(wind), "\n")
```

```
Wind rows: 681013
```

Part 2: Initial Data Cleaning and Merging

```r
# Filter and select deployments data
deployments_filtered <- deployments %>%
  filter(label_status == "Complete") %>%
  mutate(view_id = as.factor(view_id)) %>%
  select(deployment_id, view_id, wind_meter_name)

cat("Filtered deployments to", nrow(deployments_filtered), "complete records\n")
```

```
Filtered deployments to 12 complete records
```

```r
# Parse timestamps in counts data
counts_with_datetime <- counts %>%
  mutate(
    datetime_str = str_extract(image_filename, "\\d{14}"),
    datetime = ymd_hms(datetime_str)
  ) %>%
  select(-datetime_str)

cat("Parsed", sum(!is.na(counts_with_datetime$datetime)), "valid timestamps\n")
```

```
Parsed 11885 valid timestamps
```

```r
# Create master dataframe
master_df <- counts_with_datetime %>%
  left_join(deployments_filtered, by = "deployment_id") %>%
  left_join(temp, by = c("image_filename" = "filename"))

cat("Master dataframe created with", nrow(master_df), "rows\n")
```

```
Master dataframe created with 11885 rows
```

```r
cat("Columns:", paste(names(master_df), collapse = ", "), "\n")
```

```
Columns: deployment_id.x, image_filename, total_butterflies, butterflies_direct_sun, datetime
```

```r
# Check for completeness
complete_records <- master_df %>%
  filter(!is.na(view_id) & !is.na(temperature) & !is.na(datetime))
cat("Complete records with all key fields:", nrow(complete_records), "\n")
```

```
Complete records with all key fields: 11885
```

Part 3: The Lagged Data Preparation Function

```r
prepare_lag_data <- function(master_df, wind_df, lag_minutes) {
  cat("Preparing data for", lag_minutes, "minute lag...\n")

  # Create time-shifted dataframes
  df_t <- master_df %>%
    filter(!is.na(datetime) & !is.na(view_id))

  df_t_lag <- df_t %>%
    rename(
      abundance_t_minus_1 = total_butterflies,
      datetime_t_minus_1 = datetime,
      butterflies_sun_t_minus_1 = butterflies_direct_sun,
      temperature_t_minus_1 = temperature
    ) %>%
```

```r
  select(view_id, datetime_t_minus_1, abundance_t_minus_1,
         butterflies_sun_t_minus_1, temperature_t_minus_1)

# Perform sliding window join
df_t_lag <- df_t_lag %>%
  mutate(datetime_expected = datetime_t_minus_1 + minutes(lag_minutes))

final_df <- df_t %>%
  left_join(
    df_t_lag,
    by = c("view_id" = "view_id", "datetime" = "datetime_expected")
  )

# Filter invalid pairs
final_df <- final_df %>%
  filter(!is.na(abundance_t_minus_1)) %>%
  filter(!(total_butterflies == 0 & abundance_t_minus_1 == 0))

cat("Valid pairs after filtering:", nrow(final_df), "\n")

if(nrow(final_df) == 0) {
  cat("No valid pairs found, returning empty dataframe\n")
  return(data.frame())
}

# More efficient calculation using data.table-style operations
cat("Calculating interval predictors efficiently...\n")

# Pre-sort and index the data for faster lookups
wind_sorted <- wind_df %>%
  arrange(wind_meter_name, time) %>%
  mutate(wind_id = row_number())

master_sorted <- master_df %>%
  arrange(view_id, datetime) %>%
  mutate(master_id = row_number())

# Calculate time delta first (fast operation)
final_df <- final_df %>%
  mutate(time_delta_mins = as.numeric(difftime(datetime, datetime_t_minus_1, units = "mins"

# Batch process by unique combinations to reduce redundant calculations
```

```r
unique_intervals <- final_df %>%
  select(view_id, wind_meter_name, datetime_t_minus_1, datetime) %>%
  distinct() %>%
  mutate(interval_id = row_number())

cat("Processing", nrow(unique_intervals), "unique intervals...\n")

# Initialize progress tracking
pb_intervals <- nrow(unique_intervals)

# Calculate metrics for each unique interval
interval_metrics <- list()

for(i in 1:nrow(unique_intervals)) {
  if(i %% max(1, floor(pb_intervals/10)) == 0) {
    cat("Progress:", round(100*i/pb_intervals, 1), "%\n")
  }

  interval <- unique_intervals[i, ]

  # Wind metrics for this interval
  wind_subset <- wind_sorted %>%
    filter(
      wind_meter_name == interval$wind_meter_name,
      time >= interval$datetime_t_minus_1,
      time <= interval$datetime
    )

  if (nrow(wind_subset) > 0) {
    wind_metrics <- list(
      mean_wind_speed = mean(wind_subset$speed, na.rm = TRUE),
      max_wind_speed = max(wind_subset$speed, na.rm = TRUE),
      sd_wind_speed = sd(wind_subset$speed, na.rm = TRUE),
      gust_factor = mean(wind_subset$gust, na.rm = TRUE) / mean(wind_subset$speed, na.rm =
      minutes_above_2mps = sum(wind_subset$speed > 2, na.rm = TRUE)
    )
  } else {
    wind_metrics <- list(
      mean_wind_speed = NA_real_,
      max_wind_speed = NA_real_,
      sd_wind_speed = NA_real_,
      gust_factor = NA_real_,
```

```
      minutes_above_2mps = NA_real_
    )
  }

  # Temperature and sunlight metrics for this interval
  temp_subset <- master_sorted %>%
    filter(
      view_id == interval$view_id,
      datetime >= interval$datetime_t_minus_1,
      datetime <= interval$datetime
    )

  if (nrow(temp_subset) > 0) {
    # Temperature metrics
    temp_metrics <- list(
      mean_temp = mean(temp_subset$temperature, na.rm = TRUE),
      max_temp = max(temp_subset$temperature, na.rm = TRUE),
      min_temp = min(temp_subset$temperature, na.rm = TRUE),
      sd_temp = sd(temp_subset$temperature, na.rm = TRUE)
    )

    # Sunlight metrics
    total_sun <- sum(temp_subset$butterflies_direct_sun, na.rm = TRUE)
    total_count <- sum(temp_subset$total_butterflies, na.rm = TRUE)
    if (total_count > 0) {
      overall_prop <- total_sun / total_count
      obs_props <- ifelse(temp_subset$total_butterflies > 0,
                          temp_subset$butterflies_direct_sun / temp_subset$total_butterflies
                          NA_real_)
      sun_metrics <- list(
        sunlight_proportion = overall_prop,
        max_prop_sunlight = max(obs_props, na.rm = TRUE),
        min_prop_sunlight = min(obs_props, na.rm = TRUE),
        sd_prop_sunlight = sd(obs_props, na.rm = TRUE)
      )
    } else {
      sun_metrics <- list(
        sunlight_proportion = NA_real_,
        max_prop_sunlight = NA_real_,
        min_prop_sunlight = NA_real_,
        sd_prop_sunlight = NA_real_
      )
```

```
    }
  } else {
    temp_metrics <- list(
      mean_temp = NA_real_,
      max_temp = NA_real_,
      min_temp = NA_real_,
      sd_temp = NA_real_
    )
    sun_metrics <- list(
      sunlight_proportion = NA_real_,
      max_prop_sunlight = NA_real_,
      min_prop_sunlight = NA_real_,
      sd_prop_sunlight = NA_real_
    )
  }

  # Store all metrics for this interval
  interval_metrics[[i]] <- c(
    interval_id = interval$interval_id,
    wind_metrics,
    temp_metrics,
    sun_metrics
  )
}

# Convert to dataframe
cat("Combining results...\n")
metrics_df <- do.call(rbind, lapply(interval_metrics, as.data.frame))

# Join back with original data
final_df_with_intervals <- final_df %>%
  left_join(unique_intervals, by = c("view_id", "wind_meter_name", "datetime_t_minus_1", "d
  left_join(metrics_df, by = "interval_id") %>%
  select(-interval_id)

# Remove rows with too many NAs in predictors
final_clean <- final_df_with_intervals %>%
  filter(!is.na(mean_wind_speed) & !is.na(mean_temp))

cat("Final dataset rows:", nrow(final_clean), "\n")
if(nrow(final_clean) > 0) {
  cat("Time delta range:", round(min(final_clean$time_delta_mins, na.rm = TRUE), 1),
```

```
        "to", round(max(final_clean$time_delta_mins, na.rm = TRUE), 1), "minutes\n")
  }

  return(final_clean)
}
```

Part 4: Analysis and Modeling

30-Minute Lag Analysis

```
data_30m <- prepare_lag_data(master_df, wind, lag_minutes = 30)
```

```
Preparing data for 30 minute lag...
Valid pairs after filtering: 6888
Calculating interval predictors efficiently...
Processing 6888 unique intervals...
Progress: 10 %
Progress: 20 %
Progress: 30 %
Progress: 40 %
Progress: 49.9 %
Progress: 59.9 %
Progress: 69.9 %
Progress: 79.9 %
Progress: 89.9 %
Progress: 99.9 %
Combining results...
Final dataset rows: 5601
Time delta range: 30 to 30 minutes
```

Exploratory Data Analysis (30 minutes)

```
cat("Data structure:\n")
```

```
Data structure:
```

```r
glimpse(data_30m)
```

```
Rows: 5,601
Columns: 30
$ deployment_id.x           <chr> "SC1", "SC1", "SC1", "SC1", "SC1", "SC1", "S~
$ image_filename            <chr> "SC1_20231117184501.JPG", "SC1_2023111718500~
$ total_butterflies         <dbl> 54, 54, 54, 54, 54, 54, 54, 54, 54, 54, 54, ~
$ butterflies_direct_sun    <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ datetime                  <dttm> 2023-11-17 18:45:01, 2023-11-17 18:50:01, 2~
$ view_id                   <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ wind_meter_name           <chr> "JazzPlay", "JazzPlay", "JazzPlay", "JazzPla~
$ deployment_id.y           <chr> "SC1", "SC1", "SC1", "SC1", "SC1", "SC1", "S~
$ timestamp                 <dbl> 2.023112e+13, 2.023112e+13, 2.023112e+13, 2.~
$ temperature               <dbl> 15, 16, 15, 16, 15, 16, 16, 15, 16, 15, 16, ~
$ confidence                <dbl> 0.4041873, 0.5208792, 0.6782328, 0.5311629, ~
$ extraction_status         <chr> "success", "success", "success", "success", ~
$ datetime_t_minus_1        <dttm> 2023-11-17 18:15:01, 2023-11-17 18:20:01, 2~
$ abundance_t_minus_1       <dbl> 54, 54, 54, 54, 54, 54, 54, 54, 54, 54, 54, ~
$ butterflies_sun_t_minus_1 <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ temperature_t_minus_1     <dbl> 15, 15, 15, 15, 15, 15, 15, 16, 15, 16, 15, ~
$ time_delta_mins           <dbl> 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, ~
$ mean_wind_speed           <dbl> 0.400000, 0.900000, 1.030769, 1.177778, 1.24~
$ max_wind_speed            <dbl> 0.4, 1.9, 1.9, 1.9, 1.9, 2.3, 2.7, 2.9, 2.9,~
$ sd_wind_speed             <dbl> 0.0000000, 0.6524678, 0.5297556, 0.5241863, ~
$ gust_factor               <dbl> 2.250000, 1.722222, 1.656716, 1.622642, 1.58~
$ minutes_above_2mps        <dbl> 0, 0, 0, 0, 0, 1, 3, 6, 7, 12, 16, 19, 22, 2~
$ mean_temp                 <dbl> 15.00000, 15.14286, 15.14286, 15.28571, 15.2~
$ max_temp                  <dbl> 15, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16, ~
$ min_temp                  <dbl> 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, ~
$ sd_temp                   <dbl> 0.0000000, 0.3779645, 0.3779645, 0.4879500, ~
$ sunlight_proportion       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ max_prop_sunlight         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ min_prop_sunlight         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ sd_prop_sunlight          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
```

```r
cat("\n\nData summary:\n")
```

```
Data summary:
```

```r
summary(data_30m)
```

```
 deployment_id.x    image_filename     total_butterflies butterflies_direct_sun
 Length:5601        Length:5601        Min.   :  0.00    Min.   :  0.000
 Class :character   Class :character   1st Qu.: 14.00    1st Qu.:  0.000
 Mode  :character   Mode  :character   Median : 35.00    Median :  0.000
                                        Mean   : 66.52    Mean   :  1.726
                                        3rd Qu.: 73.00    3rd Qu.:  0.000
                                        Max.   :950.00    Max.   :150.000

    datetime                          view_id     wind_meter_name
 Min.   :2023-11-17 18:45:01.0   4      :2941   Length:5601
 1st Qu.:2023-12-11 22:00:01.0   2      :1039   Class :character
 Median :2023-12-30 03:50:01.0   1      : 640   Mode  :character
 Mean   :2023-12-29 12:47:03.9   5      : 444
 3rd Qu.:2024-01-23 16:50:01.0   11     : 288
 Max.   :2024-02-04 10:30:01.0   9      : 173
                                 (Other):  76
 deployment_id.y      timestamp          temperature     confidence
 Length:5601        Min.   :2.023e+13   Min.   : 4.0   Min.   :0.1222
 Class :character   1st Qu.:2.023e+13   1st Qu.:11.0   1st Qu.:0.4982
 Mode  :character   Median :2.023e+13   Median :13.0   Median :0.5951
                    Mean   :2.024e+13   Mean   :13.2   Mean   :0.5973
                    3rd Qu.:2.024e+13   3rd Qu.:15.0   3rd Qu.:0.6967
                    Max.   :2.024e+13   Max.   :33.0   Max.   :1.0000

 extraction_status  datetime_t_minus_1              abundance_t_minus_1
 Length:5601        Min.   :2023-11-17 18:15:01.0   Min.   :  0.00
 Class :character   1st Qu.:2023-12-11 21:30:01.0   1st Qu.: 14.00
 Mode  :character   Median :2023-12-30 03:20:01.0   Median : 35.00
                    Mean   :2023-12-29 12:17:03.9   Mean   : 66.69
                    3rd Qu.:2024-01-23 16:20:01.0   3rd Qu.: 73.00
                    Max.   :2024-02-04 10:00:01.0   Max.   :950.00

 butterflies_sun_t_minus_1 temperature_t_minus_1 time_delta_mins
 Min.   :  0.0             Min.   : 4.0          Min.   :30
 1st Qu.:  0.0             1st Qu.:11.0          1st Qu.:30
 Median :  0.0             Median :13.0          Median :30
 Mean   :  1.7             Mean   :13.2          Mean   :30
 3rd Qu.:  0.0             3rd Qu.:15.0          3rd Qu.:30
 Max.   :150.0             Max.   :33.0          Max.   :30
```

```
 mean_wind_speed  max_wind_speed  sd_wind_speed     gust_factor
 Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :1.149
 1st Qu.:0.1567   1st Qu.:0.700   1st Qu.:0.1029   1st Qu.:1.346
 Median :0.7700   Median :1.200   Median :0.2312   Median :1.441
 Mean   :0.8902   Mean   :1.378   Mean   :0.2383   Mean   :  Inf
 3rd Qu.:1.3400   3rd Qu.:2.000   3rd Qu.:0.3354   3rd Qu.:1.590
 Max.   :5.3333   Max.   :6.700   Max.   :1.5024   Max.   :  Inf
                                  NA's   :1        NA's   :890
 minutes_above_2mps   mean_temp       max_temp       min_temp
 Min.   : 0.000     Min.   : 4.0   Min.   : 4.0   Min.   : 4.00
 1st Qu.: 0.000     1st Qu.:11.0   1st Qu.:11.0   1st Qu.:11.00
 Median : 0.000     Median :13.0   Median :13.0   Median :13.00
 Mean   : 2.949     Mean   :13.2   Mean   :13.5   Mean   :12.89
 3rd Qu.: 0.000     3rd Qu.:15.5   3rd Qu.:16.0   3rd Qu.:15.00
 Max.   :30.000     Max.   :33.0   Max.   :33.0   Max.   :33.00

    sd_temp       sunlight_proportion max_prop_sunlight min_prop_sunlight
 Min.   :0.0000   Min.   :0.00000     Min.   :0.000     Min.   :0.00000
 1st Qu.:0.0000   1st Qu.:0.00000     1st Qu.:0.000     1st Qu.:0.00000
 Median :0.0000   Median :0.00000     Median :0.000     Median :0.00000
 Mean   :0.3613   Mean   :0.07344     Mean   :0.104     Mean   :0.05038
 3rd Qu.:0.7071   3rd Qu.:0.00000     3rd Qu.:0.000     3rd Qu.:0.00000
 Max.   :4.9497   Max.   :1.00000     Max.   :1.000     Max.   :1.00000

 sd_prop_sunlight
 Min.   :0.00000
 1st Qu.:0.00000
 Median :0.00000
 Mean   :0.03026
 3rd Qu.:0.00000
 Max.   :0.70711
 NA's   :60
```

```r
# Check for data quality issues
cat("\n\nData quality checks:\n")
```

```
Data quality checks:
```

```r
cat("Total rows:", nrow(data_30m), "\n")
```

```
Total rows: 5601
```

```r
cat("Complete cases:", sum(complete.cases(data_30m)), "\n")
```

```
Complete cases: 4679
```

```r
cat("Rows with abundance_t_minus_1 > 0:", sum(data_30m$abundance_t_minus_1 > 0, na.rm = TRUE)
```

```
Rows with abundance_t_minus_1 > 0: 5566
```

```r
cat("Rows with total_butterflies > 0:", sum(data_30m$total_butterflies > 0, na.rm = TRUE), "
```

```
Rows with total_butterflies > 0: 5547
```

```r
cat("Unique view_ids:", length(unique(data_30m$view_id)), "\n")
```

```
Unique view_ids: 9
```

```r
# Check response variable distribution
cat("\nResponse variable (total_butterflies) distribution:\n")
```

```

Response variable (total_butterflies) distribution:
```

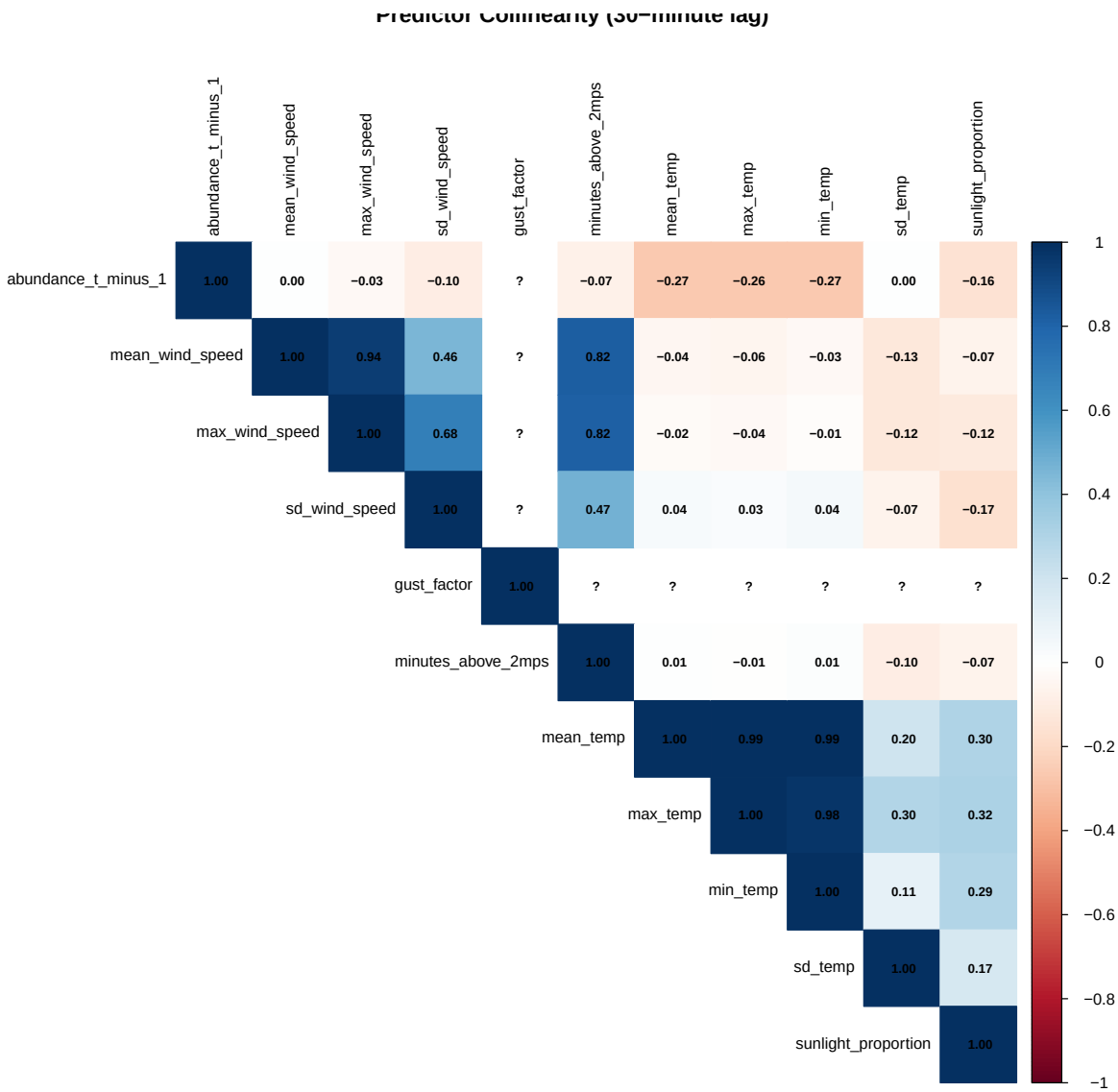
```r
table(data_30m$total_butterflies)
```

```
  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19
 54 125 188 225 109 146 103  43  78  65  24  54  57  65 111 197 164  71  39 171
 20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39
 13  43  20  52 134  30  89  14  14   8  12  16  16  19 194 142  16  61   6  35
 40  41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59
 80  38  17  14  36  22  18  19  66  12  44  55 168  11 149  68  10  30   6   7
 60  61  62  63  64  65  66  67  68  69  70  71  72  73  74  75  76  77  78  79
  2  11   6  10   5   5   8 166   4   4   3   3  80   9   5  61   9   4  32   8
```

```
 80  81  82  83  84  85  86  87  88  89  90  91  92  93  94  95  96  97  98  99
  3   5   5   6   6  33   4   5   3   1 136  32   7   6  10   6   4   3  29   6
100 101 102 103 104 105 106 108 109 110 111 112 113 114 115 116 117 118 119 120
  3   5   4  10   5   8   7   7  28   1   5   1   3   1   3   1   2   3   4   4
121 123 124 125 126 127 128 129 130 131 132 133 134 135 137 139 140 141 143 144
  2   2   2   1   2   1   1  58   5   6   1   3   1   4   1   2   2   3   2   1
146 148 149 150 152 153 157 159 160 162 164 165 168 170 174 184 185 187 188 191
  2   1   2   1   4   4   2   1   1   1   3   1  26   1   1   1   1   1   1   1
192 193 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213
  2   3   4   3   2   2   1   2  10   9  26   3  29   1   2   4  33  33  36  14
214 215 216 217 218 222 223 225 226 227 228 229 230 231 232 233 234 236 237 238
  1   3   3   2   1   2   5  30   3   1   2   4   1   3   4   9  12  57   4   2
242 243 244 245 247 248 249 250 251 252 253 254 255 256 257 262 263 264 266 267
 34  32   2   4   3   2   4   4   2  32  31   7   2   2   2   1  30   2   1   1
268 269 270 271 272 274 276 284 285 286 288 293 315 318 322 329 337 352 371 372
  1   1   2  28   1   1   3  30   1   2   3   1   1   1   1   1   1   1   1   1
374 497 499 528 551 569 619 740 950
  1   1  25   1   1   1   1   1  22
```

```r
# Select numeric predictors for correlation analysis
predictors_30m <- data_30m %>%
  select(abundance_t_minus_1, mean_wind_speed, max_wind_speed, sd_wind_speed,
         gust_factor, minutes_above_2mps, mean_temp, max_temp, min_temp, sd_temp,
         sunlight_proportion) %>%
  na.omit()

if (nrow(predictors_30m) > 1) {
  # Calculate correlation matrix
  cor_matrix_30m <- cor(predictors_30m, use = "complete.obs")

  # Check if correlation matrix has any NA/NaN/Inf values
  if (!any(is.na(cor_matrix_30m) | is.infinite(cor_matrix_30m))) {
    corrplot(cor_matrix_30m, method = "color", type = "upper",
             order = "hclust", tl.cex = 0.8, tl.col = "black",
             addCoef.col = "black", number.cex = 0.7,
             title = "Predictor Collinearity (30-minute lag)")
  } else {
    # If hclust fails, use original order
    corrplot(cor_matrix_30m, method = "color", type = "upper",
             order = "original", tl.cex = 0.8, tl.col = "black",
             addCoef.col = "black", number.cex = 0.7,
             title = "Predictor Collinearity (30-minute lag)")
```

```
  }
} else {
  cat("Not enough data for correlation analysis\n")
}
```



Model Fitting (30 minutes)

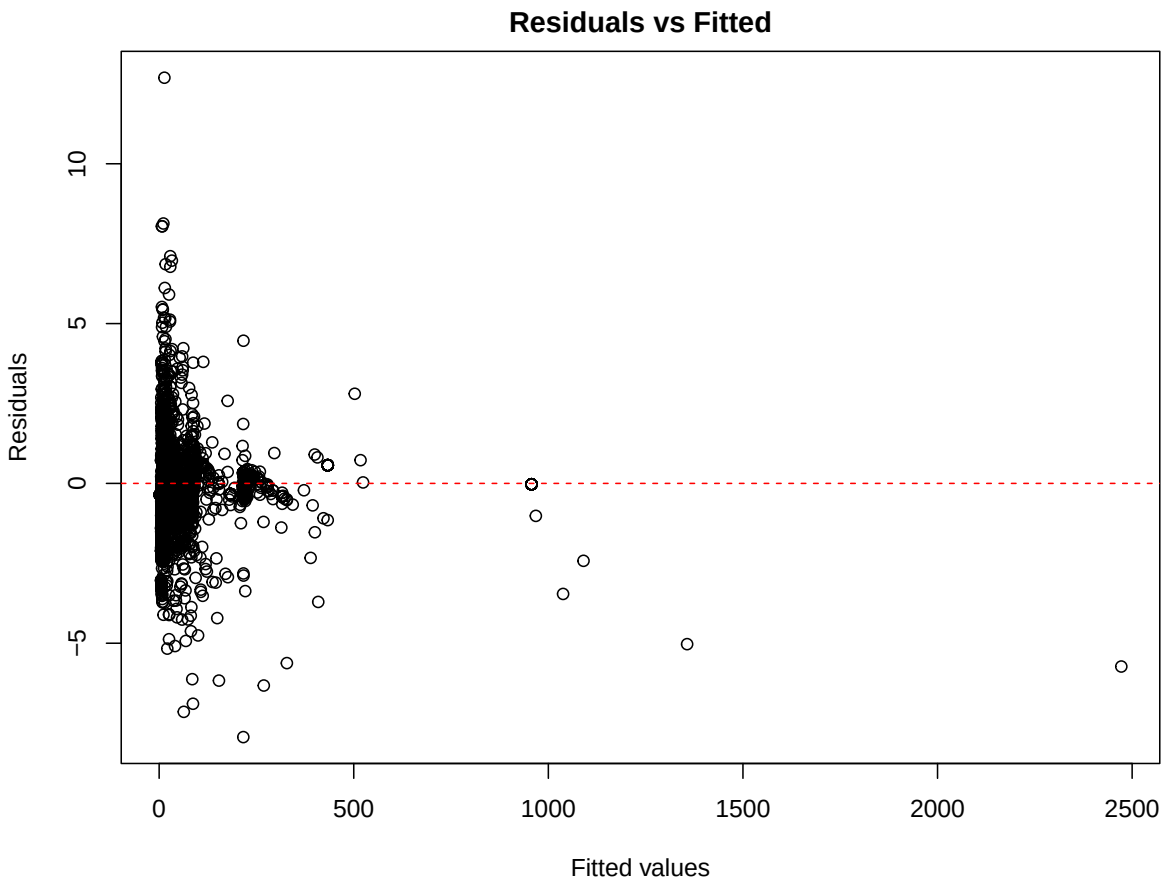
```
# Define comprehensive model formulas
model_formulas_30m <- list(
  # 1. Baseline Model
  "1. Baseline" = total_butterflies ~ s(abundance_t_minus_1) + s(view_id, bs = "re"),

  # 2. Single Environmental Factor Models
  "2a. Wind only" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(view_
  "2b. Temperature only" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(vie
  "2c. Sunlight only" = total_butterflies ~ s(abundance_t_minus_1) + s(sunlight_proportion) +

  # 3. Additive Models (No Interactions)
  "3a. Wind + Temp" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(mea
  "3b. Wind + Sunlight" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s
  "3c. Temp + Sunlight" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(sunli
  "3d. All additive" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(me

  # 4. Single Interaction Models
  "4a. Wind x Temp interaction" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_sp
  "4b. Wind x Sunlight interaction" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_win
  "4c. Temp x Sunlight interaction" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_win

  # 5. Multiple Interaction Models
  "5a. Wind x Temp + Wind x Sun" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_s
  "5b. Wind x Temp + Temp x Sun" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_s
  "5c. Wind x Sun + Temp x Sun" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_sp

  # 6. Global Model with All Interactions
  "6. Global model" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(mea
)

# Storage for results
models_30m <- list()
aic_values_30m <- numeric()

# Fit each model
for (i in seq_along(model_formulas_30m)) {
  model_name <- names(model_formulas_30m)[i]
  formula <- model_formulas_30m[[i]]

  cat("\n### Model", i, ":", model_name, "\n\n")
```

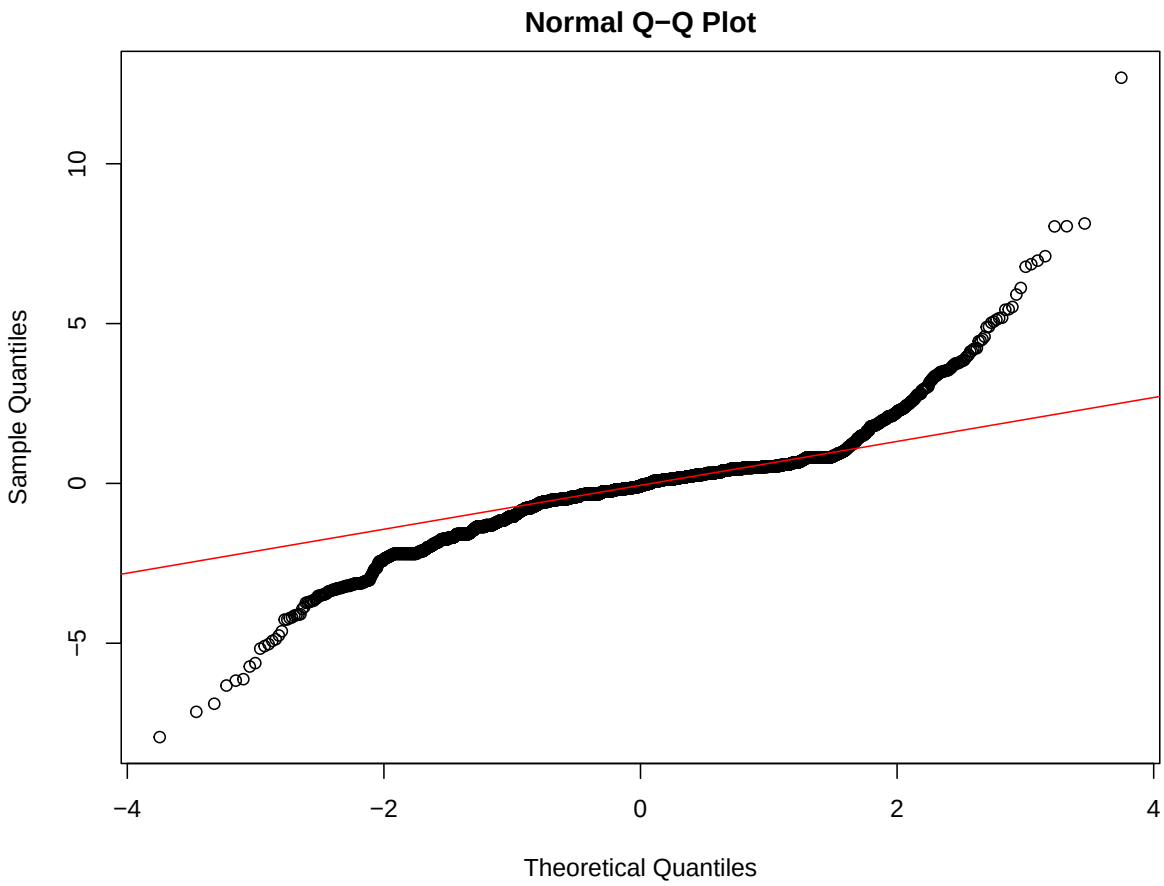
```r
tryCatch({
  # Check data before fitting
  model_vars <- all.vars(formula)
  model_data <- data_30m %>%
    select(all_of(model_vars), view_id) %>%
    na.omit()

  cat("Model data dimensions after removing NAs:", nrow(model_data), "x", ncol(model_data)
  cat("Variables in model:", paste(model_vars, collapse = ", "), "\n")

  # Check if we have enough data
  if(nrow(model_data) < 50) {
    cat("Warning: Very few observations (", nrow(model_data), ") for this model\n")
  }

  # Check for sufficient variation in view_id
  view_counts <- table(model_data$view_id)
  cat("View IDs with observations:", length(view_counts), "\n")
  cat("Min/Max observations per view:", min(view_counts), "/", max(view_counts), "\n")

  # Fit GAM (simpler and faster than GAMM)
  gam_model <- gam(
    formula,
    family = nb(),
    data = model_data
  )

  cat("GAM fitted successfully\n")

  # Store model
  models_30m[[model_name]] <- gam_model
  aic_values_30m[model_name] <- AIC(gam_model)

  # Print summary
  cat("Model Summary:\n")
  print(summary(gam_model))

  # Diagnostic plots - individual plots for better visibility in PDF
  cat("\n\nDiagnostic plots:\n")

  # Set up for individual plots
  par(mfrow = c(1, 1), mar = c(4, 4, 2, 1))
```

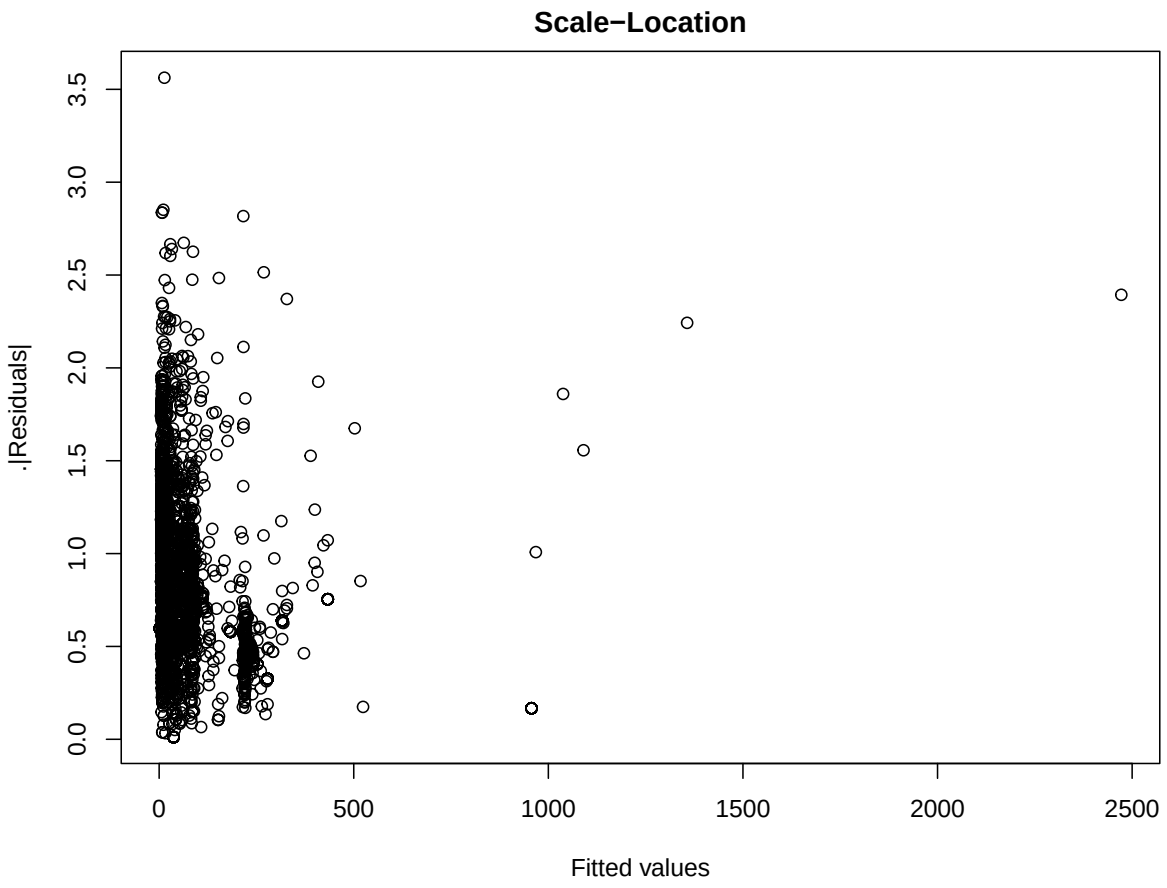
```r
    # Plot 1: Residuals vs Fitted
    plot(fitted(gam_model), residuals(gam_model),
         xlab = "Fitted values", ylab = "Residuals",
         main = "Residuals vs Fitted")
    abline(h = 0, col = "red", lty = 2)

    # Plot 2: Q-Q plot
    qqnorm(residuals(gam_model))
    qqline(residuals(gam_model), col = "red")

    # Plot 3: Scale-Location (if appropriate)
    plot(fitted(gam_model), sqrt(abs(residuals(gam_model))),
         xlab = "Fitted values", ylab = "√|Residuals|",
         main = "Scale-Location")

    # Plot 4: Response vs Fitted
    plot(fitted(gam_model), model_data$total_butterflies,
         xlab = "Fitted values", ylab = "Observed values",
         main = "Response vs Fitted")
    abline(a = 0, b = 1, col = "red", lty = 2)

  }, error = function(e) {
    cat("Error fitting model:", e$message, "\n")
    cat("This might be due to insufficient data or convergence issues\n")
    aic_values_30m[model_name] <- NA
  })
}
```

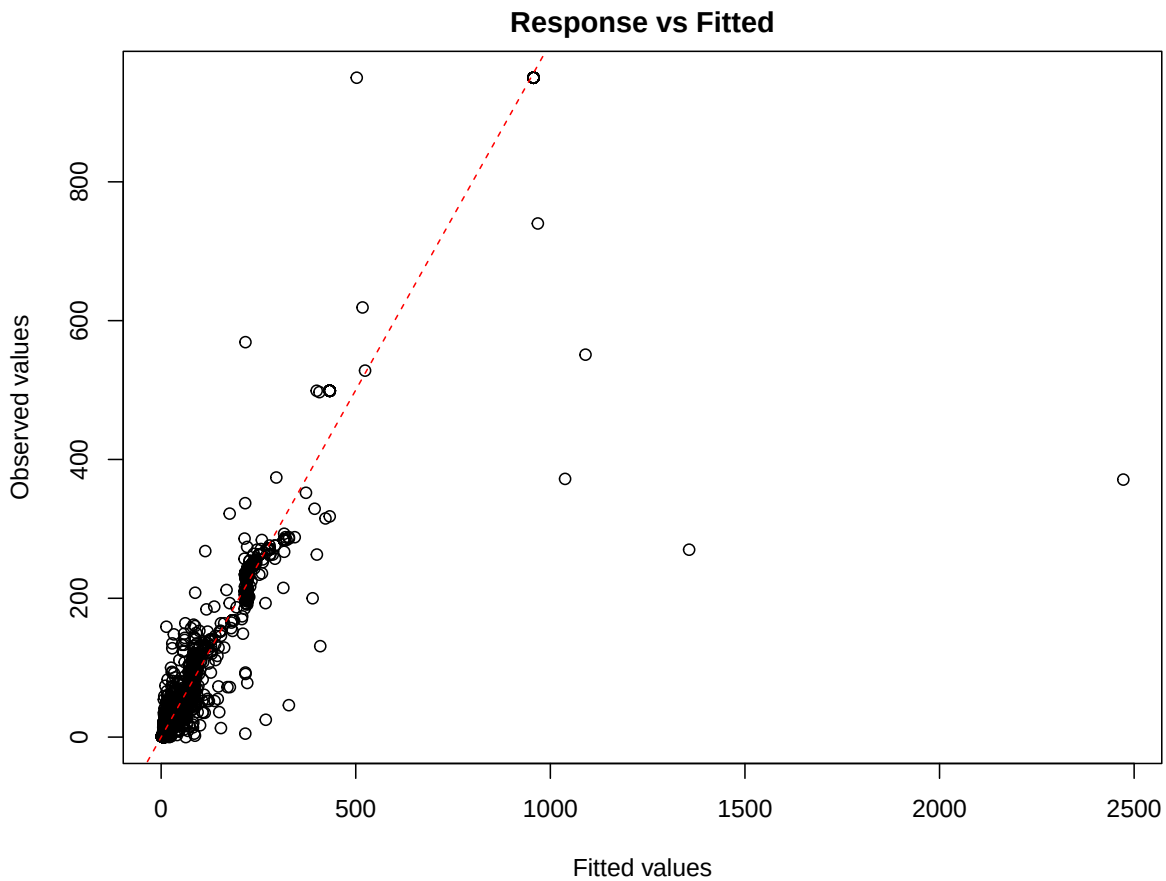
```
### Model 1 : 1. Baseline

Model data dimensions after removing NAs: 5601 x 3
Variables in model: total_butterflies, abundance_t_minus_1, view_id
View IDs with observations: 9
Min/Max observations per view: 13 / 2941
GAM fitted successfully
Model Summary:

Family: Negative Binomial(15.91)
Link function: log
```

```
Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3619     0.1704   19.73   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                         edf Ref.df  Chi.sq p-value
s(abundance_t_minus_1) 8.963  8.999 63809.5  <2e-16 ***
s(view_id)             7.785  8.000   113.3  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.843   Deviance explained = 92.8%
-REML =  21673  Scale est. = 1         n = 5601


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
10
5
0
−5
0
500
1000
1500
2000
2500

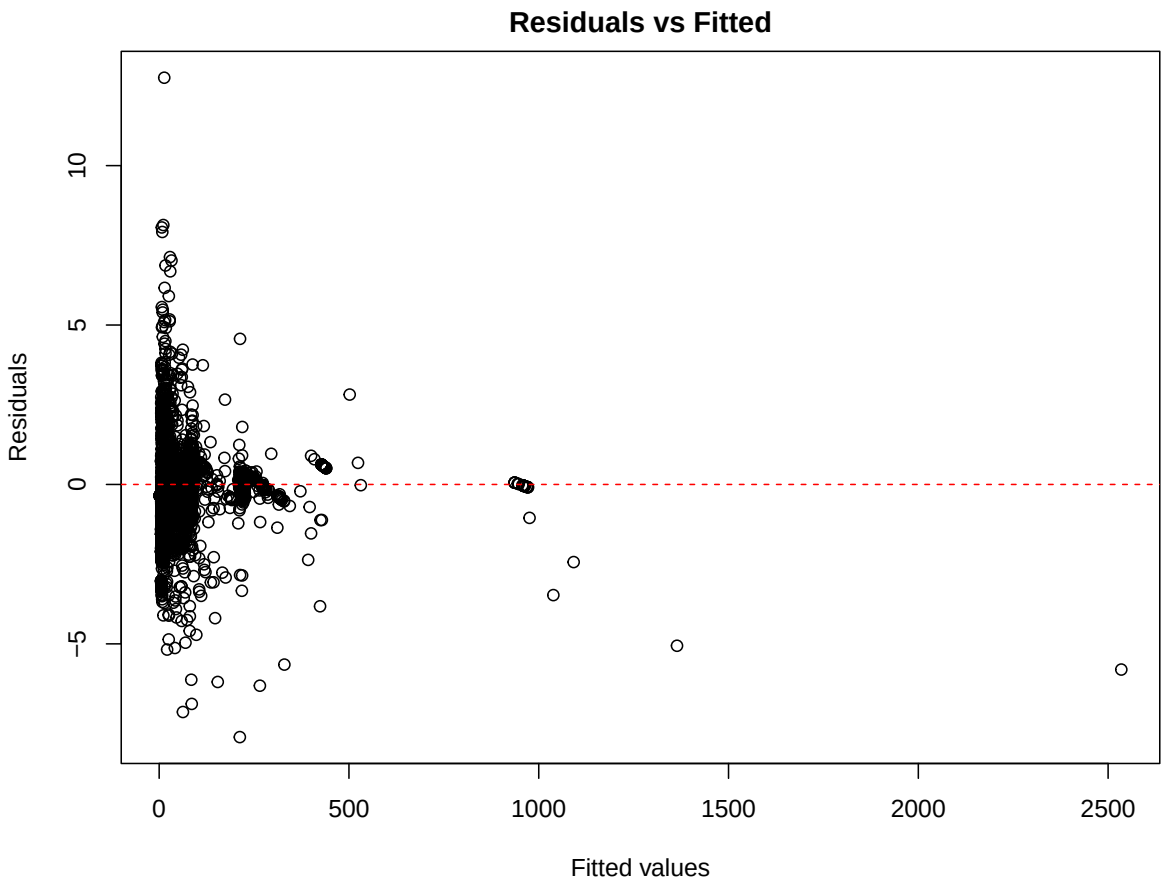

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4

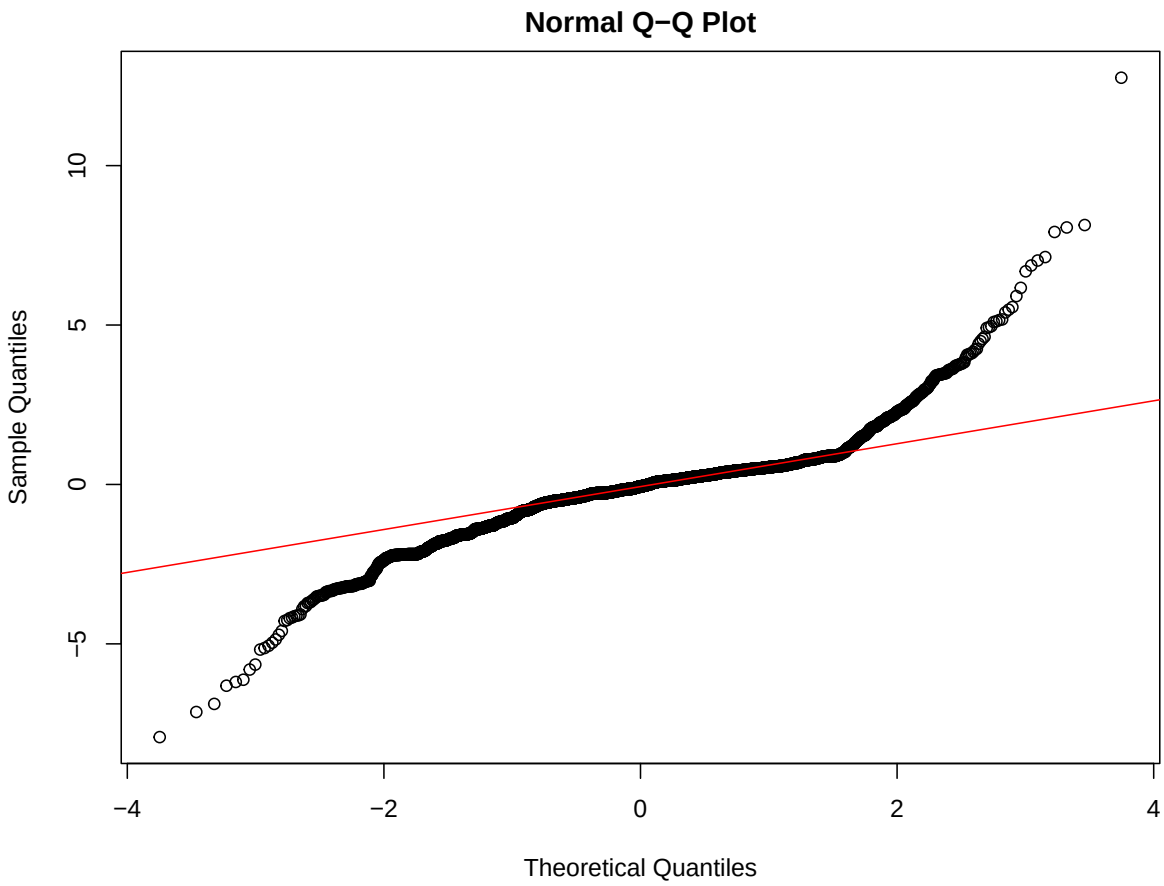

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
500
1000
1500
2000
2500
Fitted values

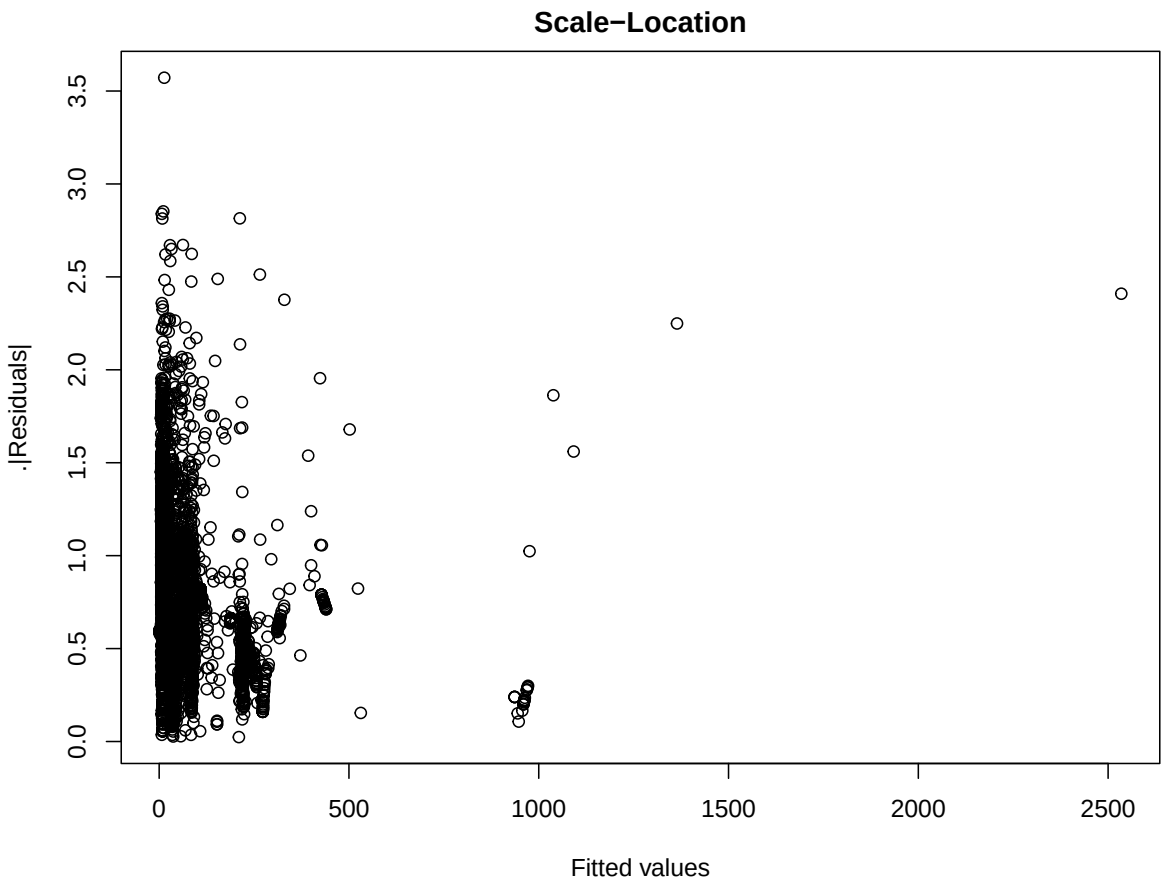


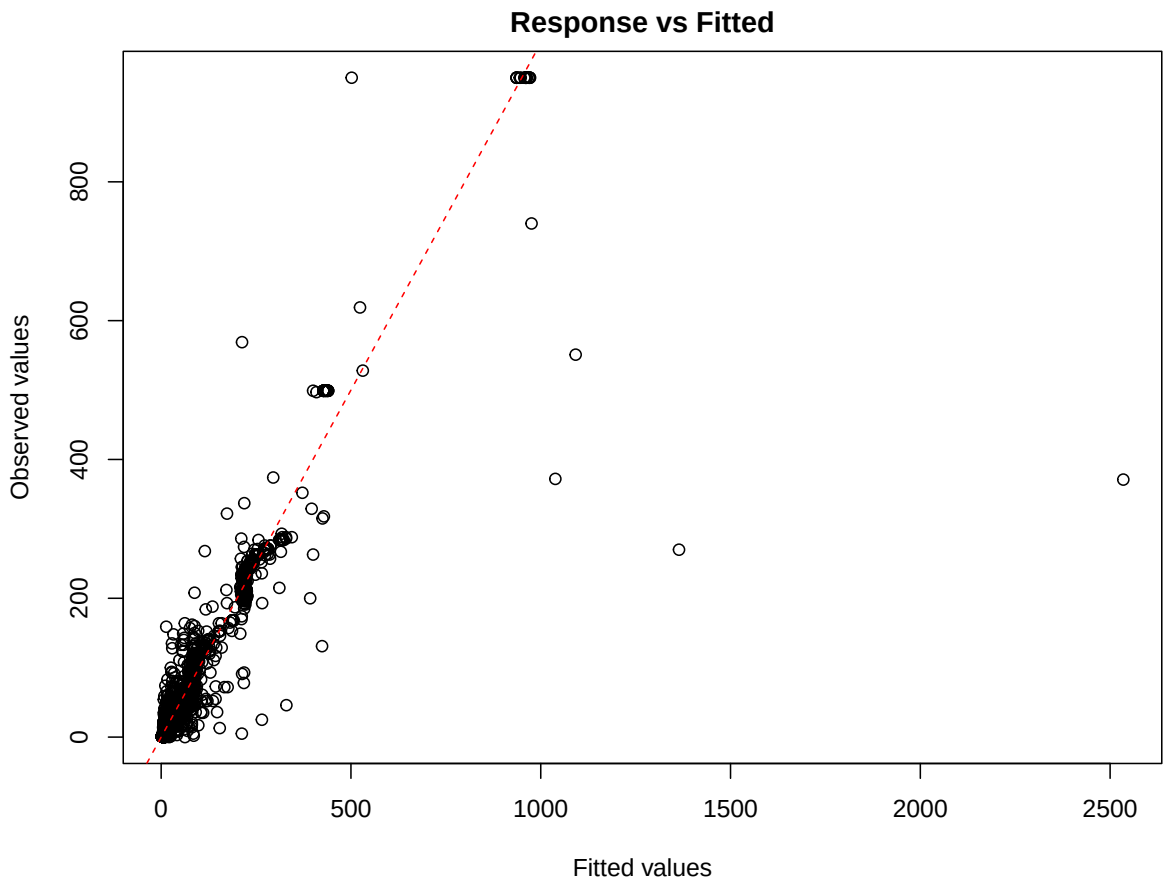

```
### Model 2 : 2a. Wind only

Model data dimensions after removing NAs: 5601 x 4
Variables in model: total_butterflies, abundance_t_minus_1, mean_wind_speed, view_id
View IDs with observations: 9
Min/Max observations per view: 13 / 2941
GAM fitted successfully
Model Summary:

Family: Negative Binomial(16.01)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(view_id, bs = "re")
```

```
Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    3.367      0.172   19.57   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                         edf Ref.df  Chi.sq p-value
s(abundance_t_minus_1) 8.964  8.999 62497.8 < 2e-16 ***
s(mean_wind_speed)     3.391  4.232    15.3 0.00534 **
s(view_id)             7.788  8.000   115.7 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.837   Deviance explained = 92.9%
-REML =  21672  Scale est. = 1         n = 5601


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
10
5
0
−5
0
500
1000
1500
2000
2500

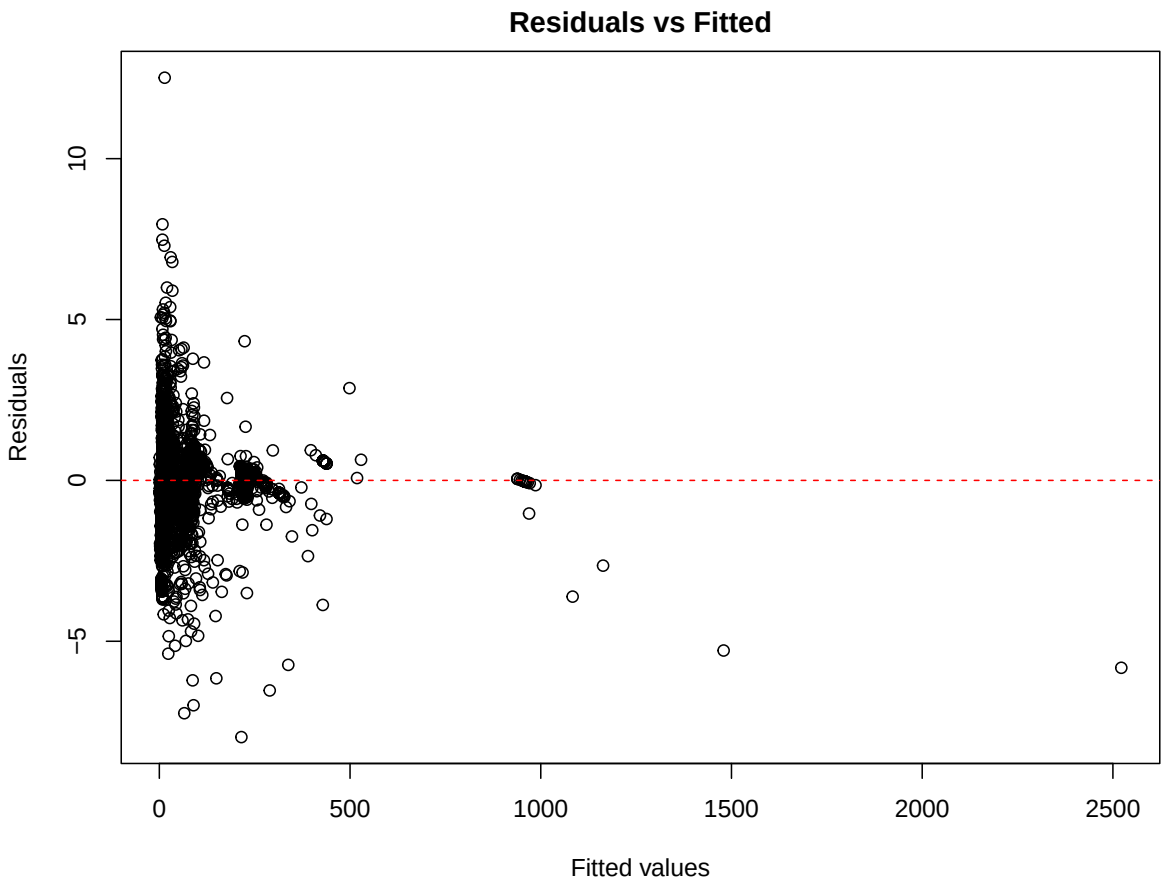

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4

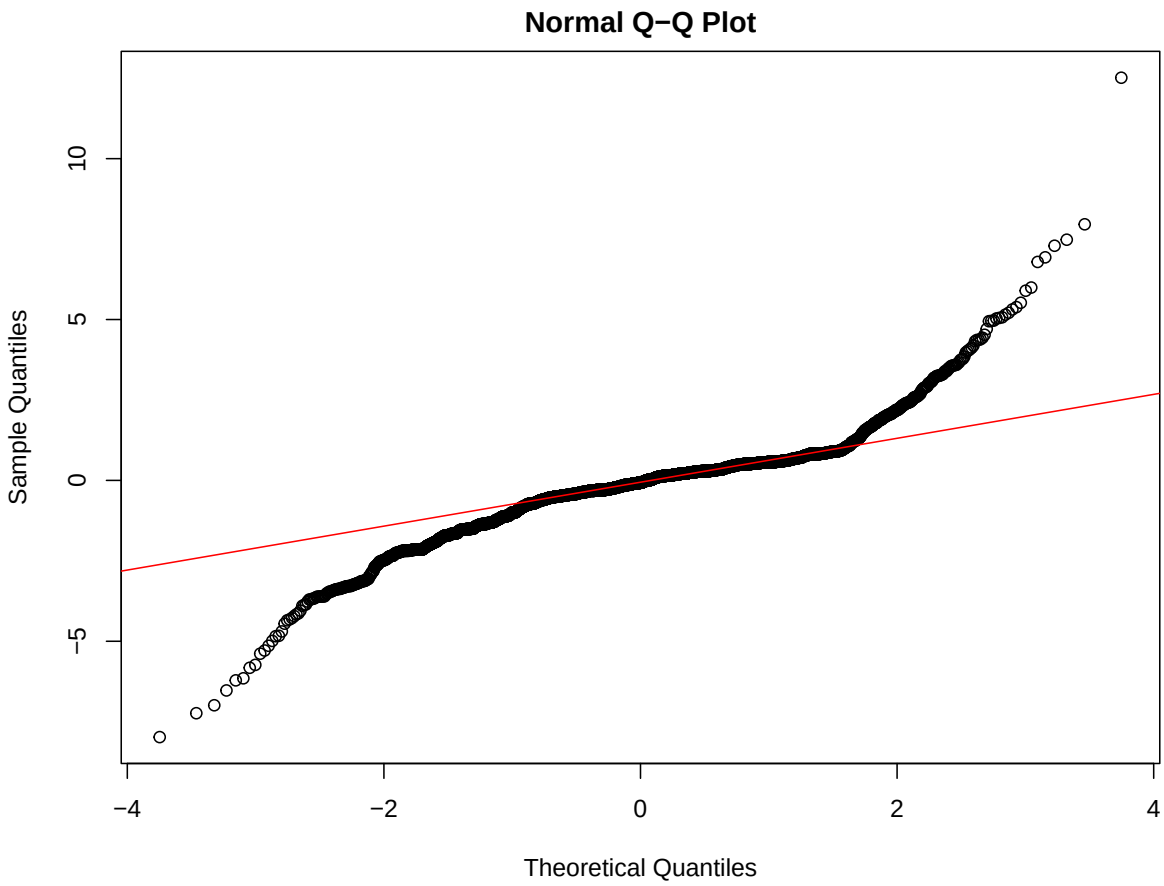

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
500
1000
1500
2000
2500
Fitted values

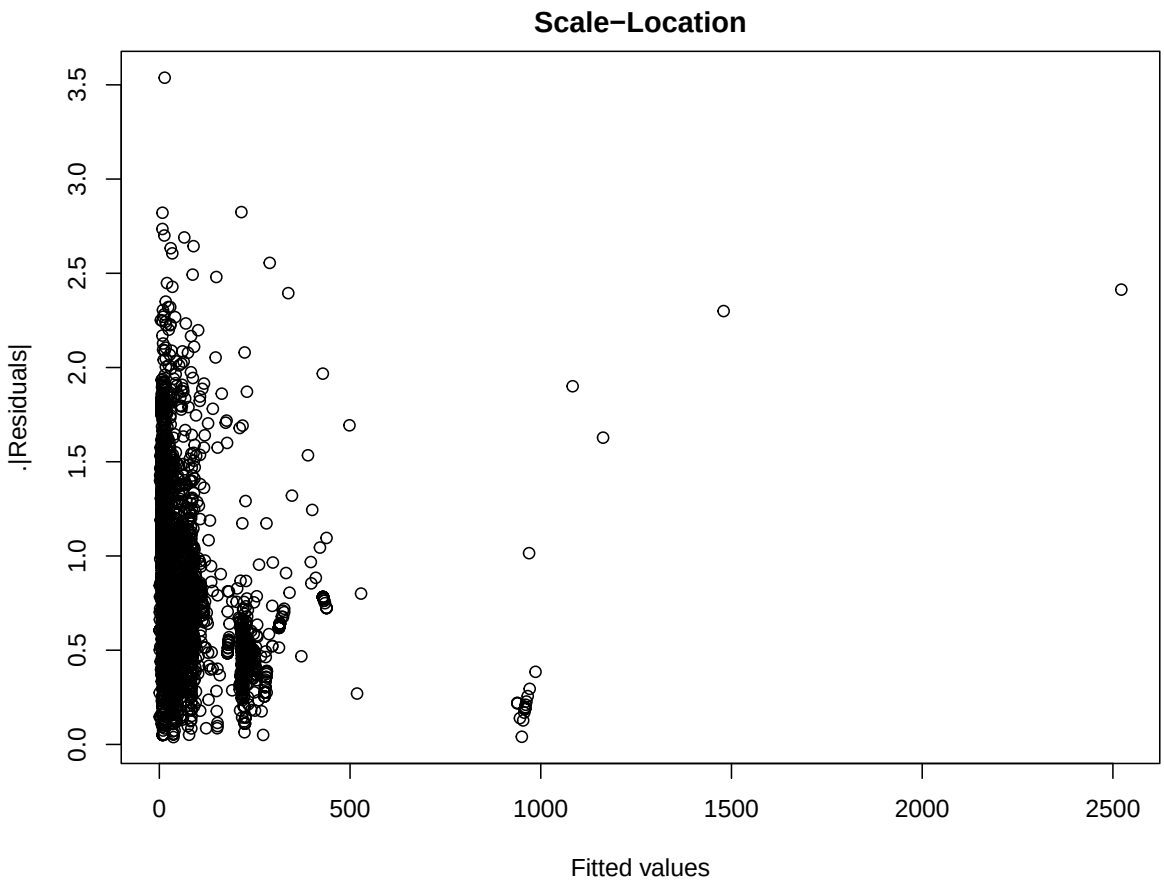


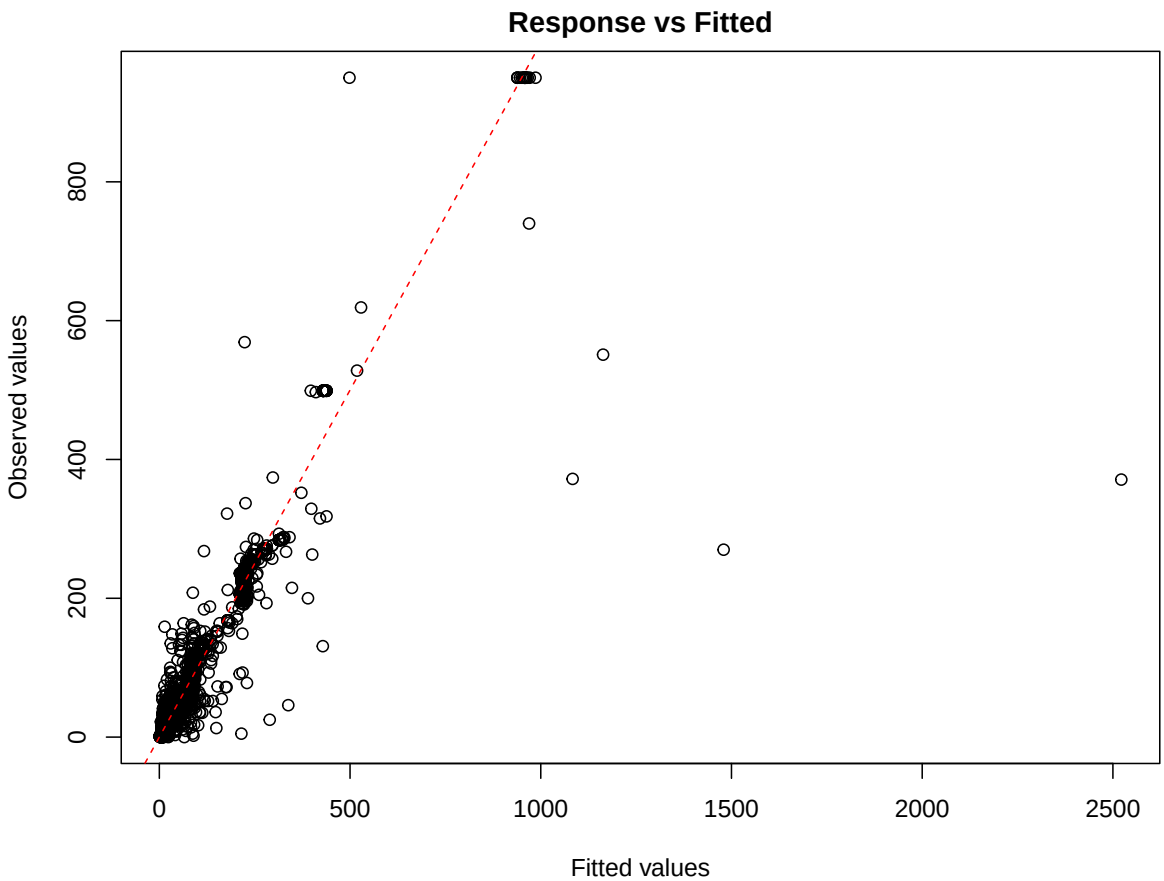

```
### Model 3 : 2b. Temperature only

Model data dimensions after removing NAs: 5601 x 4
Variables in model: total_butterflies, abundance_t_minus_1, mean_temp, view_id
View IDs with observations: 9
Min/Max observations per view: 13 / 2941
GAM fitted successfully
Model Summary:

Family: Negative Binomial(16.183)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(view_id,
    bs = "re")
```

```
Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3536     0.1659   20.22   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                         edf Ref.df  Chi.sq p-value
s(abundance_t_minus_1) 8.964  8.999 60495.4  <2e-16 ***
s(mean_temp)           7.720  8.389   102.3  <2e-16 ***
s(view_id)             7.775  8.000   126.5  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.829   Deviance explained =   93%
-REML =  21631  Scale est. = 1         n = 5601


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
10
5
0
−5
0
500
1000
1500
2000
2500

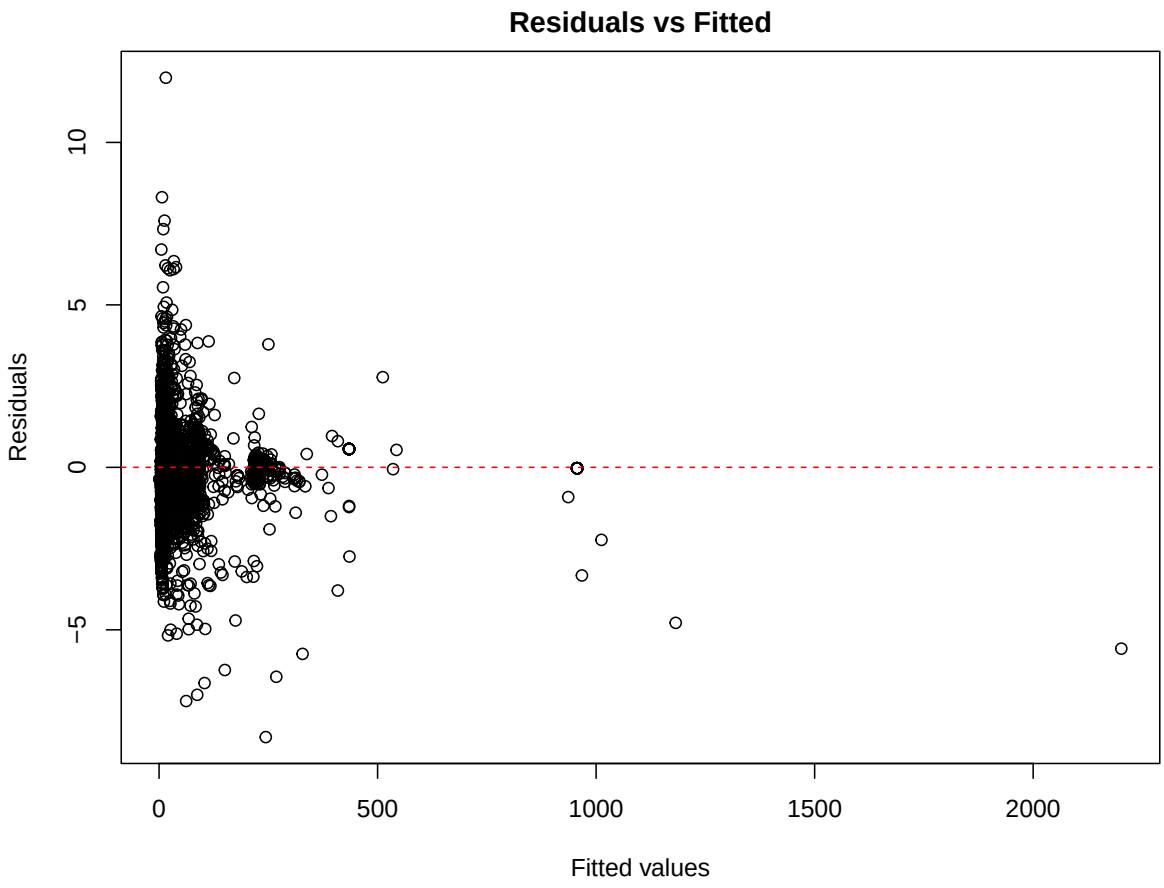

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4

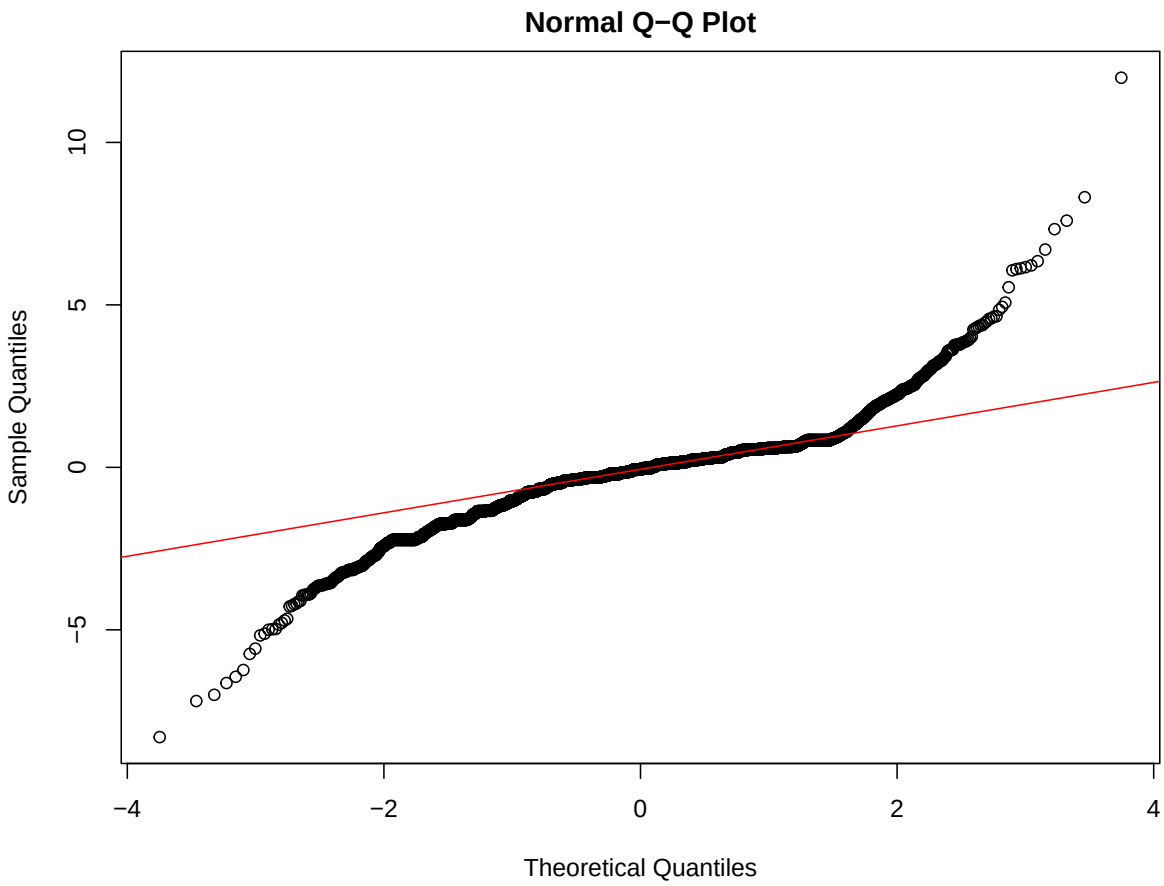

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
500
1000
1500
2000
2500
Fitted values

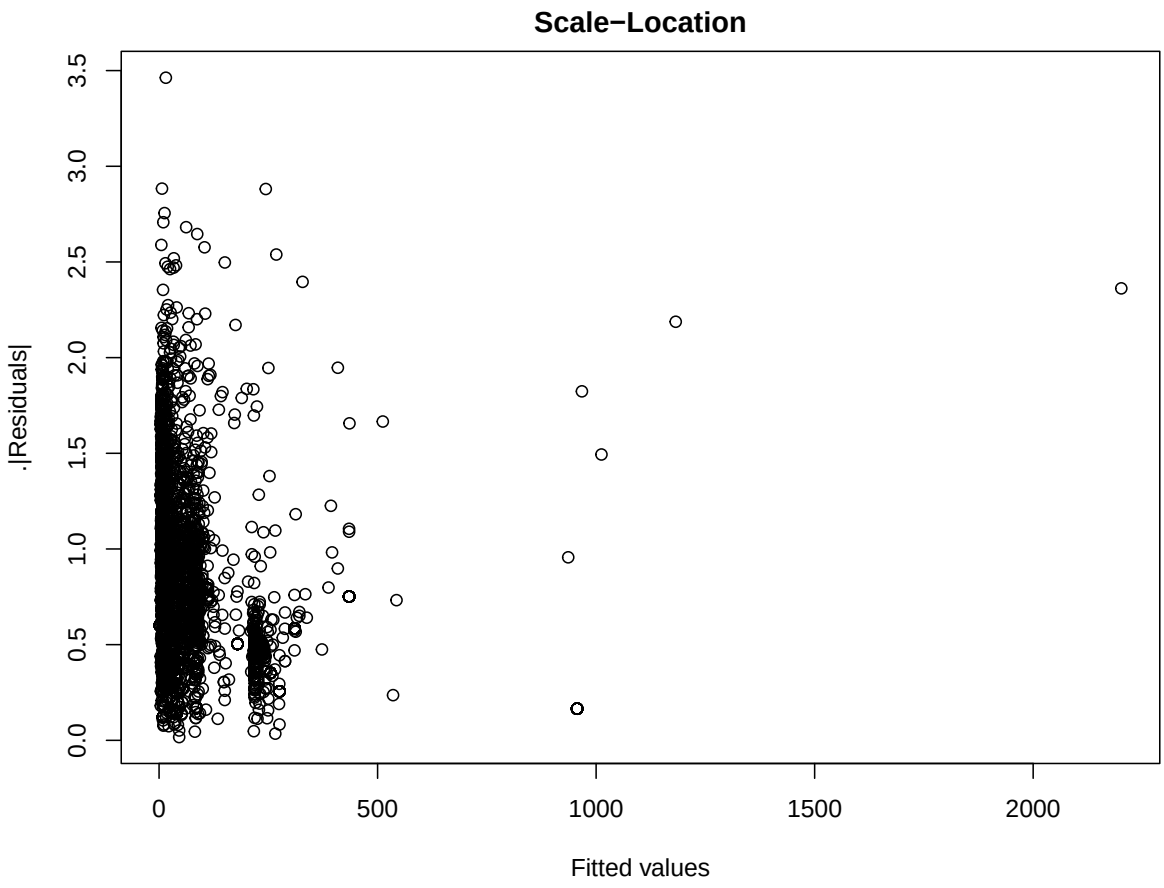


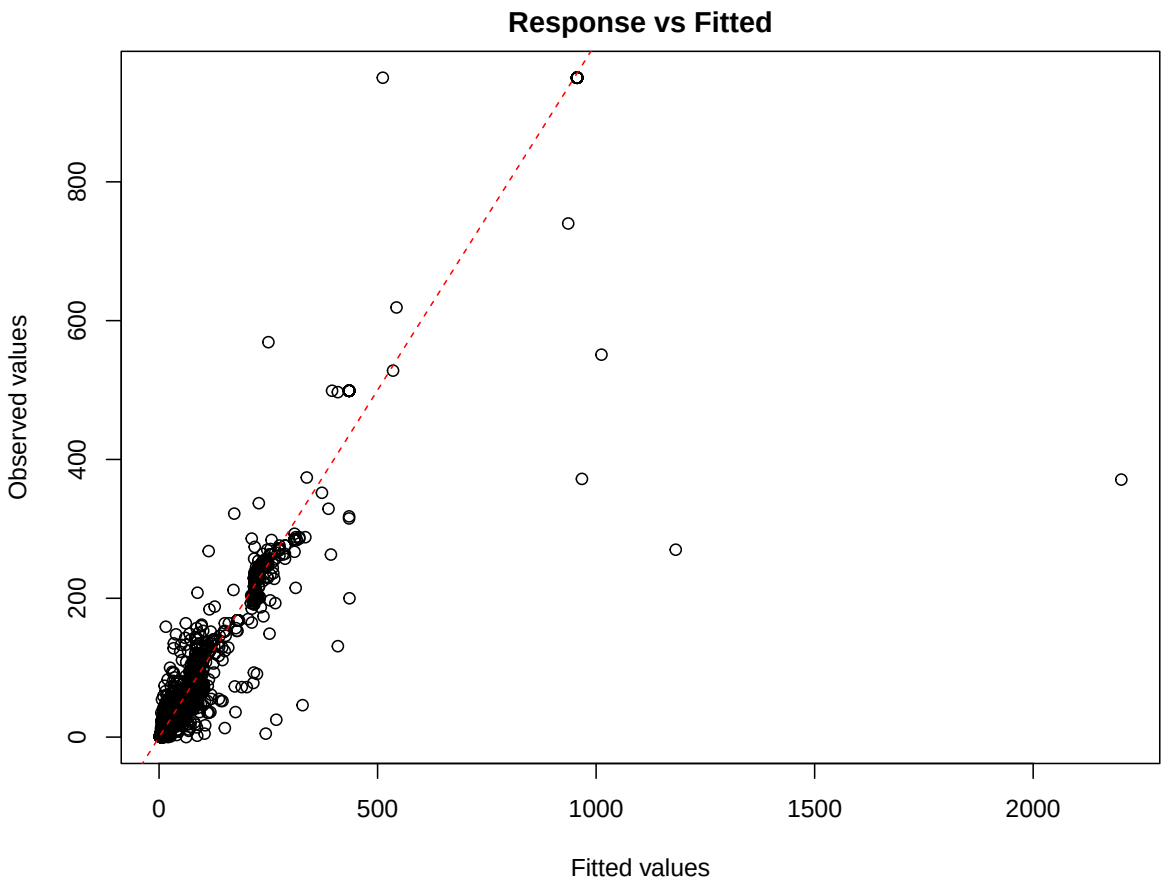

```
### Model 4 : 2c. Sunlight only

Model data dimensions after removing NAs: 5601 x 4
Variables in model: total_butterflies, abundance_t_minus_1, sunlight_proportion, view_id
View IDs with observations: 9
Min/Max observations per view: 13 / 2941
GAM fitted successfully
Model Summary:

Family: Negative Binomial(16.659)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(sunlight_proportion) +
    s(view_id, bs = "re")
```

```
Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    3.360      0.169   19.88   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                         edf Ref.df   Chi.sq p-value
s(abundance_t_minus_1) 8.960  8.999 60950.20  <2e-16 ***
s(sunlight_proportion) 8.280  8.850   177.09  <2e-16 ***
s(view_id)             7.784  8.000    97.03  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.873   Deviance explained = 93.1%
-REML =  21603  Scale est. = 1         n = 5601


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
10
5
0
−5
0
500
1000
1500
2000

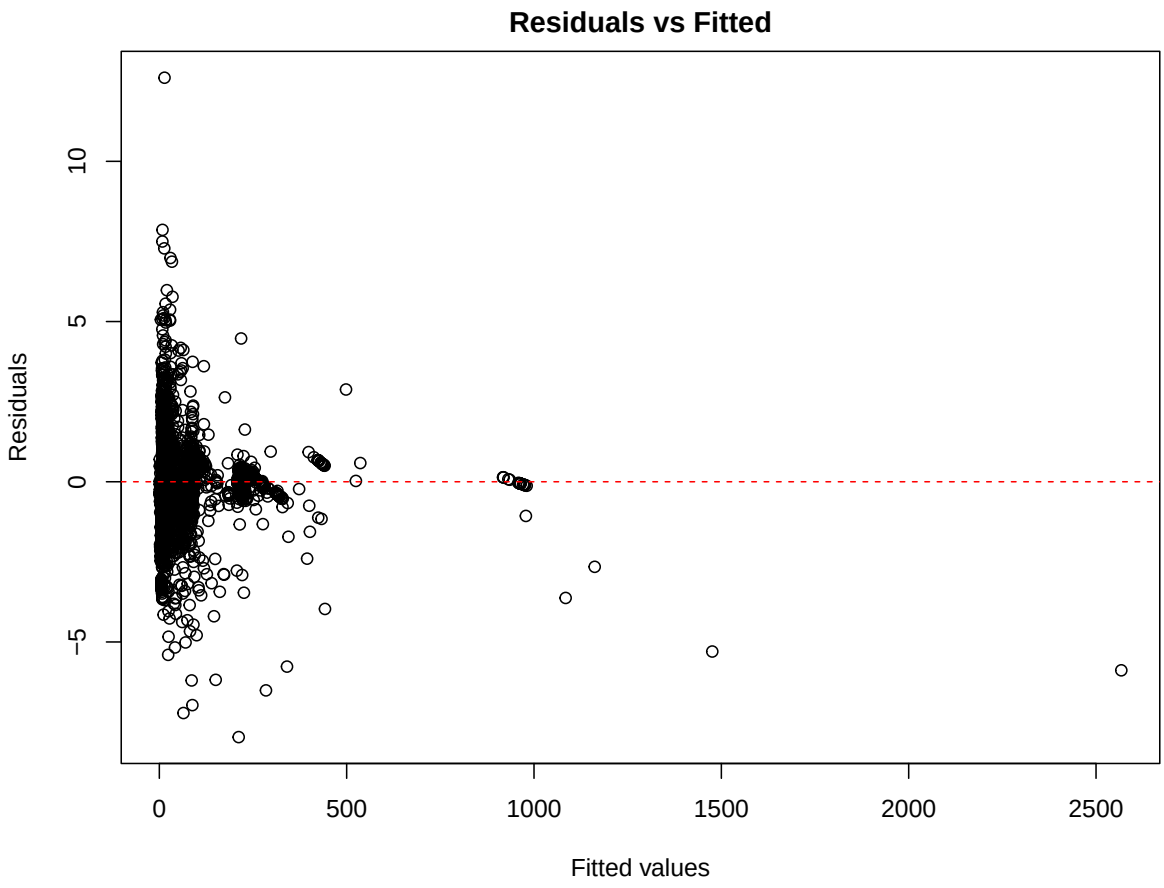

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4

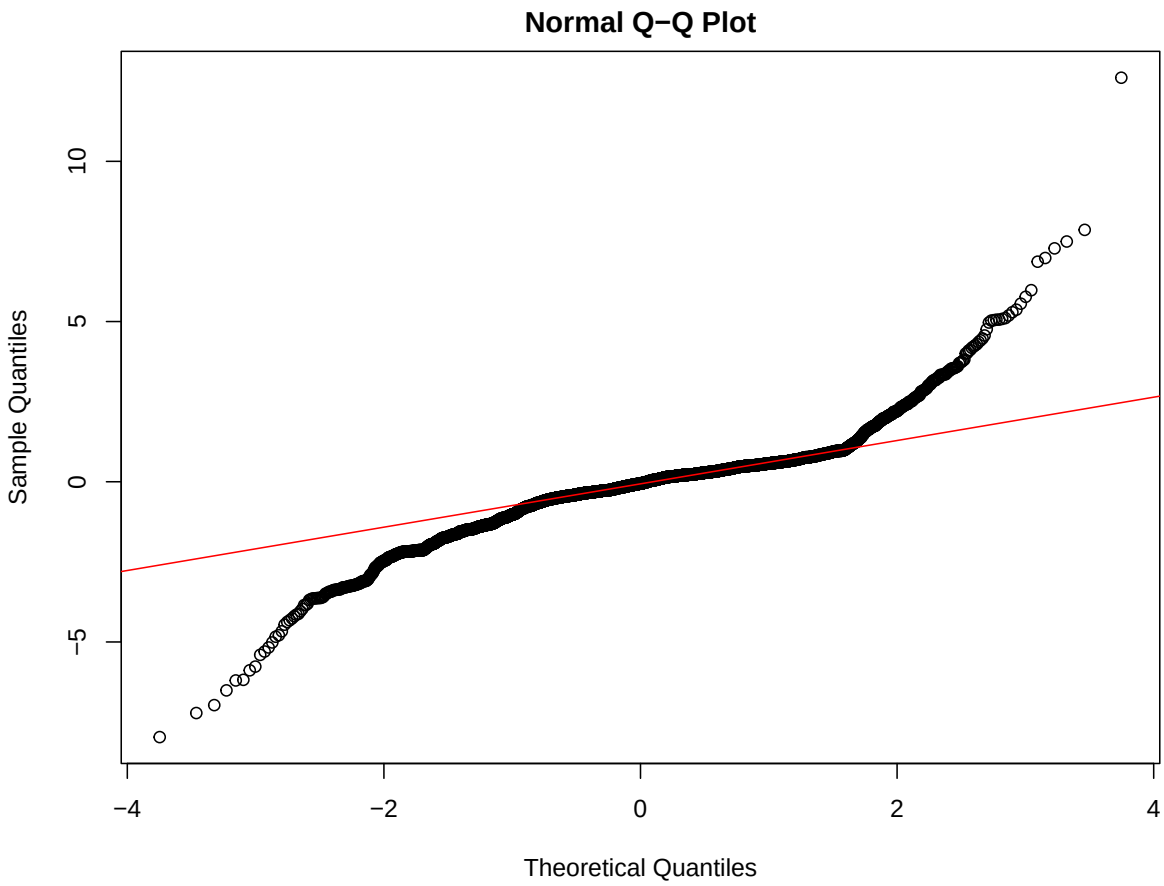

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
500
1000
1500
2000
Fitted values

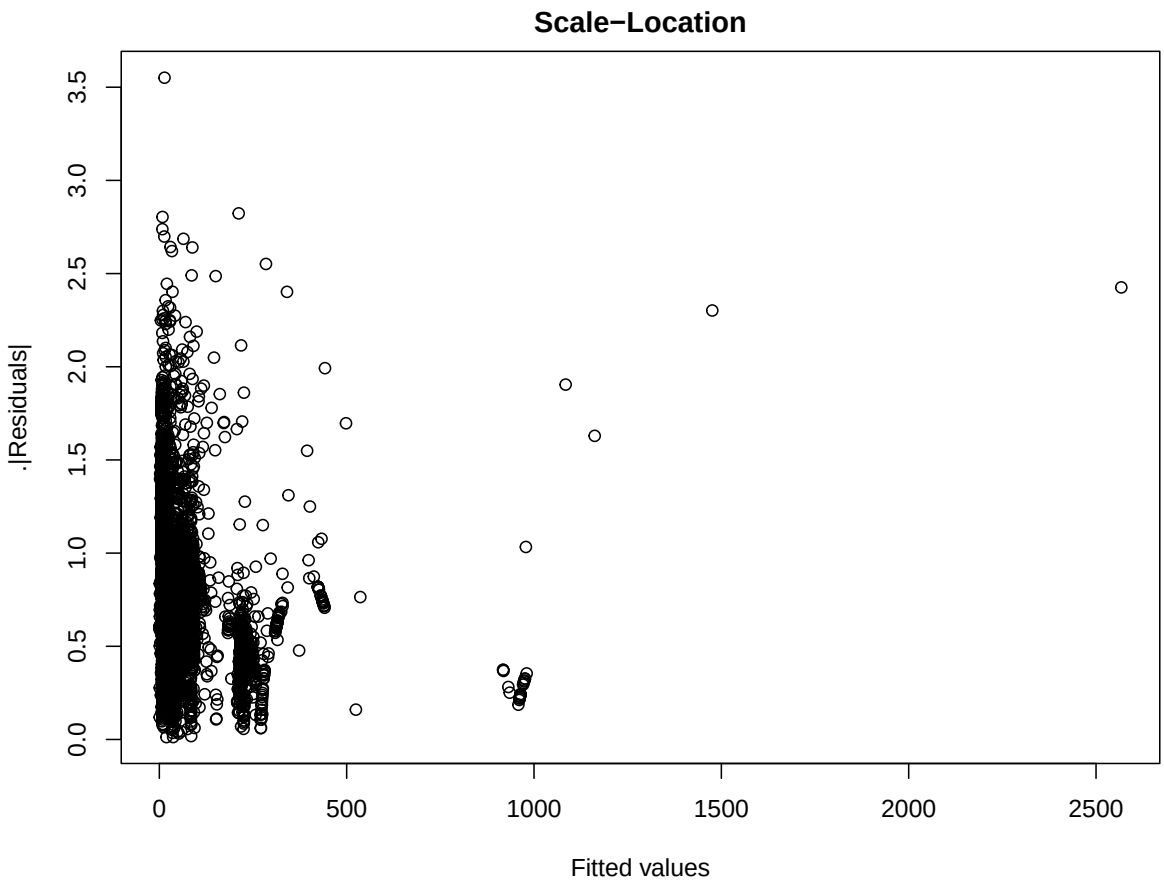


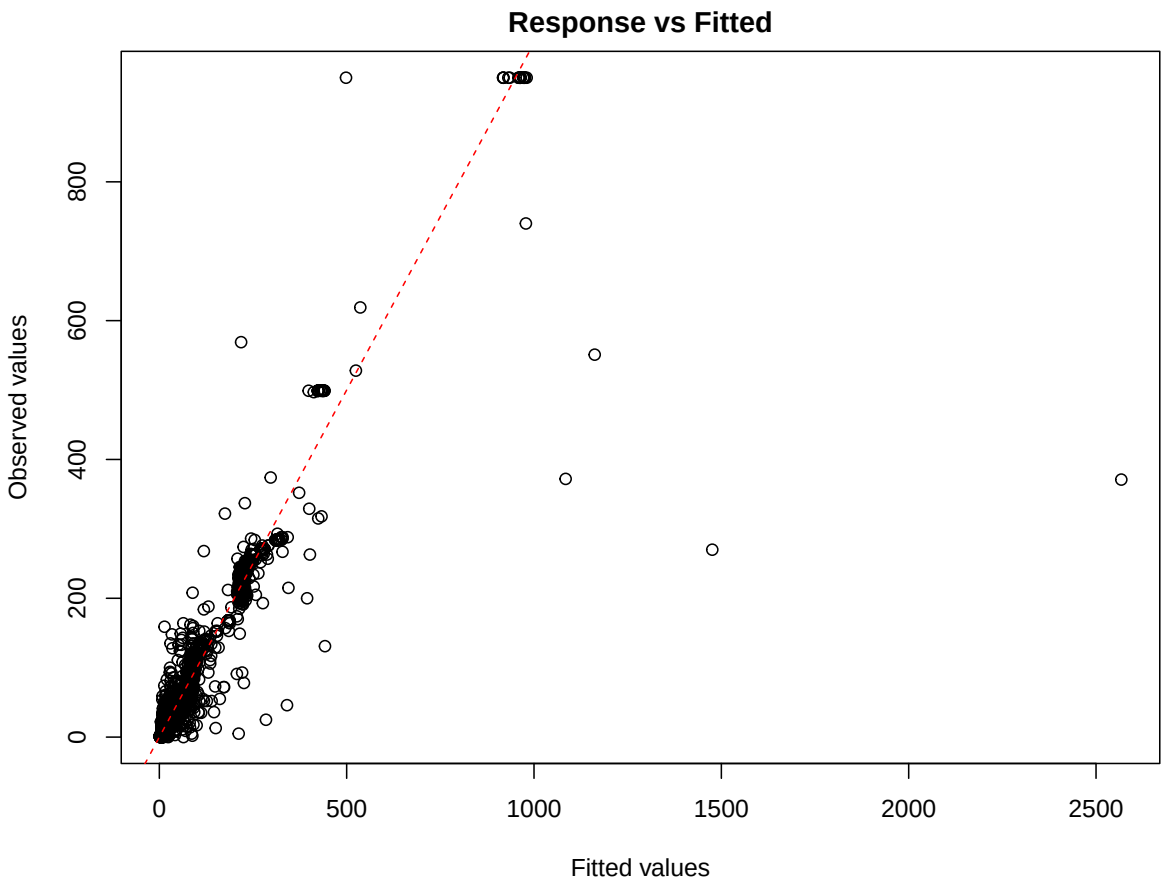

```
### Model 5 : 3a. Wind + Temp

Model data dimensions after removing NAs: 5601 x 5
Variables in model: total_butterflies, abundance_t_minus_1, mean_wind_speed, mean_temp, view_
View IDs with observations: 9
Min/Max observations per view: 13 / 2941
GAM fitted successfully
Model Summary:

Family: Negative Binomial(16.286)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(view_id, bs = "re")
```

```
Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3596     0.1677   20.03   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                         edf Ref.df  Chi.sq p-value
s(abundance_t_minus_1) 8.965  8.999 59396.2 < 2e-16 ***
s(mean_wind_speed)     3.279  4.099    15.2 0.00471 **
s(mean_temp)           7.735  8.400   102.2 < 2e-16 ***
s(view_id)             7.779  8.000   125.4 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.825   Deviance explained =   93%
-REML =  21630  Scale est. = 1         n = 5601


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
10
5
0
−5
0
500
1000
1500
2000
2500

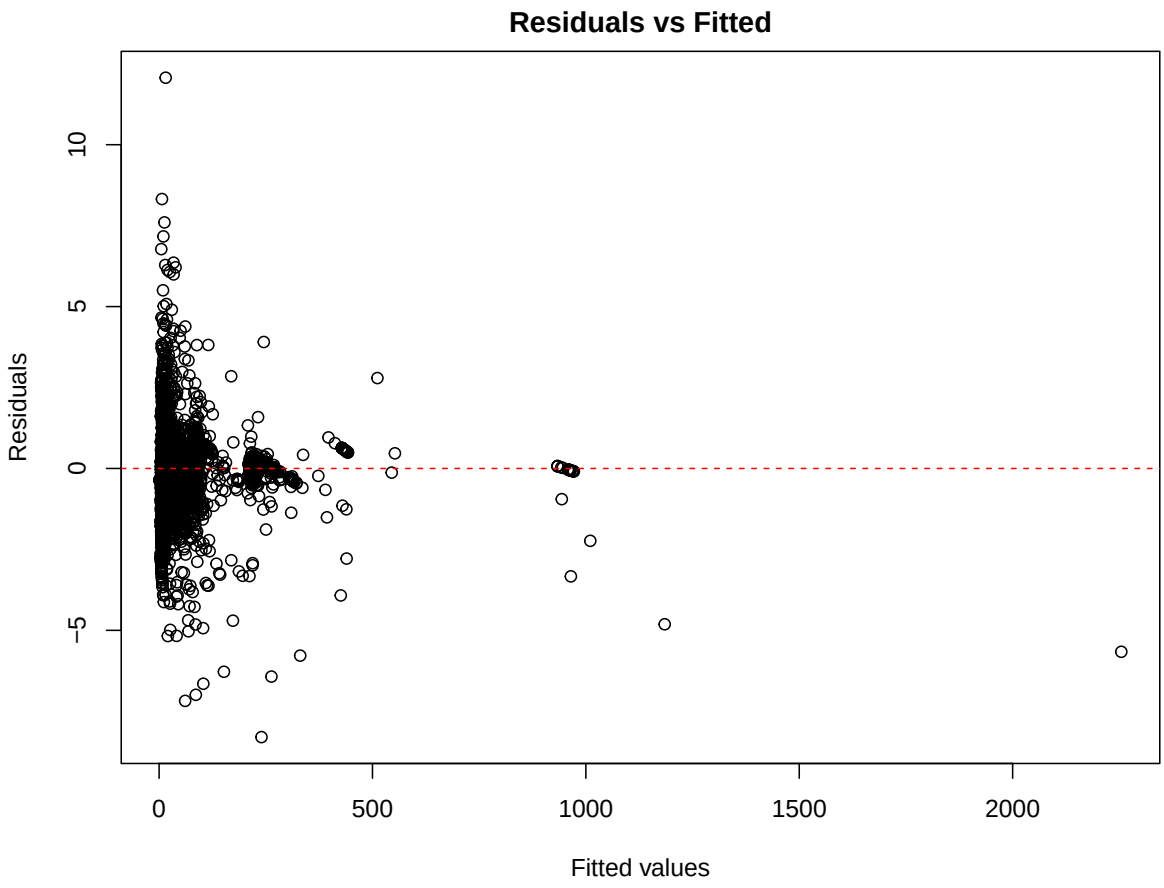

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4

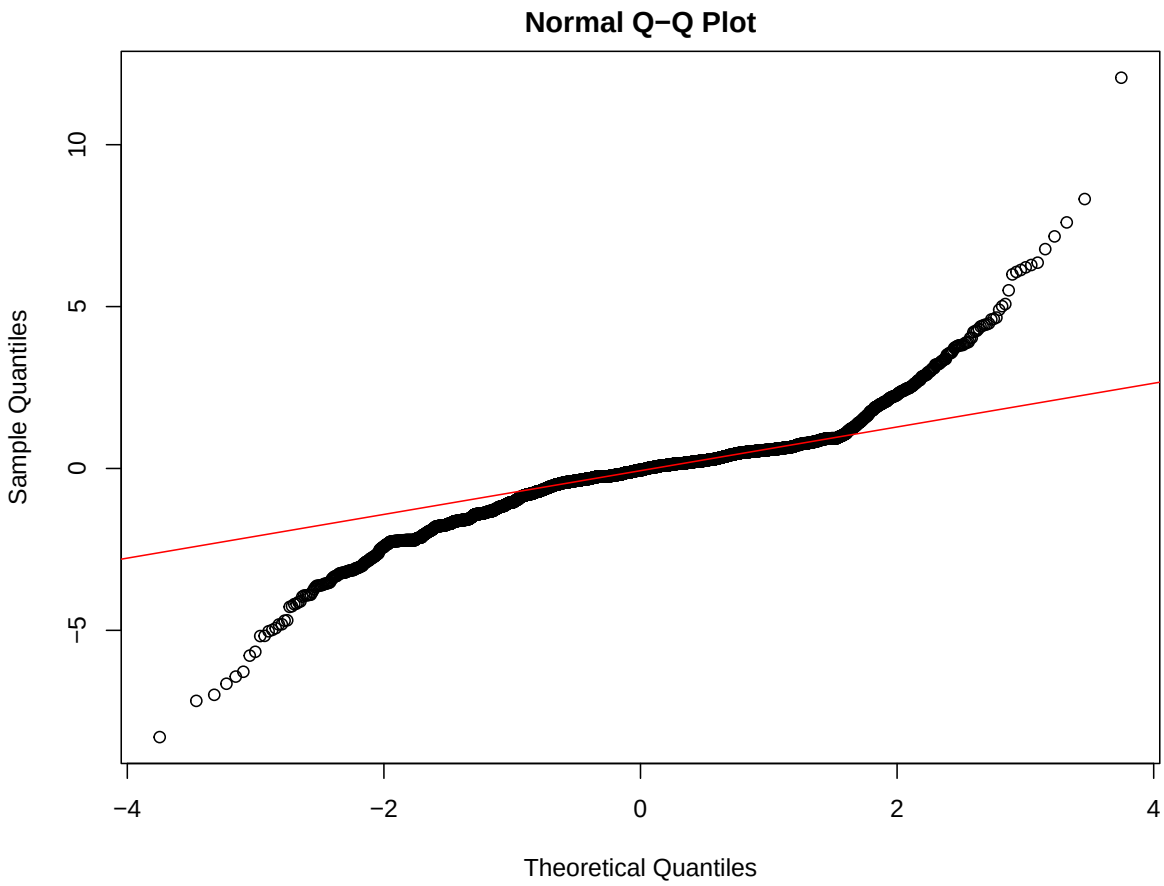

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
500
1000
1500
2000
2500
Fitted values

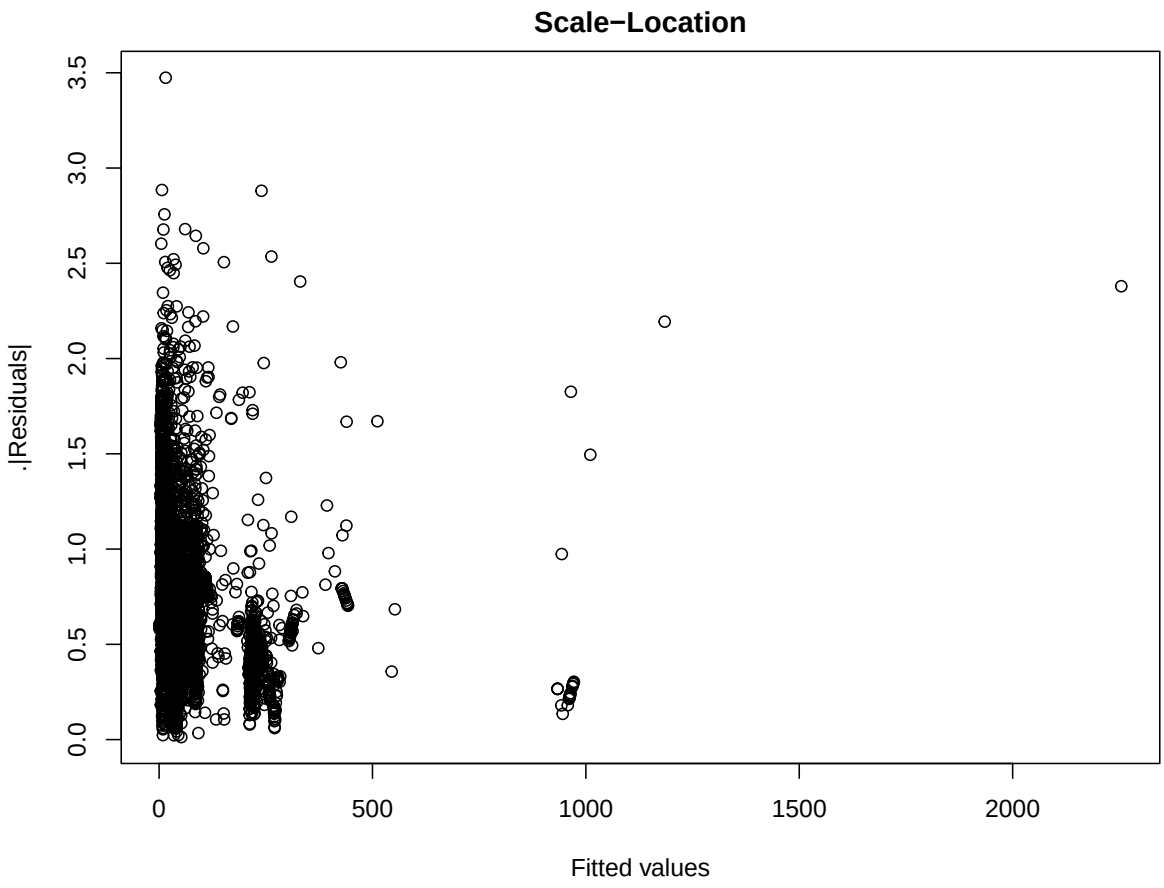


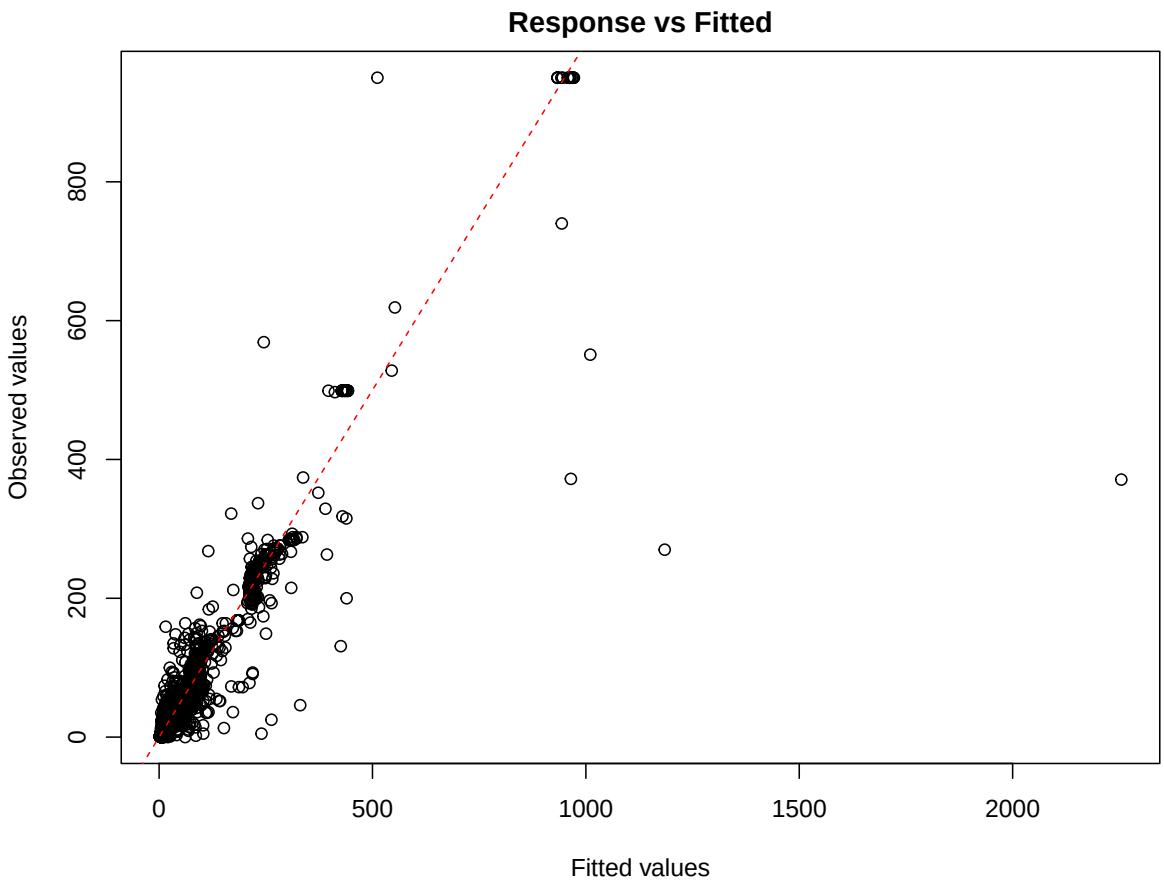

```
### Model 6 : 3b. Wind + Sunlight

Model data dimensions after removing NAs: 5601 x 5
Variables in model: total_butterflies, abundance_t_minus_1, mean_wind_speed, sunlight_propor
View IDs with observations: 9
Min/Max observations per view: 13 / 2941
GAM fitted successfully
Model Summary:

Family: Negative Binomial(16.814)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(sunlight_proportion) + s(view_id, bs = "re")
```

```
Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    3.365      0.171   19.68   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                         edf Ref.df   Chi.sq  p-value
s(abundance_t_minus_1) 8.961  8.999 59479.98  < 2e-16 ***
s(mean_wind_speed)     4.818  5.886    23.42 0.000746 ***
s(sunlight_proportion) 8.292  8.854   182.72  < 2e-16 ***
s(view_id)             7.788  8.000   104.28  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.869   Deviance explained = 93.1%
-REML =  21599  Scale est. = 1         n = 5601


Diagnostic plots:
```


Residuals vs Fitted
Residuals
10
5
0
−5
0
500
1000
1500
2000
Fitted values

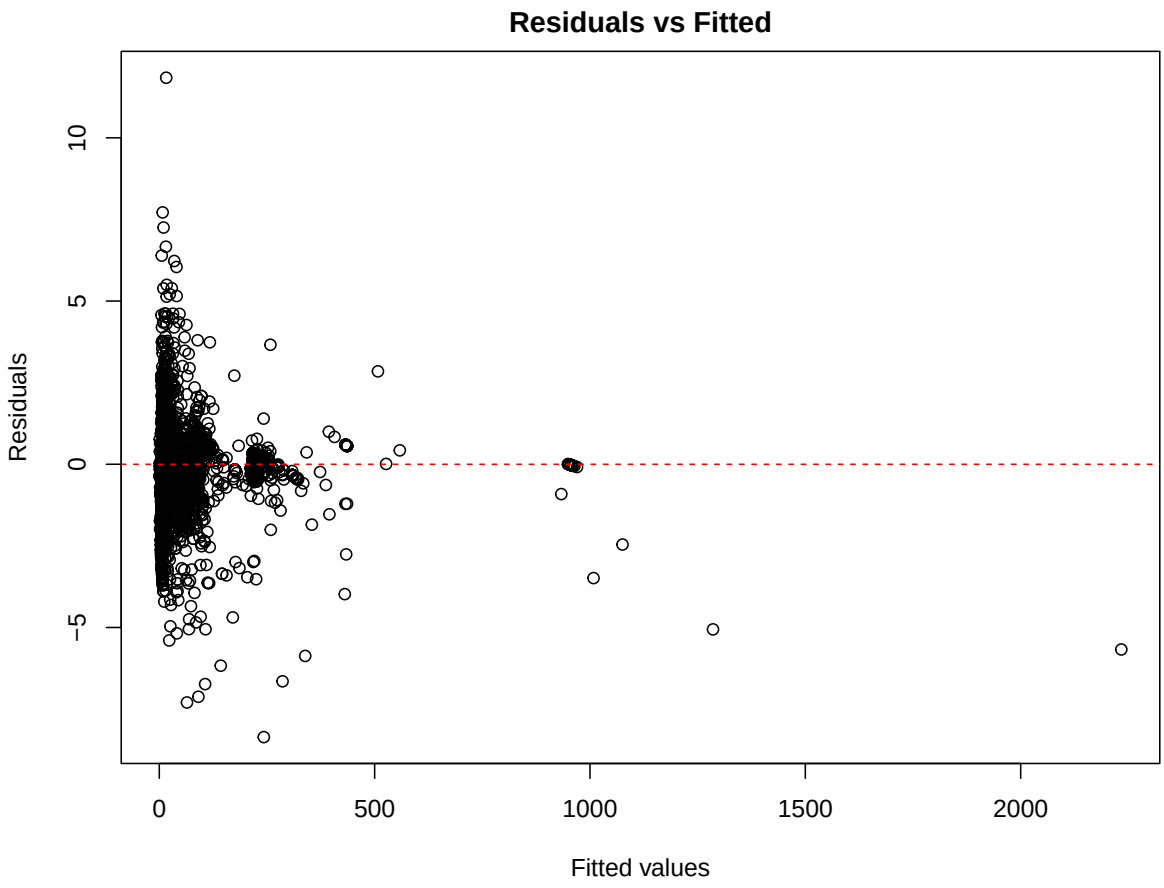

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4

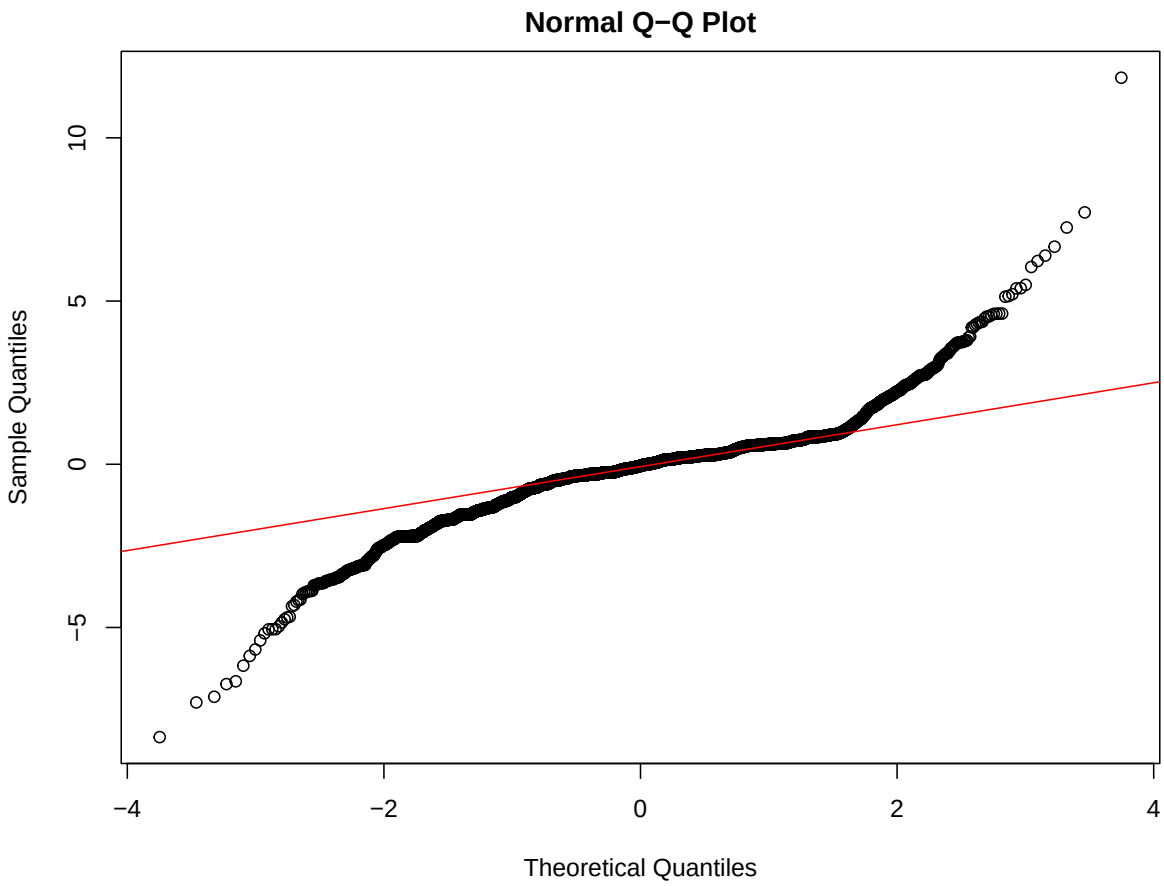

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
500
1000
1500
2000
Fitted values

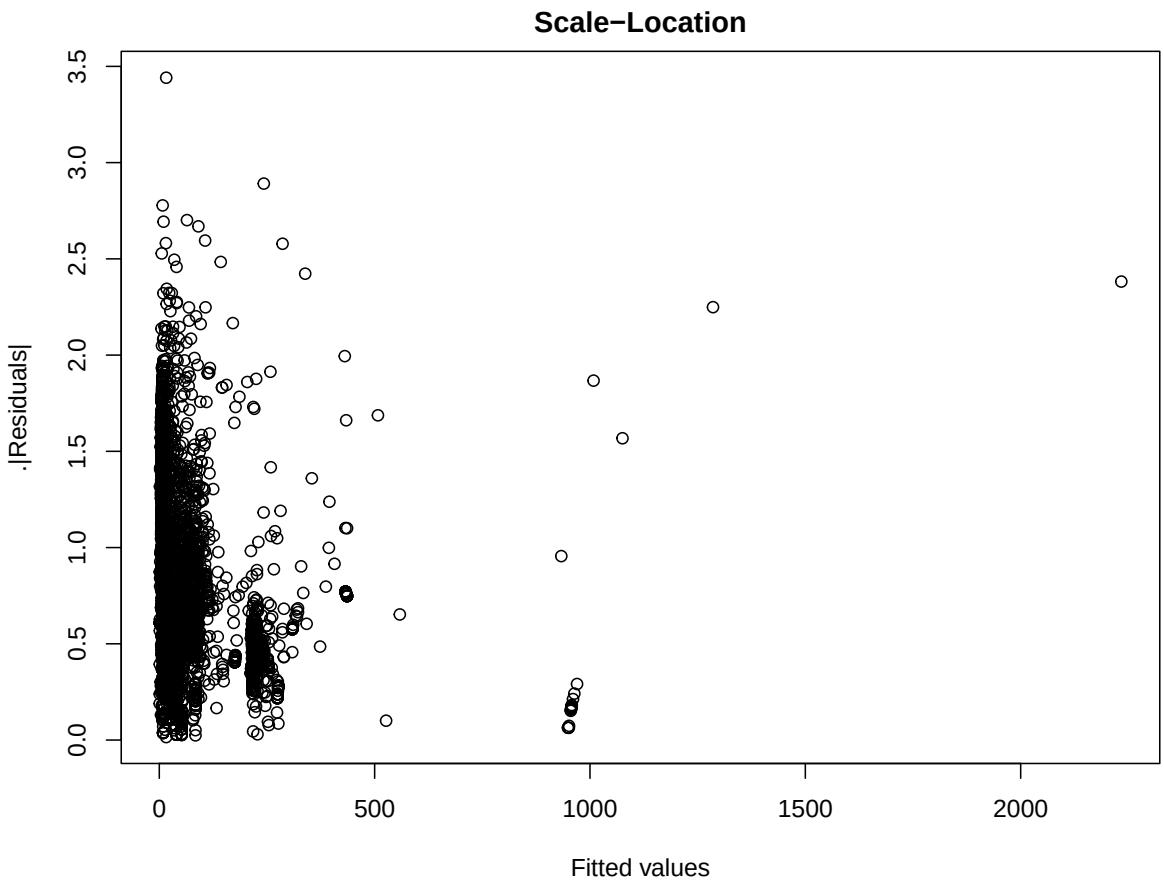


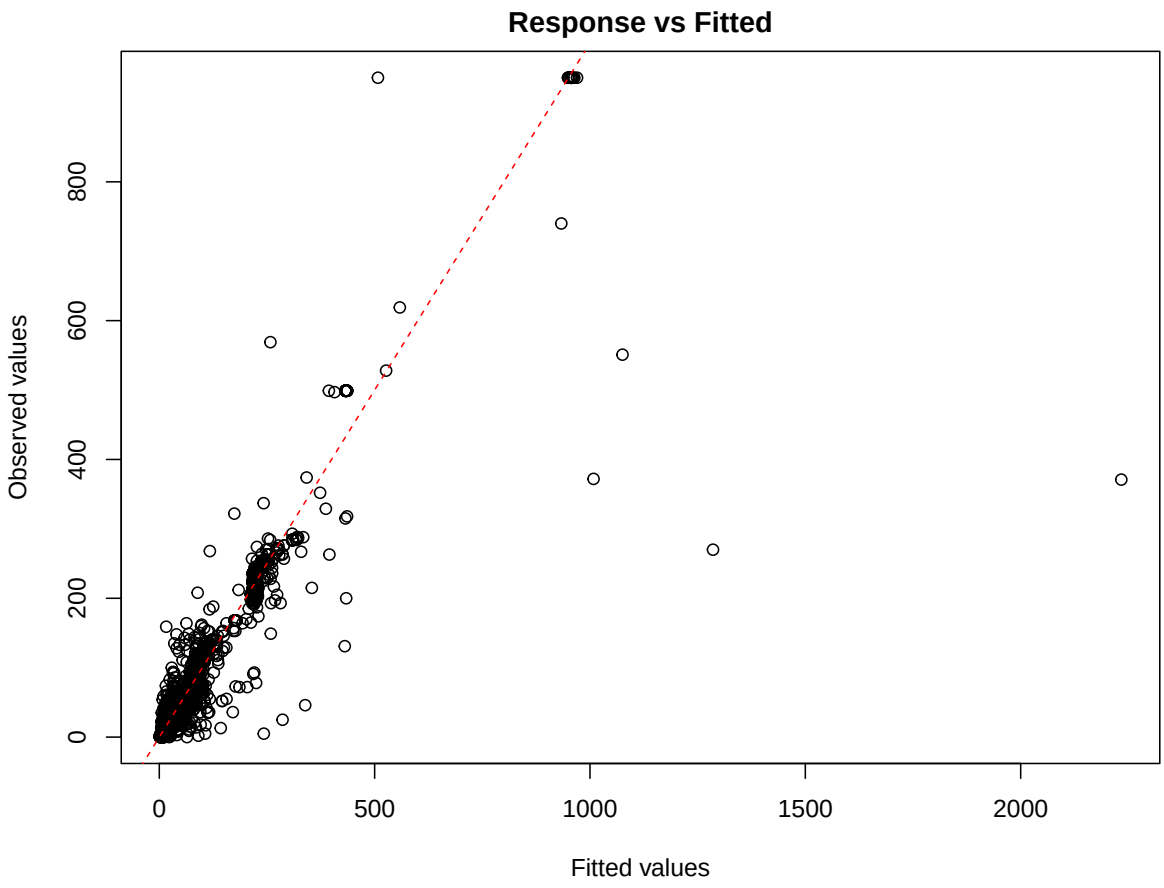

```
### Model 7 : 3c. Temp + Sunlight

Model data dimensions after removing NAs: 5601 x 5
Variables in model: total_butterflies, abundance_t_minus_1, mean_temp, sunlight_proportion, v
View IDs with observations: 9
Min/Max observations per view: 13 / 2941
GAM fitted successfully
Model Summary:

Family: Negative Binomial(17.021)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(sunlight_proportion) +
    s(view_id, bs = "re")
```

```
Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3509     0.1644   20.38   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                         edf Ref.df  Chi.sq p-value
s(abundance_t_minus_1) 8.961  8.999 59612.2  <2e-16 ***
s(mean_temp)           7.816  8.455   113.9  <2e-16 ***
s(sunlight_proportion) 8.252  8.839   187.1  <2e-16 ***
s(view_id)             7.774  8.000   115.6  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.864   Deviance explained = 93.3%
-REML =  21555  Scale est. = 1         n = 5601


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
10
5
0
−5
0
500
1000
1500
2000

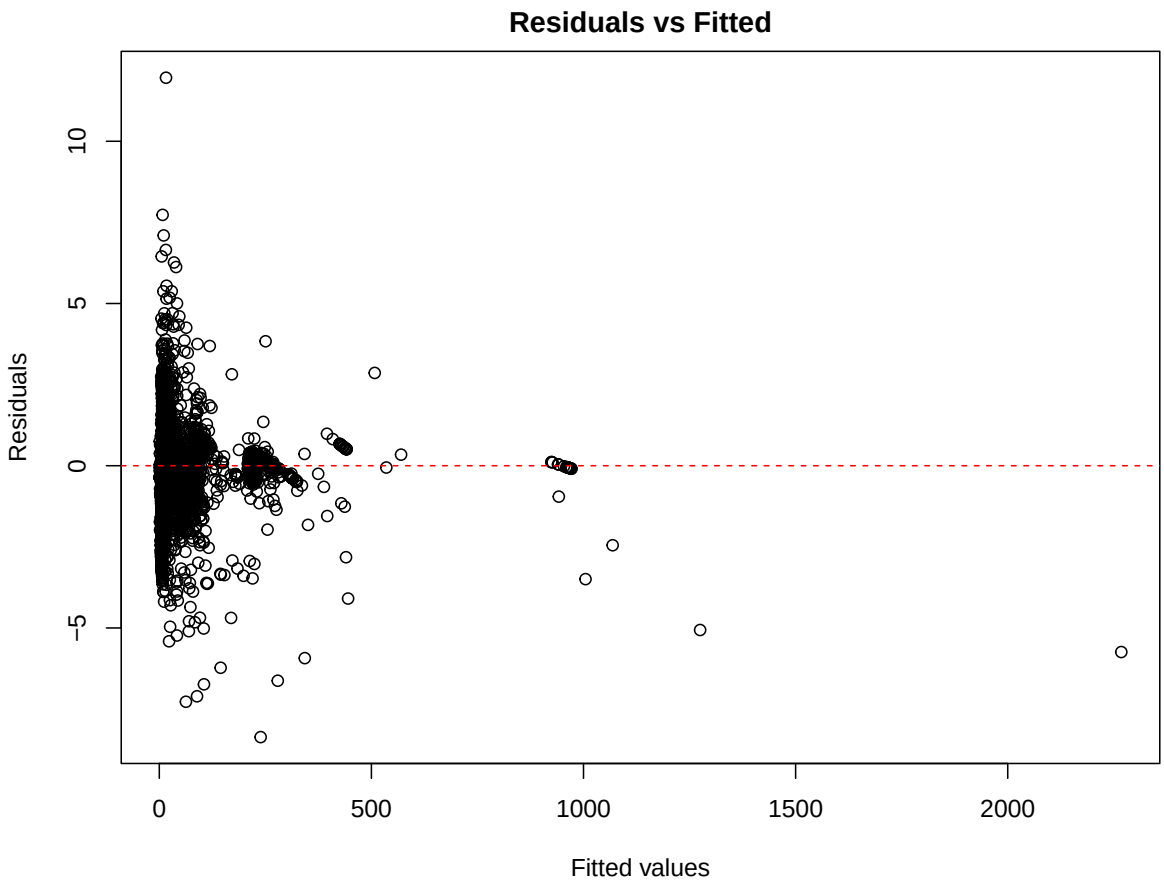

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4

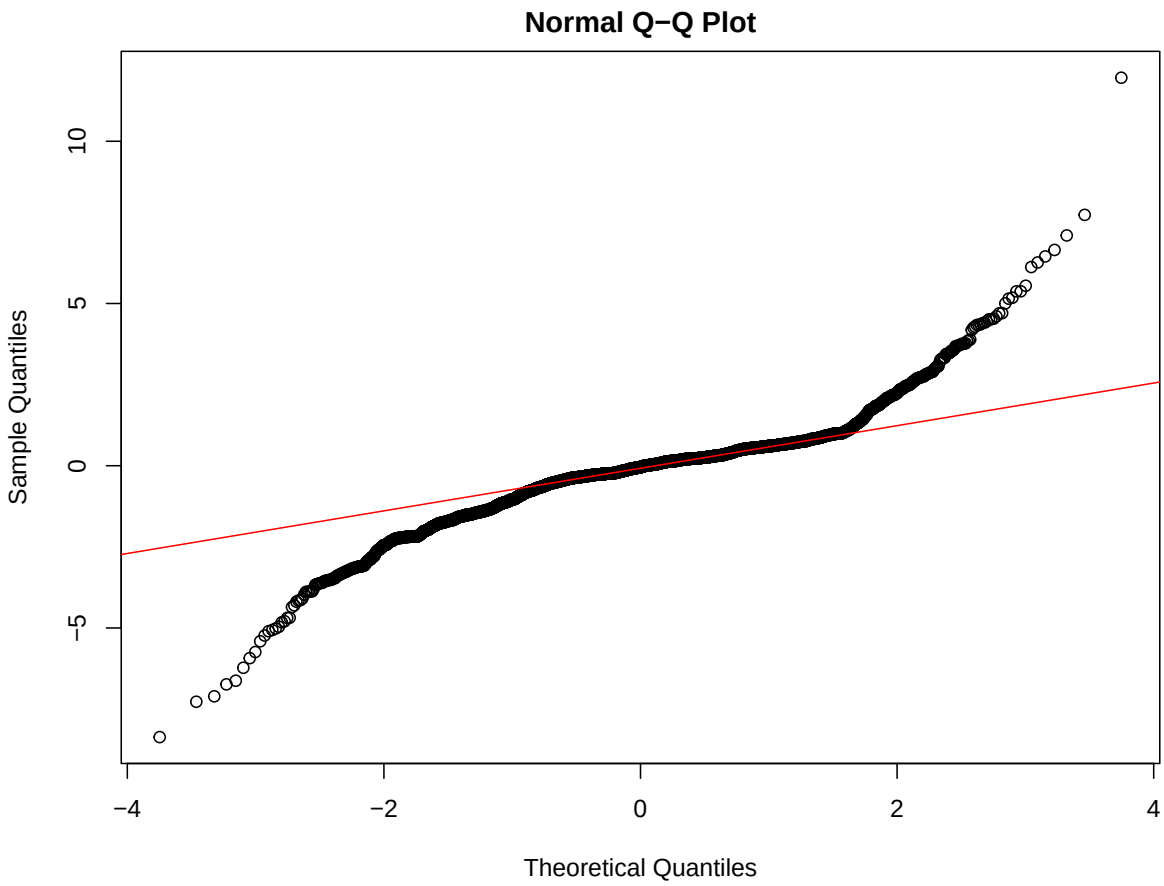

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
500
1000
1500
2000
Fitted values

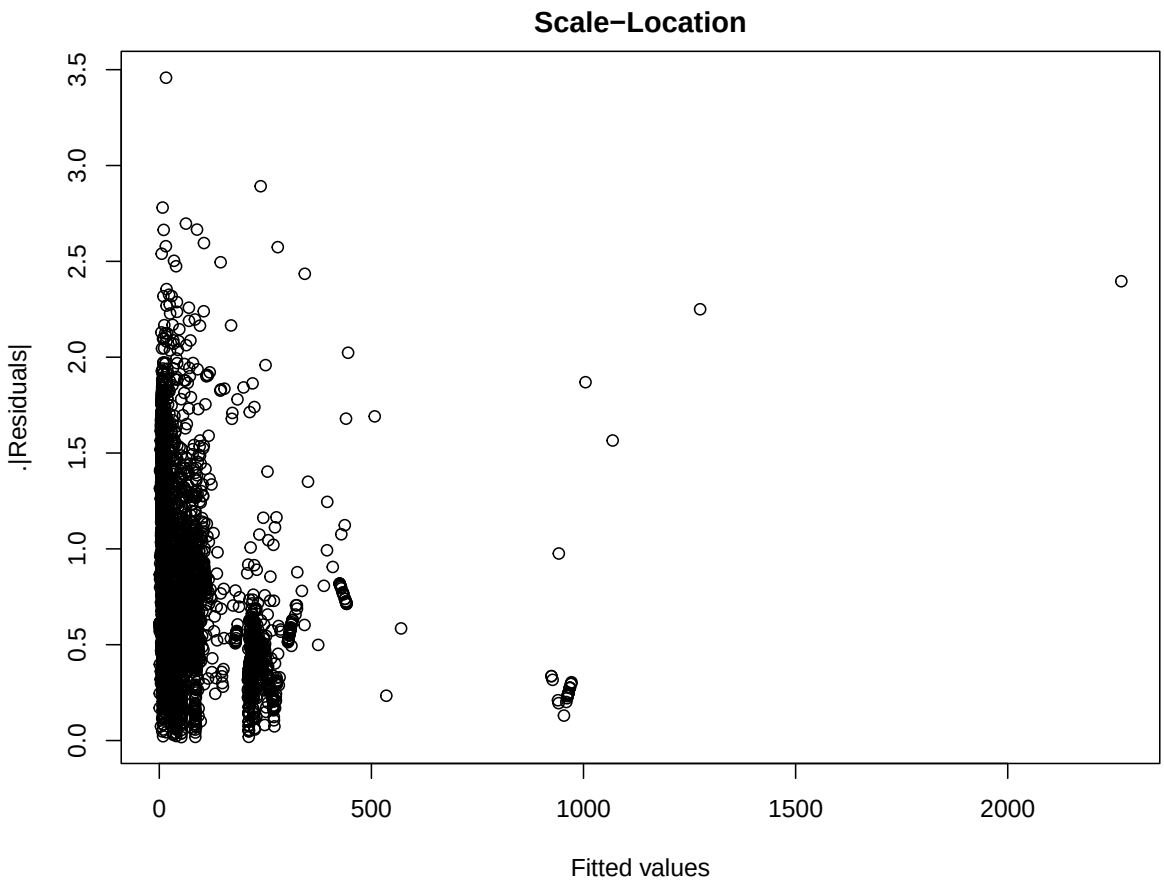


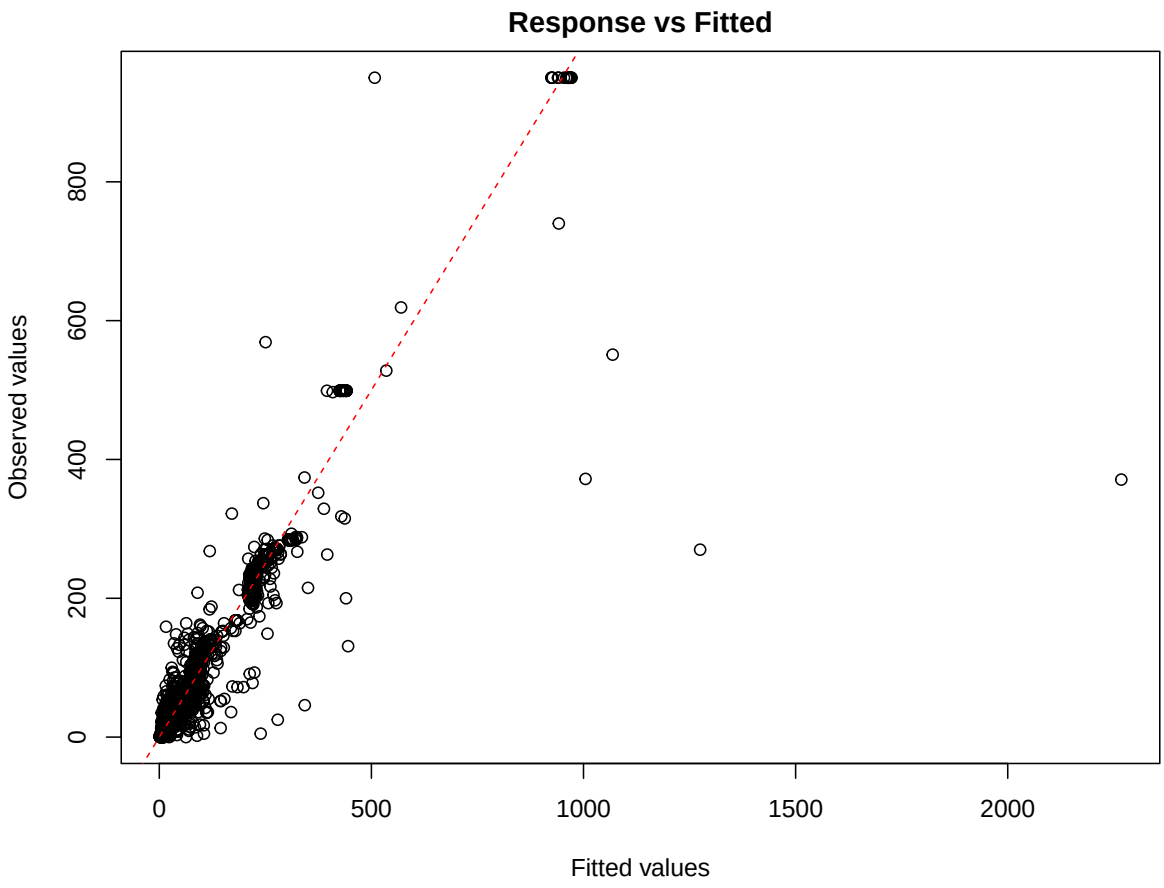

```
### Model 8 : 3d. All additive

Model data dimensions after removing NAs: 5601 x 6
Variables in model: total_butterflies, abundance_t_minus_1, mean_wind_speed, mean_temp, sunli
View IDs with observations: 9
Min/Max observations per view: 13 / 2941
GAM fitted successfully
Model Summary:

Family: Negative Binomial(17.208)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + s(view_id, bs = "re")
```

```
Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3580     0.1668   20.14   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                         edf Ref.df   Chi.sq  p-value
s(abundance_t_minus_1) 8.962  8.999 58294.08  < 2e-16 ***
s(mean_wind_speed)     6.076  7.191    32.73 8.98e-05 ***
s(mean_temp)           7.830  8.465   114.95  < 2e-16 ***
s(sunlight_proportion) 8.263  8.843   194.49  < 2e-16 ***
s(view_id)             7.779  8.000   119.55  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.862   Deviance explained = 93.3%
-REML =  21550  Scale est. = 1         n = 5601


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
10
5
0
−5
0
500
1000
1500
2000

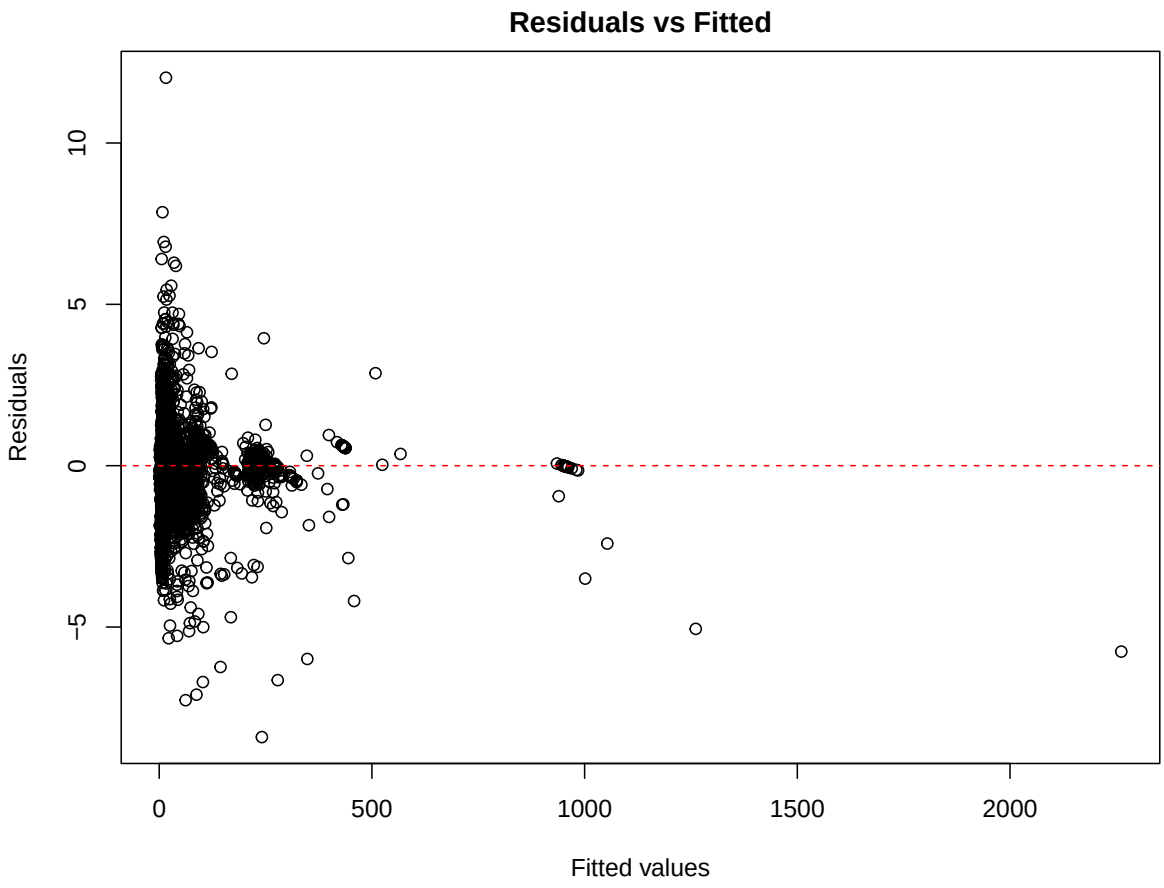

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4

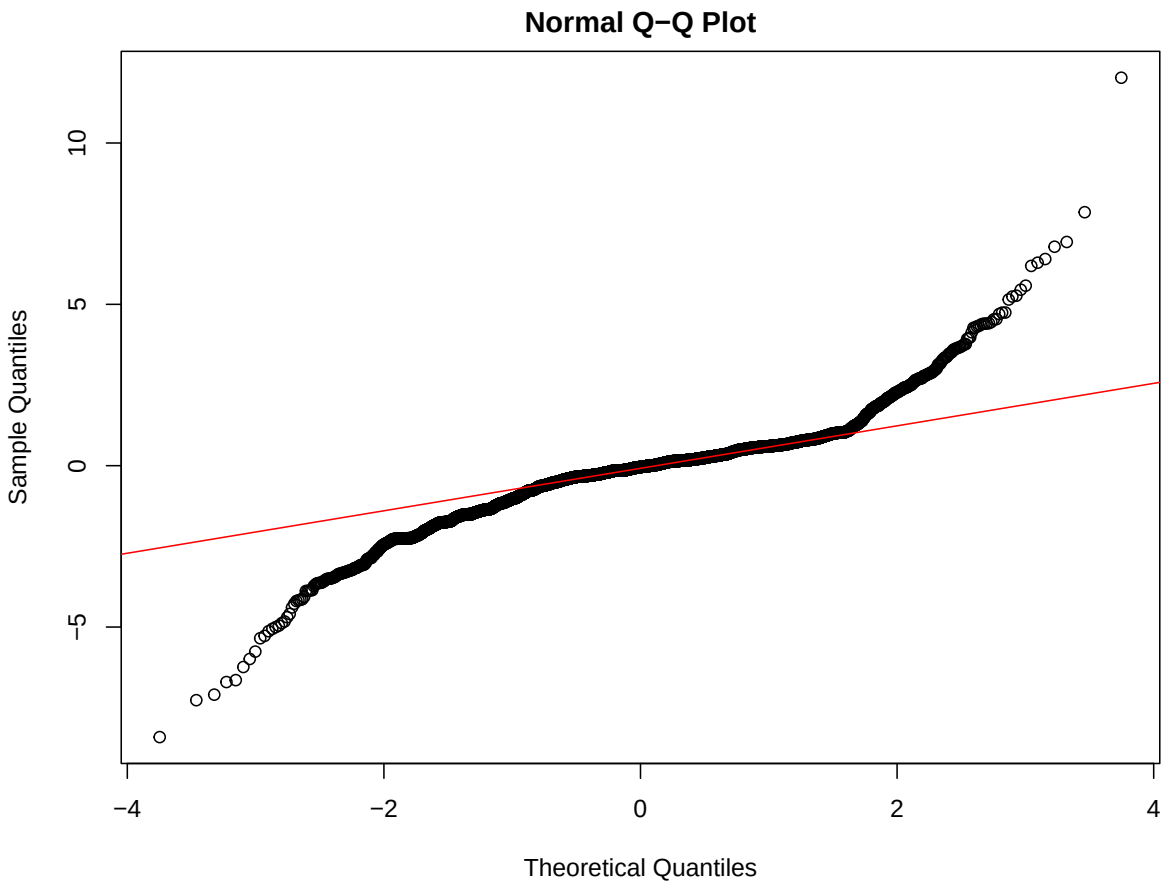

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
500
1000
1500
2000
Fitted values

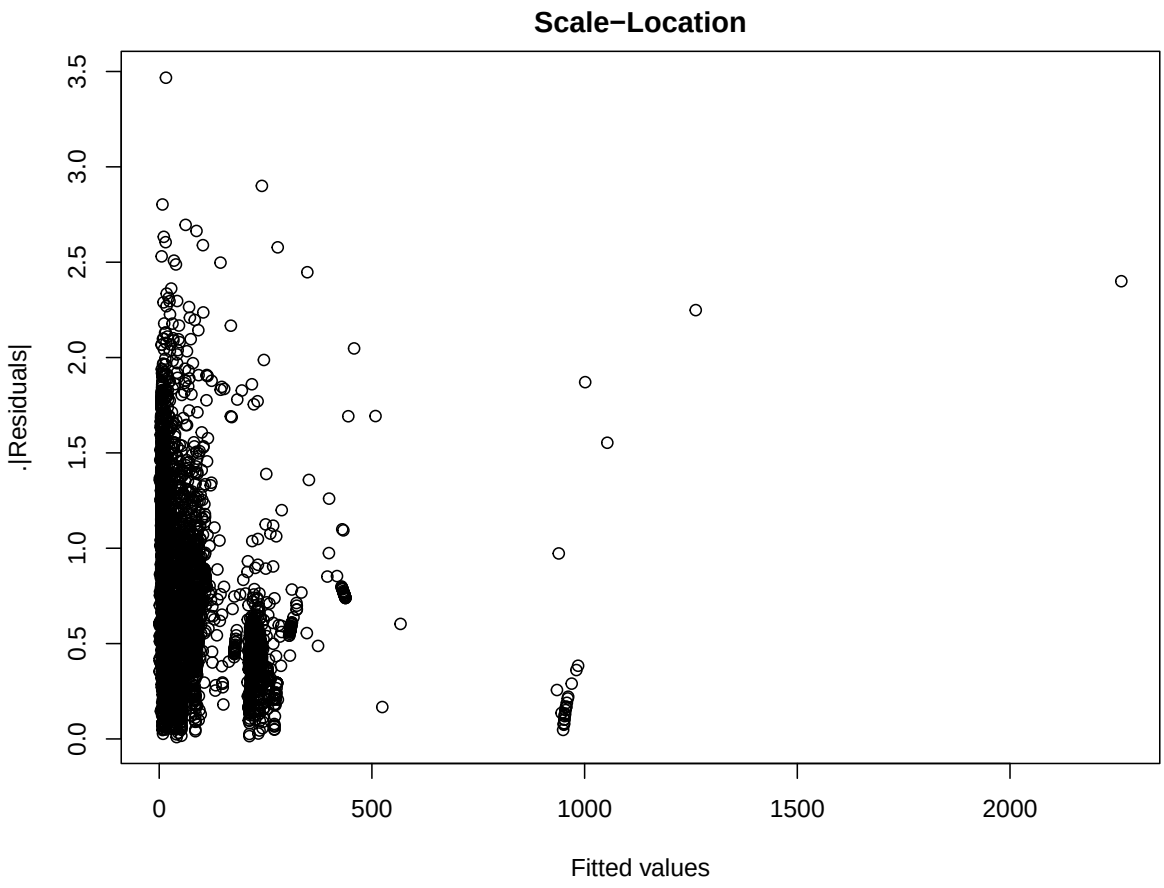


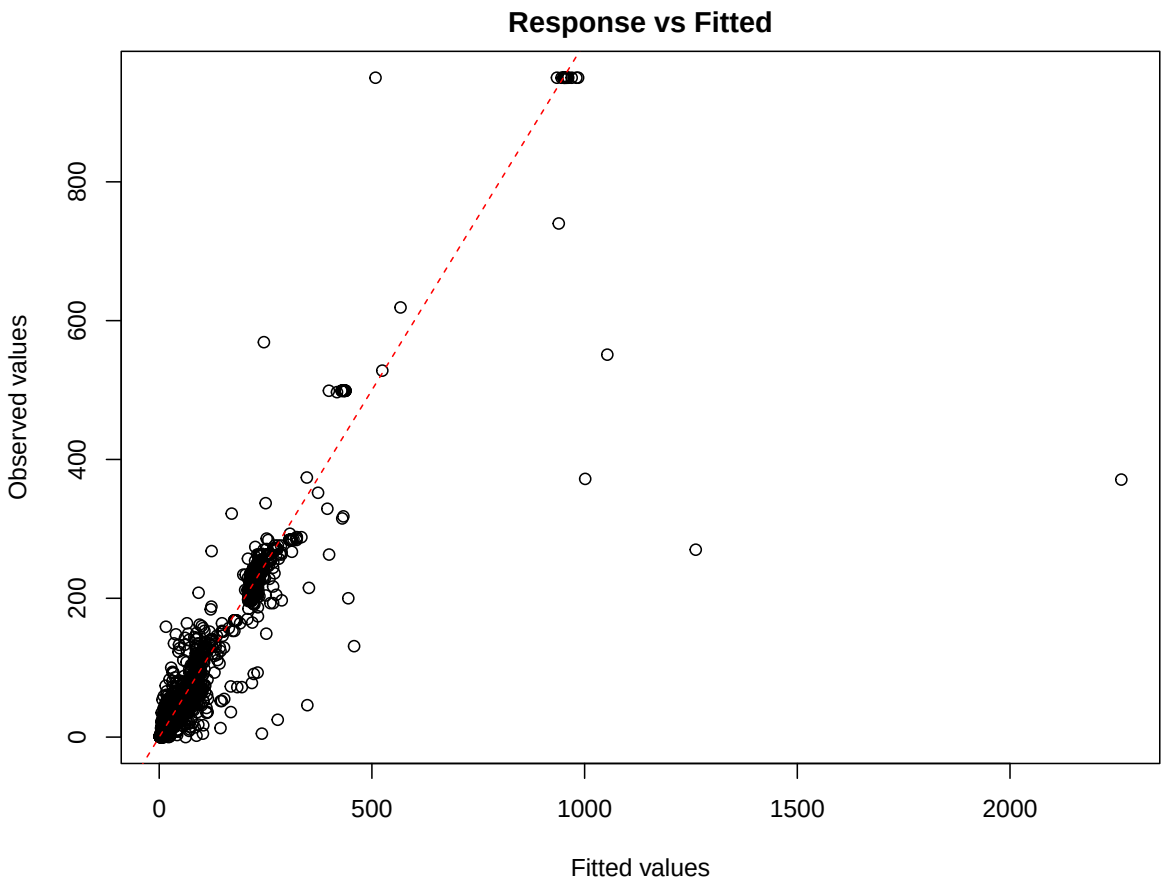

```
### Model 9 : 4a. Wind x Temp interaction

Model data dimensions after removing NAs: 5601 x 6
Variables in model: total_butterflies, abundance_t_minus_1, mean_wind_speed, mean_temp, sunli
View IDs with observations: 9
Min/Max observations per view: 13 / 2941
GAM fitted successfully
Model Summary:

Family: Negative Binomial(17.364)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + s(view_id, bs = "re")
```

```
Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3600     0.1681   19.98   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                edf Ref.df   Chi.sq  p-value
s(abundance_t_minus_1)        8.962  8.999 55680.65  < 2e-16 ***
s(mean_wind_speed)            5.213  6.329    26.29 0.000187 ***
s(mean_temp)                  7.941  8.546   114.56  < 2e-16 ***
s(sunlight_proportion)        8.268  8.844   196.75  < 2e-16 ***
ti(mean_wind_speed,mean_temp) 8.436  9.988    34.35 0.000156 ***
s(view_id)                    7.780  8.000   122.40  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.863   Deviance explained = 93.4%
-REML =  21541  Scale est. = 1         n = 5601


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
10
5
0
−5
0
500
1000
1500
2000

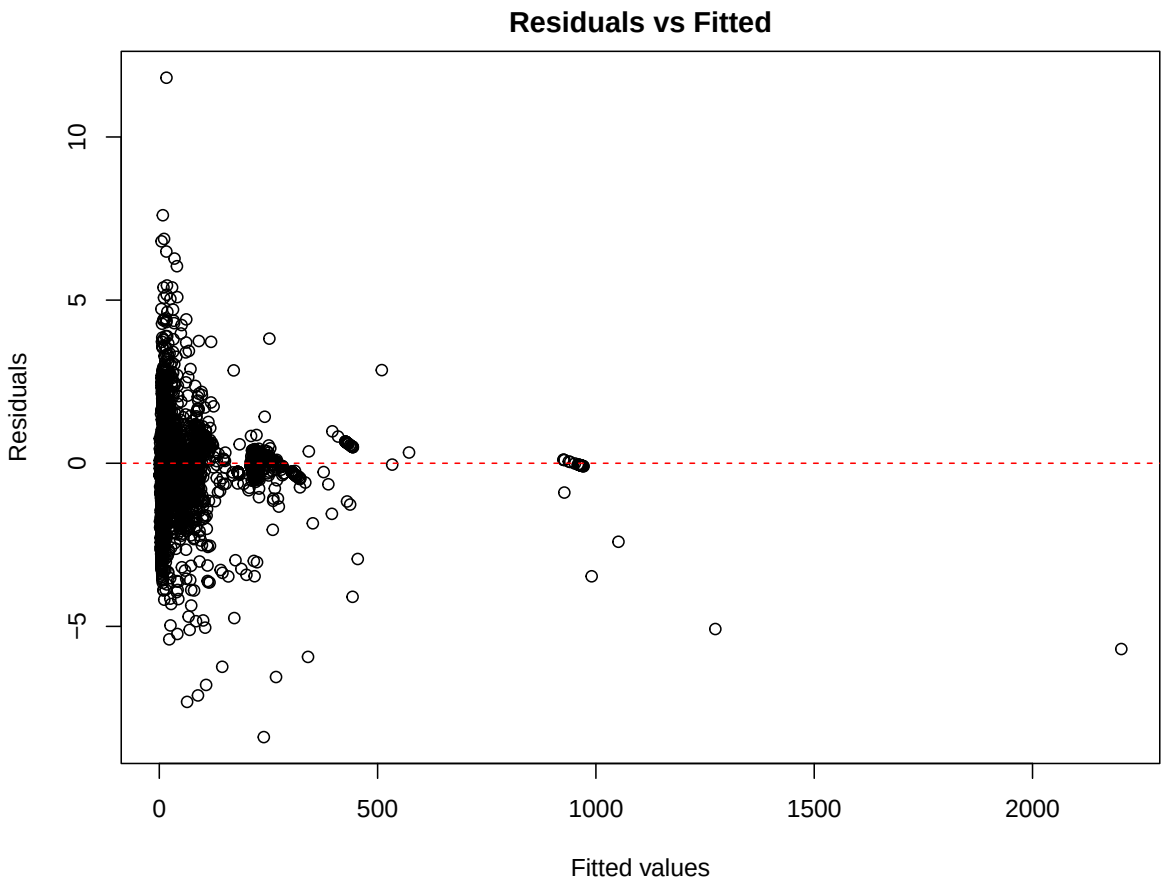

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4

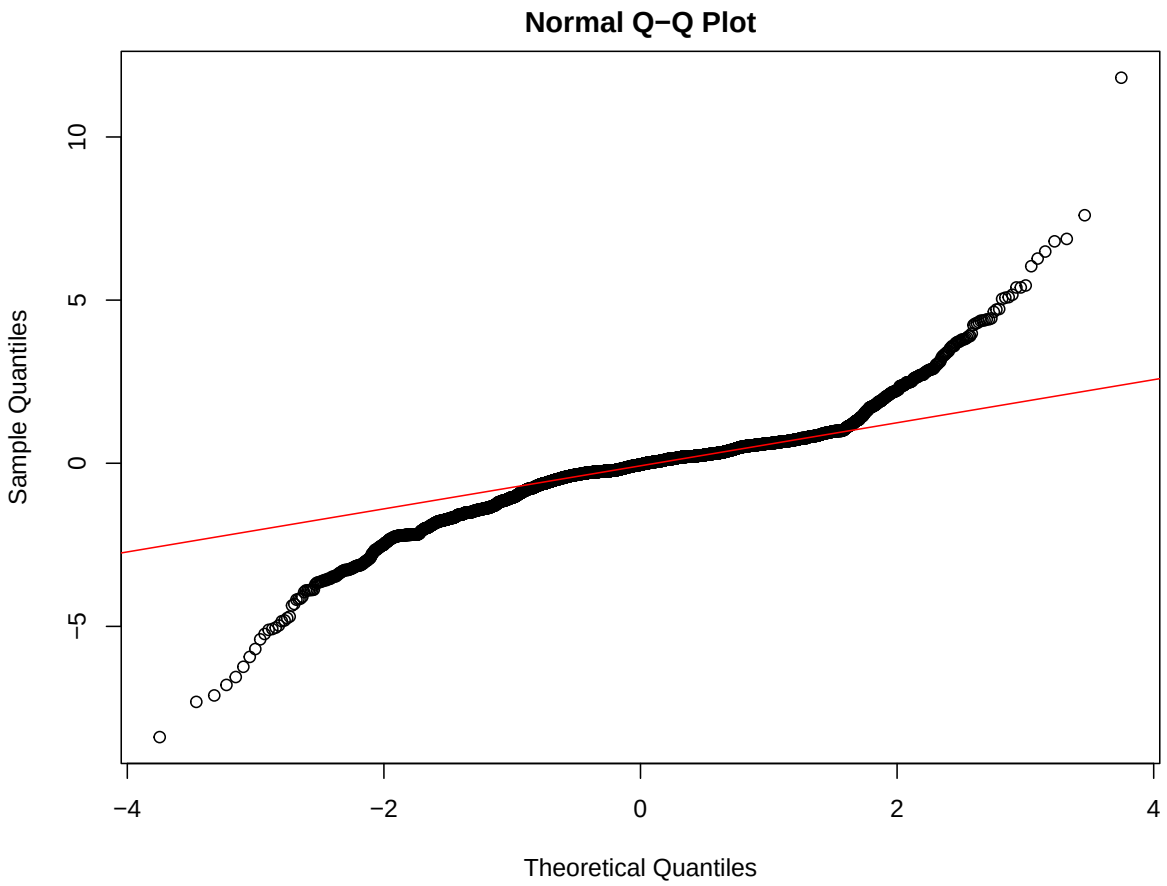

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
500
1000
1500
2000
Fitted values

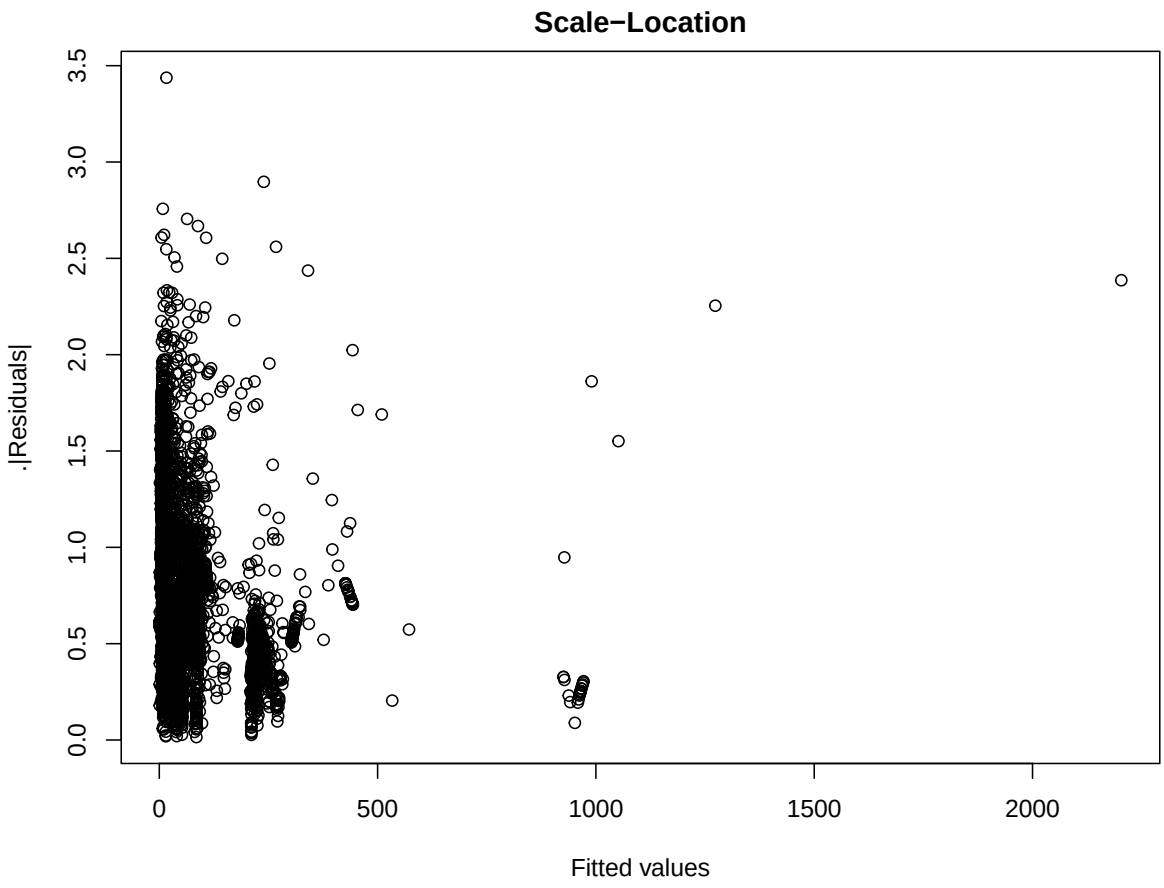


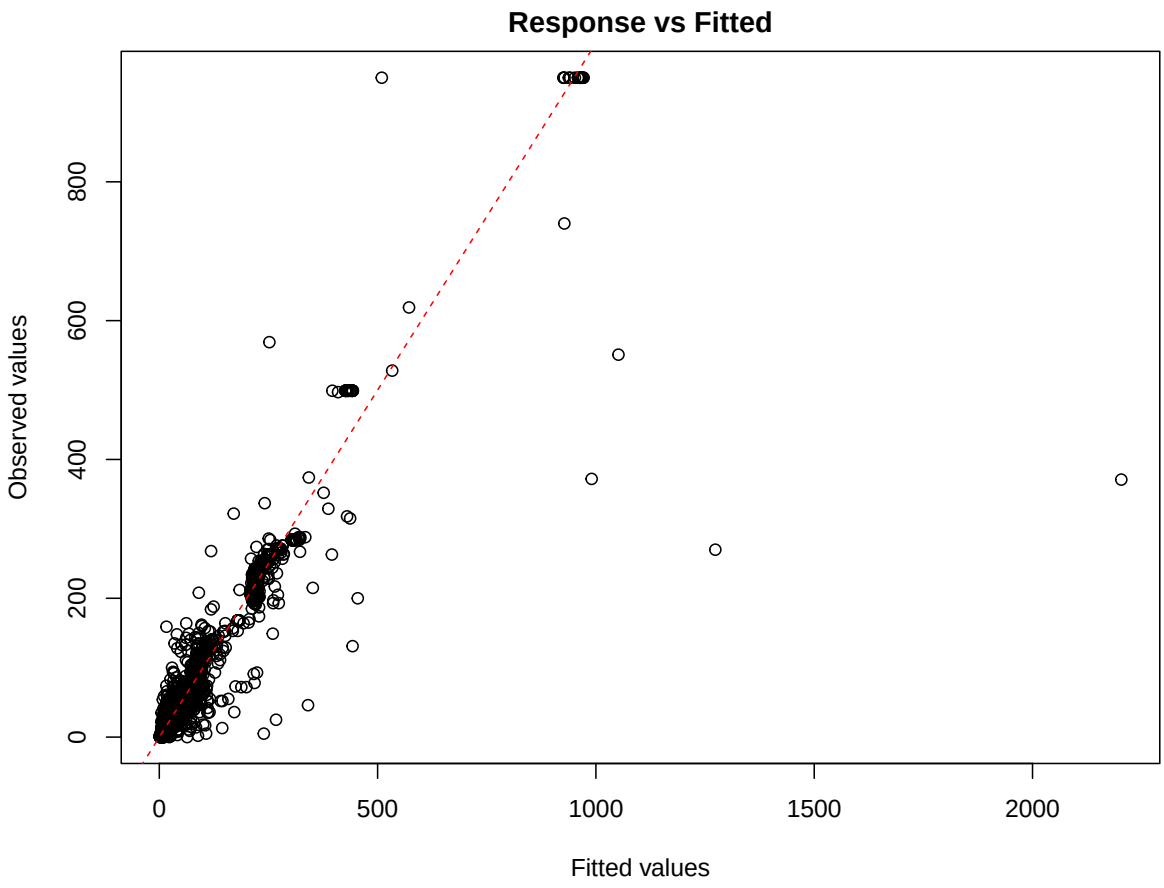

```
### Model 10 : 4b. Wind x Sunlight interaction

Model data dimensions after removing NAs: 5601 x 6
Variables in model: total_butterflies, abundance_t_minus_1, mean_wind_speed, mean_temp, sunli
View IDs with observations: 9
Min/Max observations per view: 13 / 2941
GAM fitted successfully
Model Summary:

Family: Negative Binomial(17.357)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    sunlight_proportion) + s(view_id, bs = "re")
```

```
Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3558     0.1665   20.15   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                          edf Ref.df   Chi.sq  p-value
s(abundance_t_minus_1)                  8.962  8.999 58232.79  < 2e-16 ***
s(mean_wind_speed)                      5.932  7.052    33.17 1.36e-05 ***
s(mean_temp)                            7.870  8.491   114.82  < 2e-16 ***
s(sunlight_proportion)                  8.219  8.822   191.28  < 2e-16 ***
ti(mean_wind_speed,sunlight_proportion) 7.468  9.472    24.30  0.00599 **
s(view_id)                              7.778  8.000   119.47  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.867   Deviance explained = 93.4%
-REML =  21545  Scale est. = 1         n = 5601


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
10
5
0
−5
0
500
1000
1500
2000

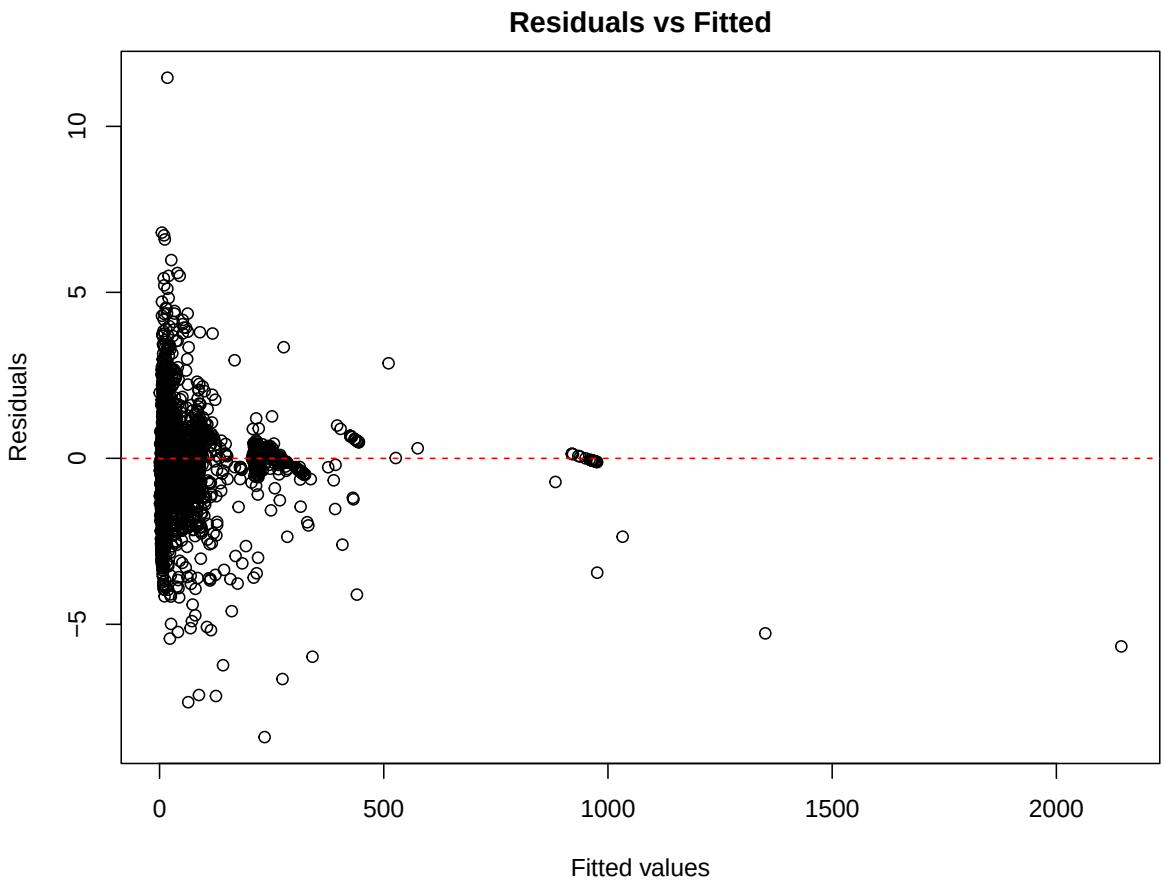

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4

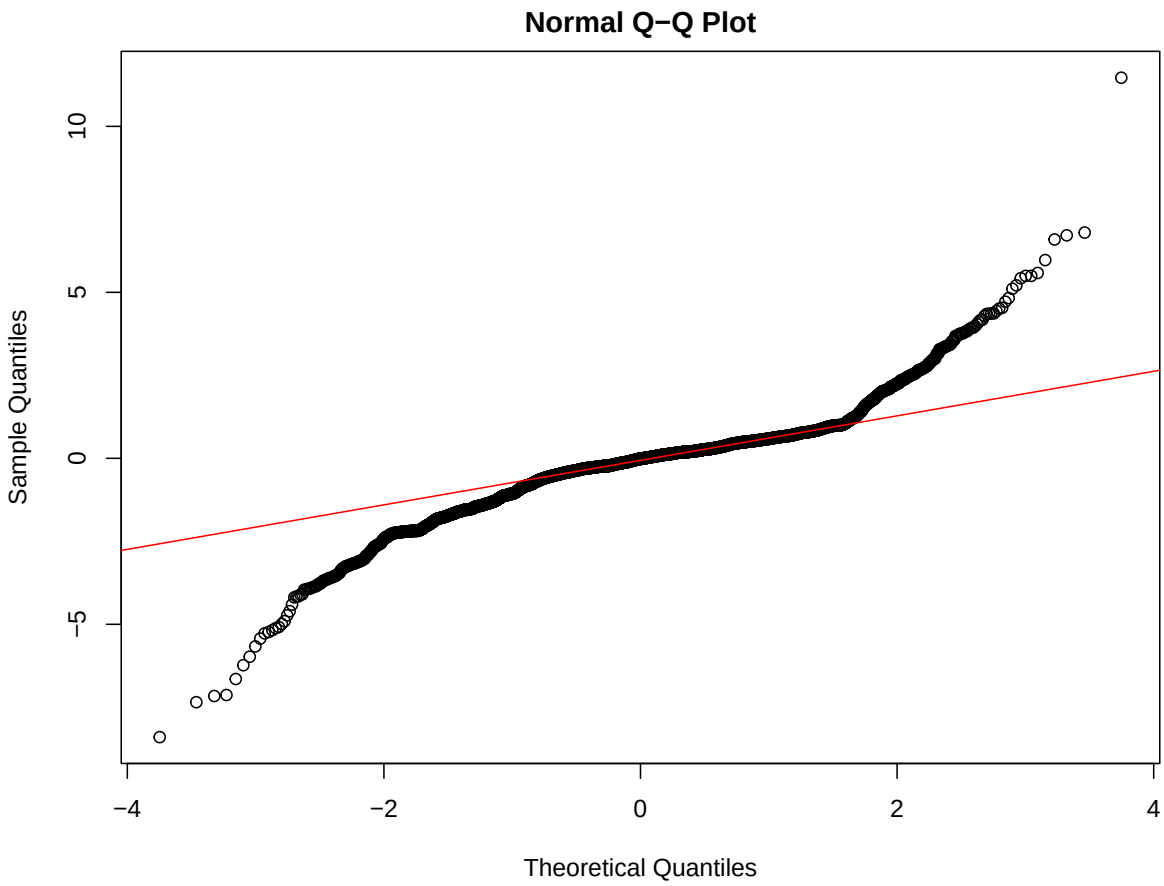

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
500
1000
1500
2000
Fitted values

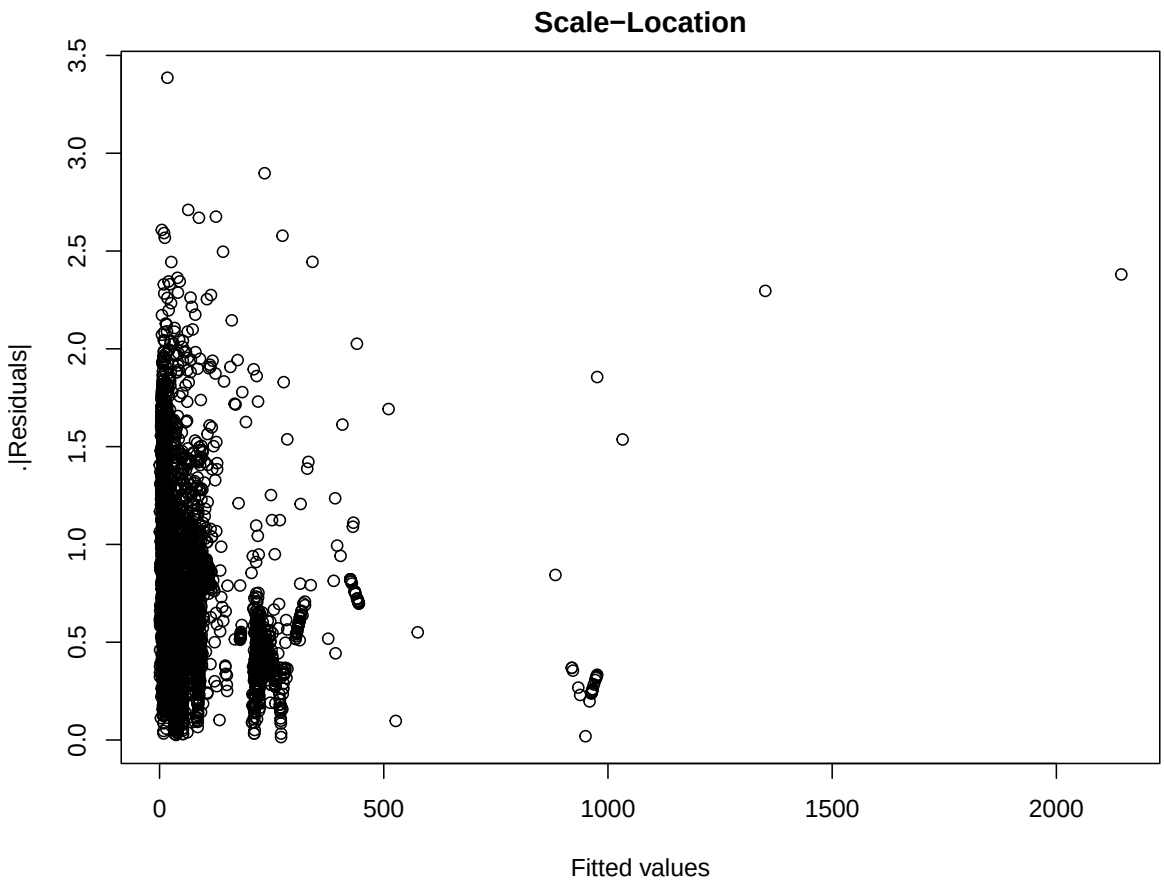


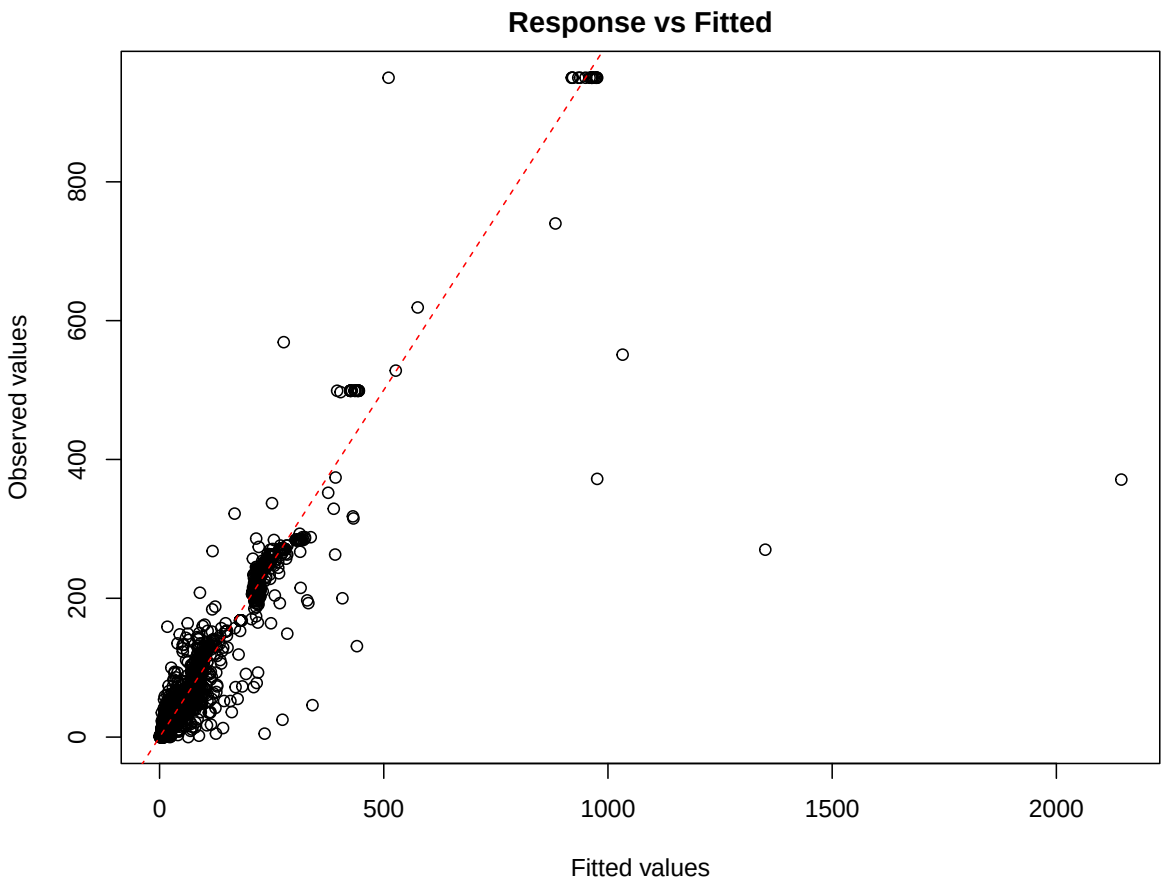

```
### Model 11 : 4c. Temp x Sunlight interaction

Model data dimensions after removing NAs: 5601 x 6
Variables in model: total_butterflies, abundance_t_minus_1, mean_wind_speed, mean_temp, sunli
View IDs with observations: 9
Min/Max observations per view: 13 / 2941
GAM fitted successfully
Model Summary:

Family: Negative Binomial(17.604)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_temp, sunlight_proportion) +
    s(view_id, bs = "re")
```

```
Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3396     0.1645    20.3   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                     edf Ref.df   Chi.sq  p-value
s(abundance_t_minus_1)             8.962  8.999 57694.24  < 2e-16 ***
s(mean_wind_speed)                 3.744  4.653    23.78 0.000232 ***
s(mean_temp)                       7.968  8.539    91.88  < 2e-16 ***
s(sunlight_proportion)             8.020  8.729   105.97  < 2e-16 ***
ti(mean_temp,sunlight_proportion) 10.293 12.324   114.45  < 2e-16 ***
s(view_id)                         7.775  8.000   122.49  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.869   Deviance explained = 93.5%
-REML =  21505  Scale est. = 1         n = 5601


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
10
5
0
−5
0
500
1000
1500
2000

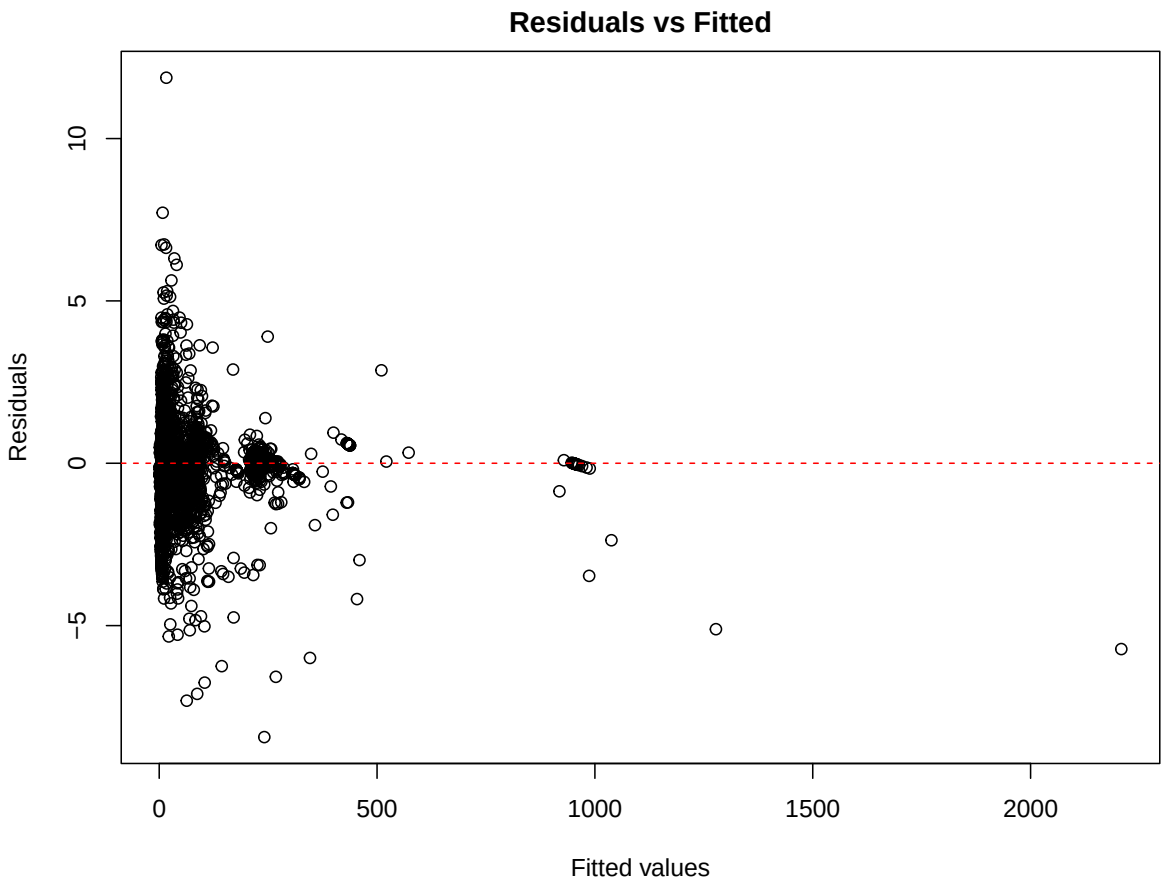

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4

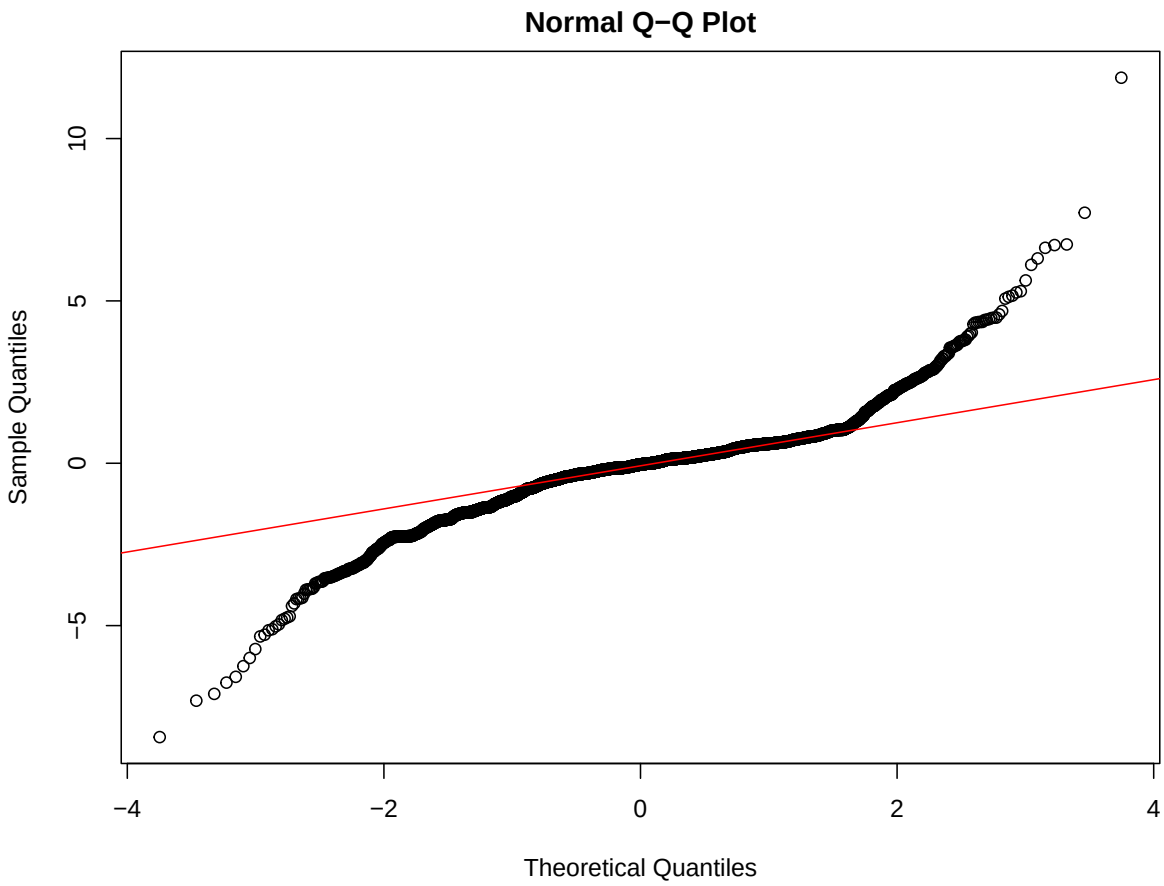

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
500
1000
1500
2000
Fitted values

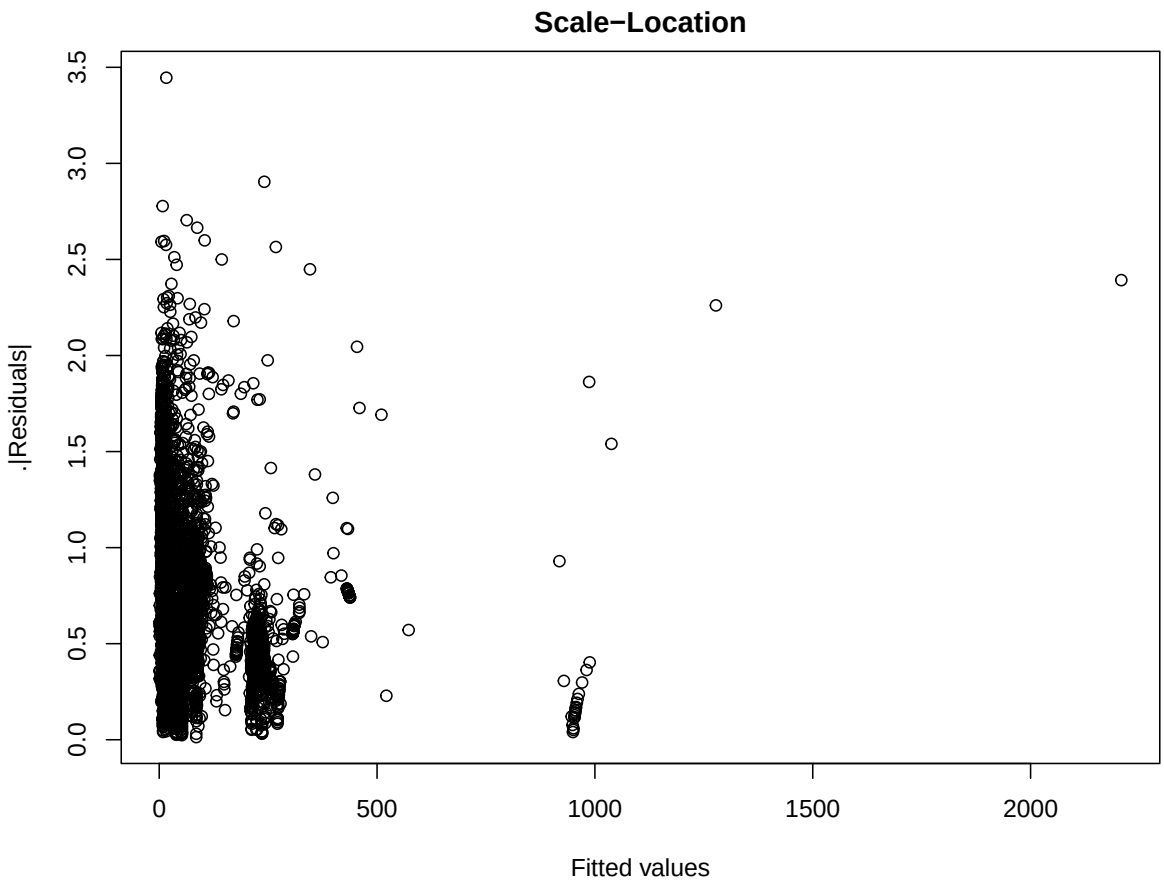


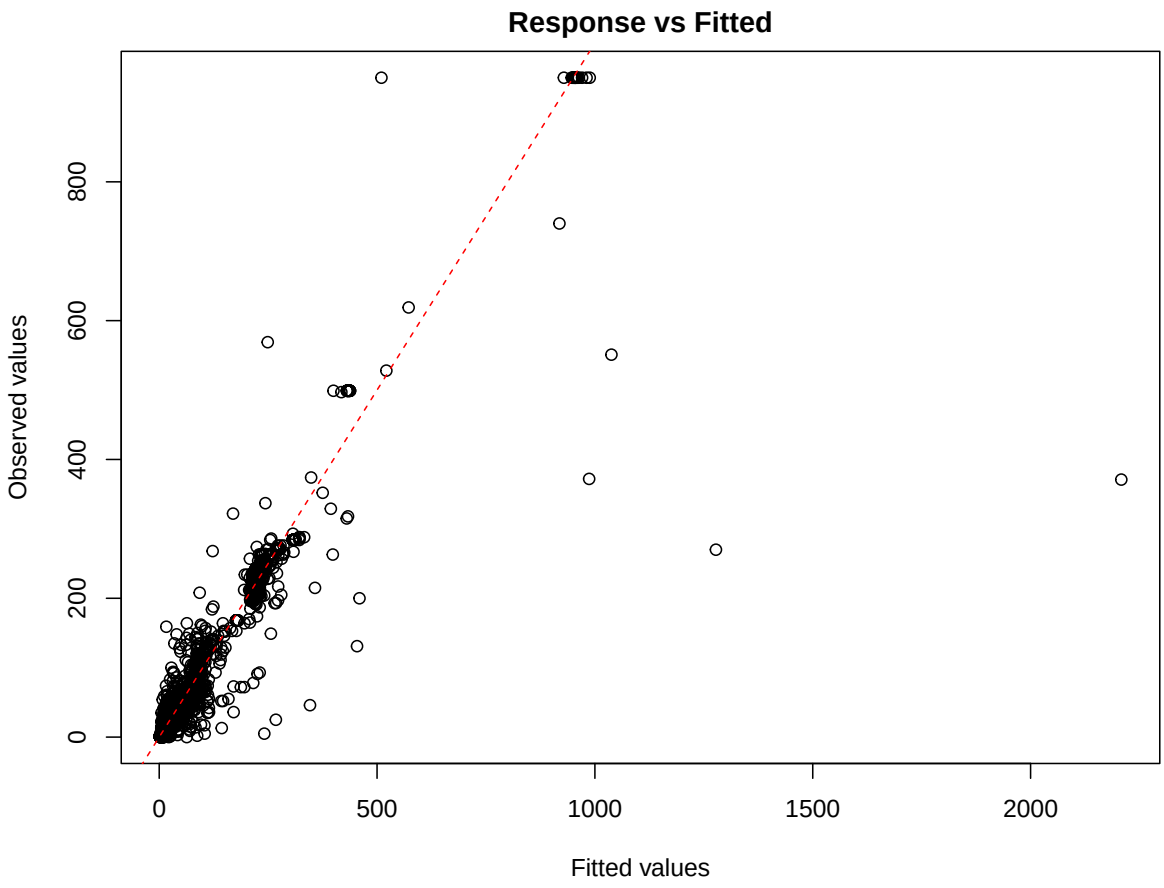

```
### Model 12 : 5a. Wind x Temp + Wind x Sun

Model data dimensions after removing NAs: 5601 x 6
Variables in model: total_butterflies, abundance_t_minus_1, mean_wind_speed, mean_temp, sunli
View IDs with observations: 9
Min/Max observations per view: 13 / 2941
GAM fitted successfully
Model Summary:

Family: Negative Binomial(17.501)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + ti(mean_wind_speed, sunlight_proportion) + s(view_id,
```

```
    bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3564     0.1676   20.03   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                          edf Ref.df   Chi.sq  p-value
s(abundance_t_minus_1)                  8.962  8.999 55404.29  < 2e-16 ***
s(mean_wind_speed)                      5.434  6.557    30.81 2.98e-05 ***
s(mean_temp)                            7.949  8.551   112.97  < 2e-16 ***
s(sunlight_proportion)                  8.230  8.826   193.99  < 2e-16 ***
ti(mean_wind_speed,mean_temp)           8.551 10.073    35.09 0.000197 ***
ti(mean_wind_speed,sunlight_proportion) 7.104  9.085    22.07 0.009670 **
s(view_id)                              7.779  8.000   121.95  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.867   Deviance explained = 93.4%
-REML =  21536  Scale est. = 1         n = 5601


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
10
5
0
−5
0
500
1000
1500
2000

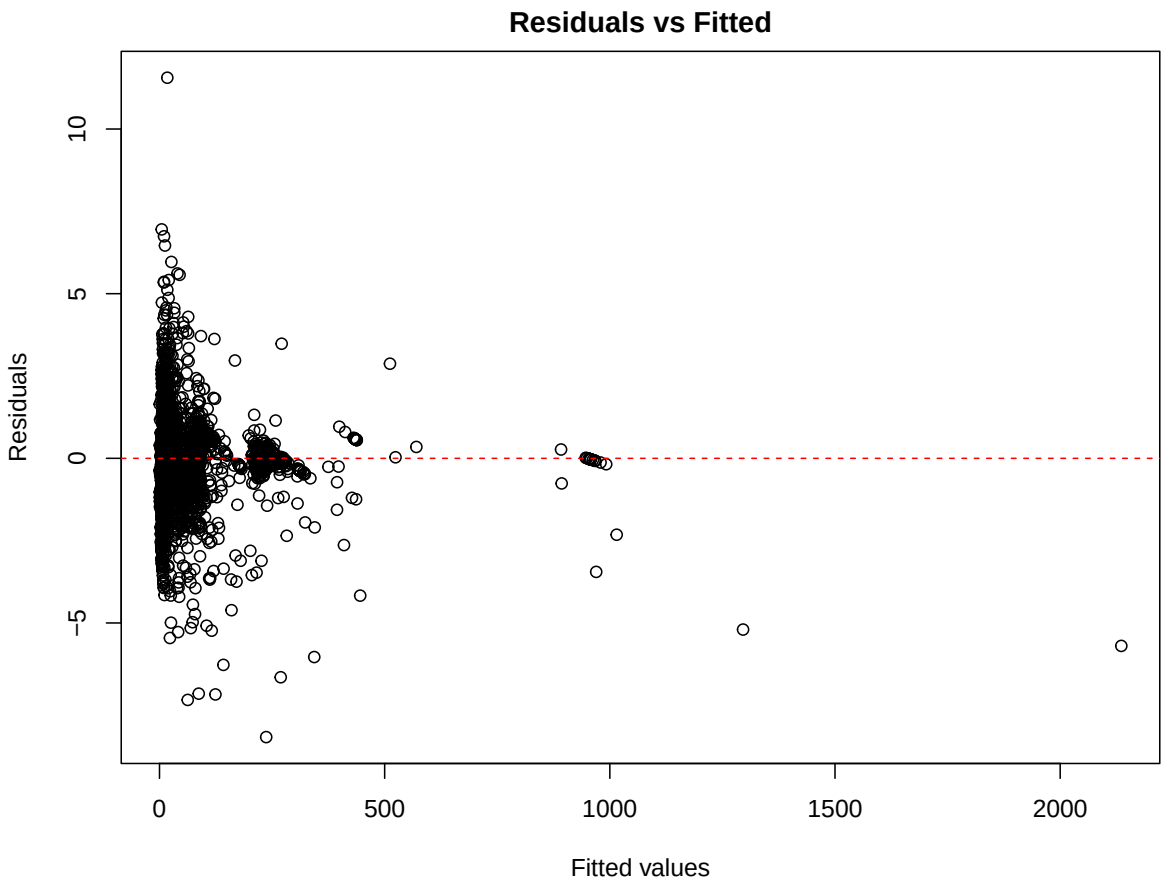

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4

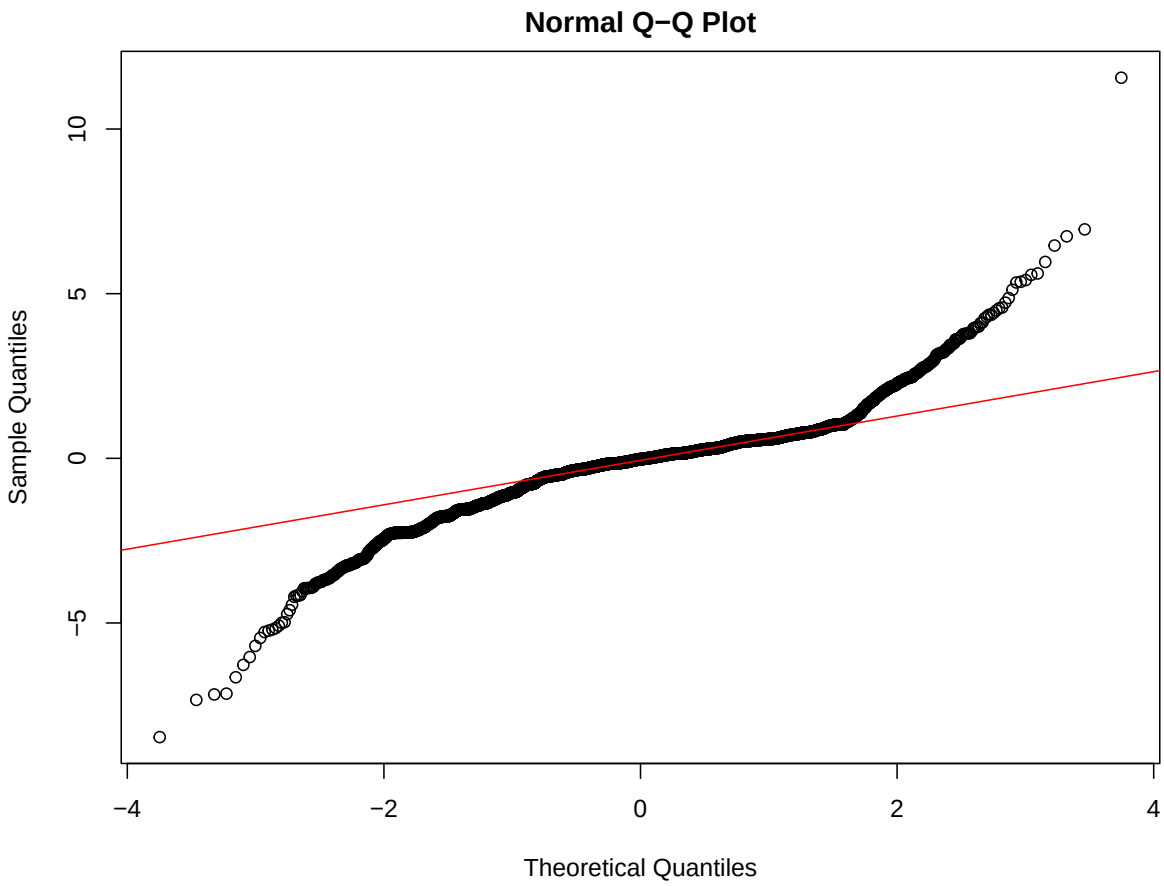

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
500
1000
1500
2000
Fitted values

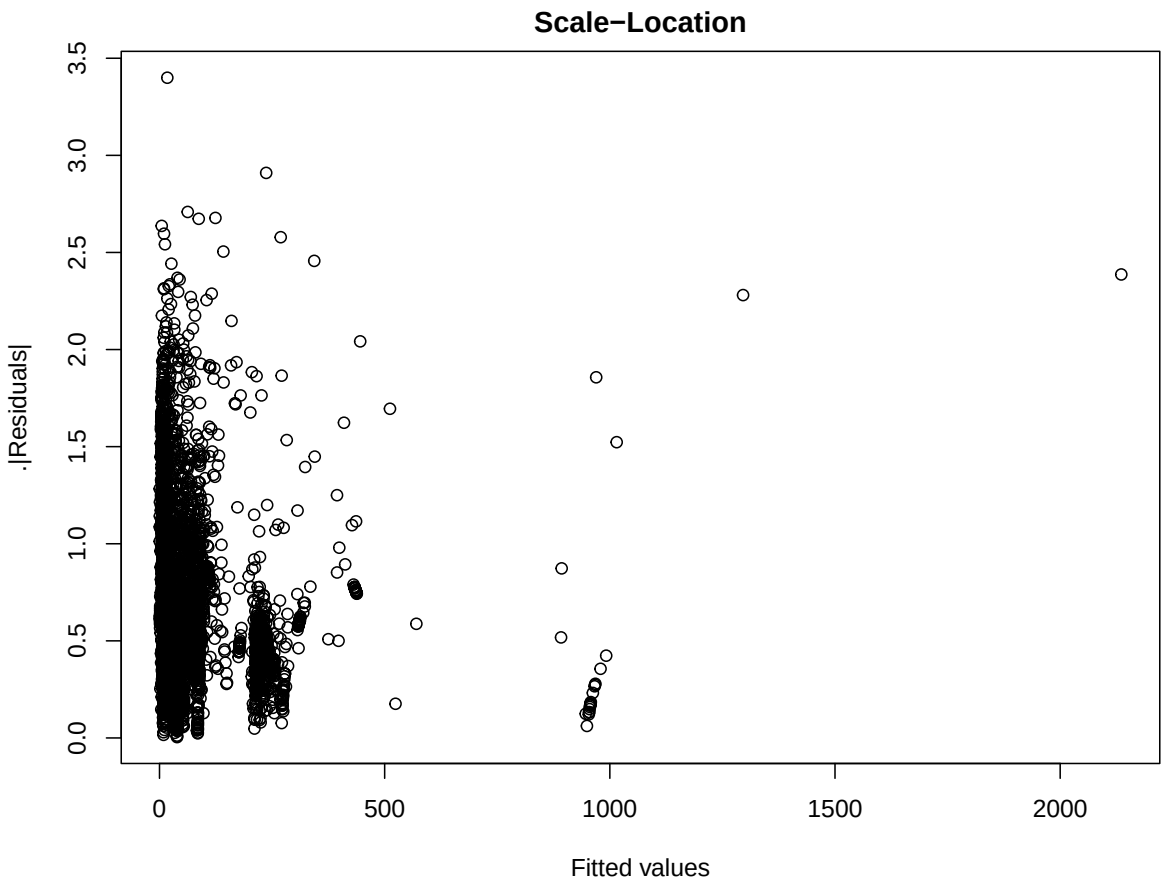


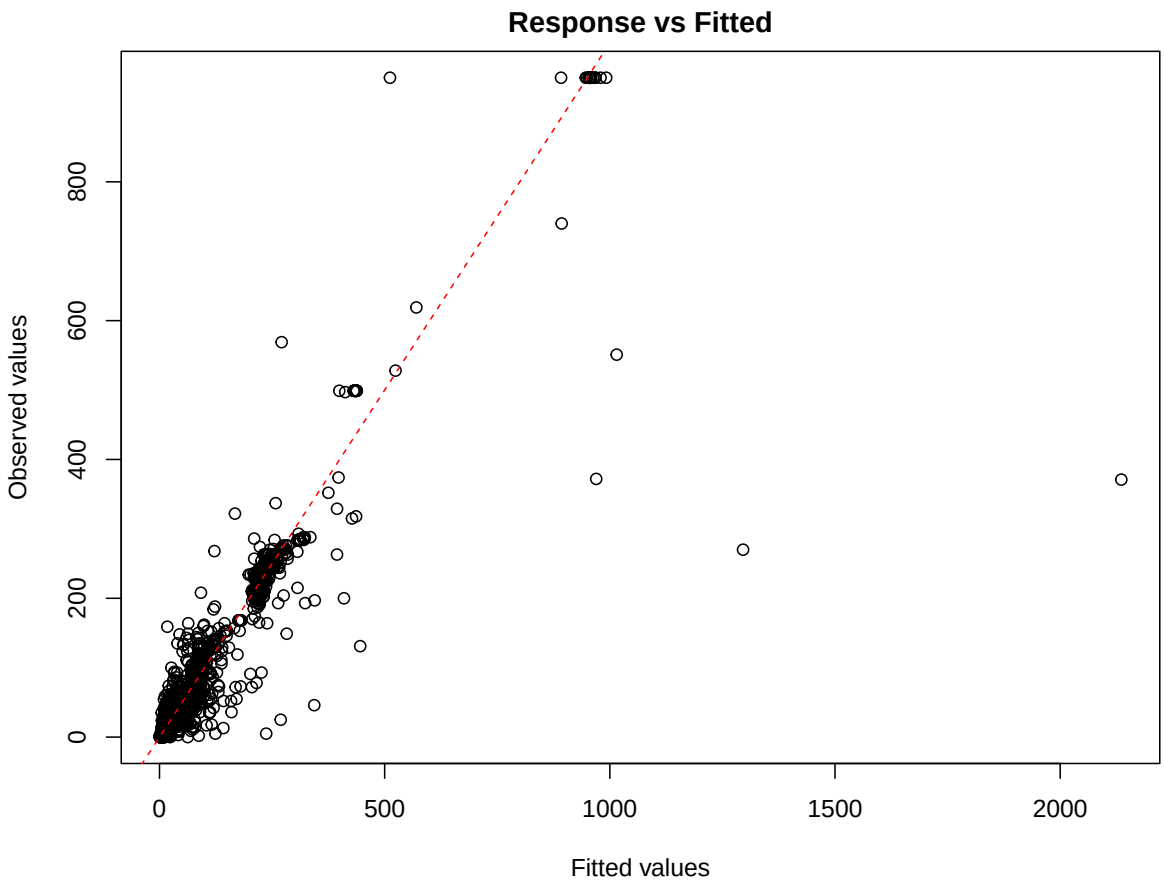

```
### Model 13 : 5b. Wind x Temp + Temp x Sun

Model data dimensions after removing NAs: 5601 x 6
Variables in model: total_butterflies, abundance_t_minus_1, mean_wind_speed, mean_temp, sunli
View IDs with observations: 9
Min/Max observations per view: 13 / 2941
GAM fitted successfully
Model Summary:

Family: Negative Binomial(17.868)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + ti(mean_temp, sunlight_proportion) + s(view_id,
```

```
    bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3476     0.1651   20.28   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                    edf Ref.df   Chi.sq  p-value
s(abundance_t_minus_1)            8.962  8.999 55365.60  < 2e-16 ***
s(mean_wind_speed)                4.277  5.276    32.90 6.06e-06 ***
s(mean_temp)                      8.131  8.653    99.42  < 2e-16 ***
s(sunlight_proportion)            7.988  8.713   105.73  < 2e-16 ***
ti(mean_wind_speed,mean_temp)     6.088  7.367    43.92  < 2e-16 ***
ti(mean_temp,sunlight_proportion) 10.778 12.744   125.56  < 2e-16 ***
s(view_id)                        7.775  8.000   122.14  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.872   Deviance explained = 93.5%
-REML =  21490  Scale est. = 1         n = 5601


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
10
5
0
−5
0
500
1000
1500
2000

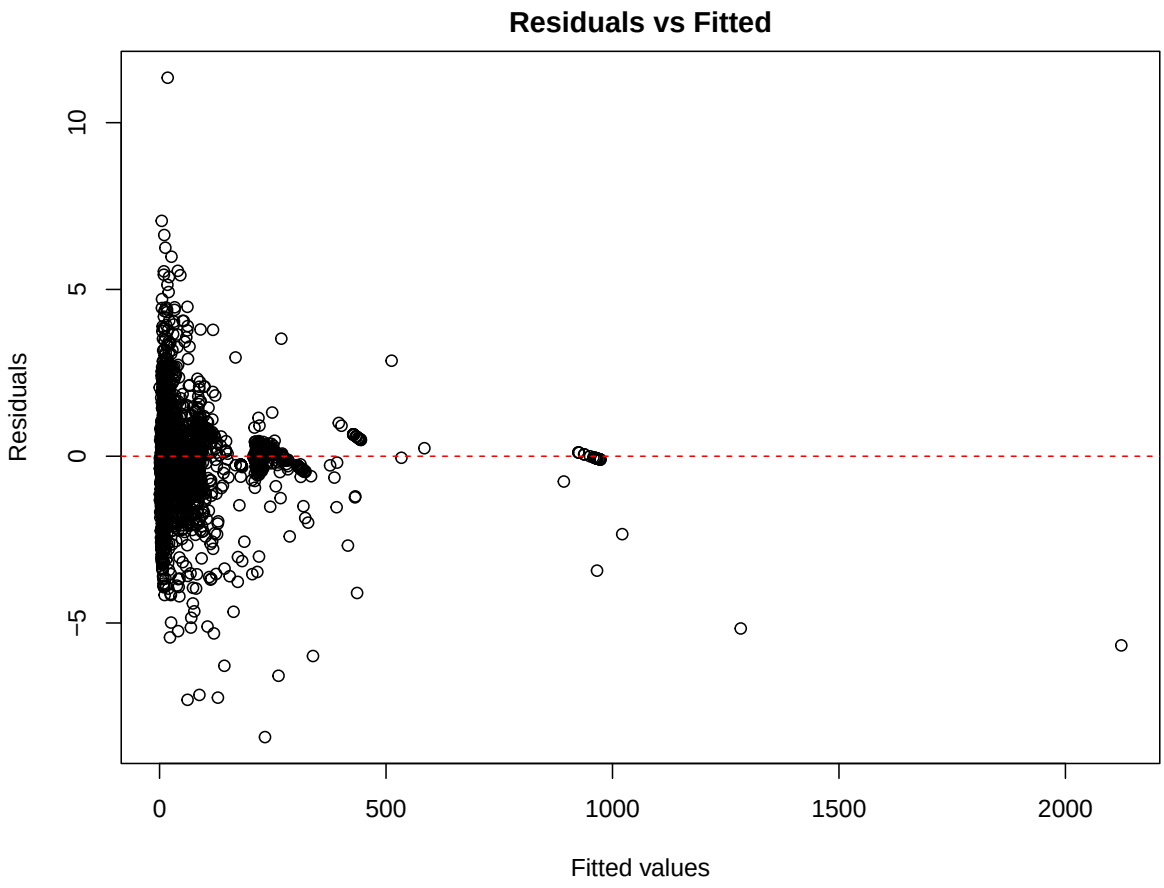

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4

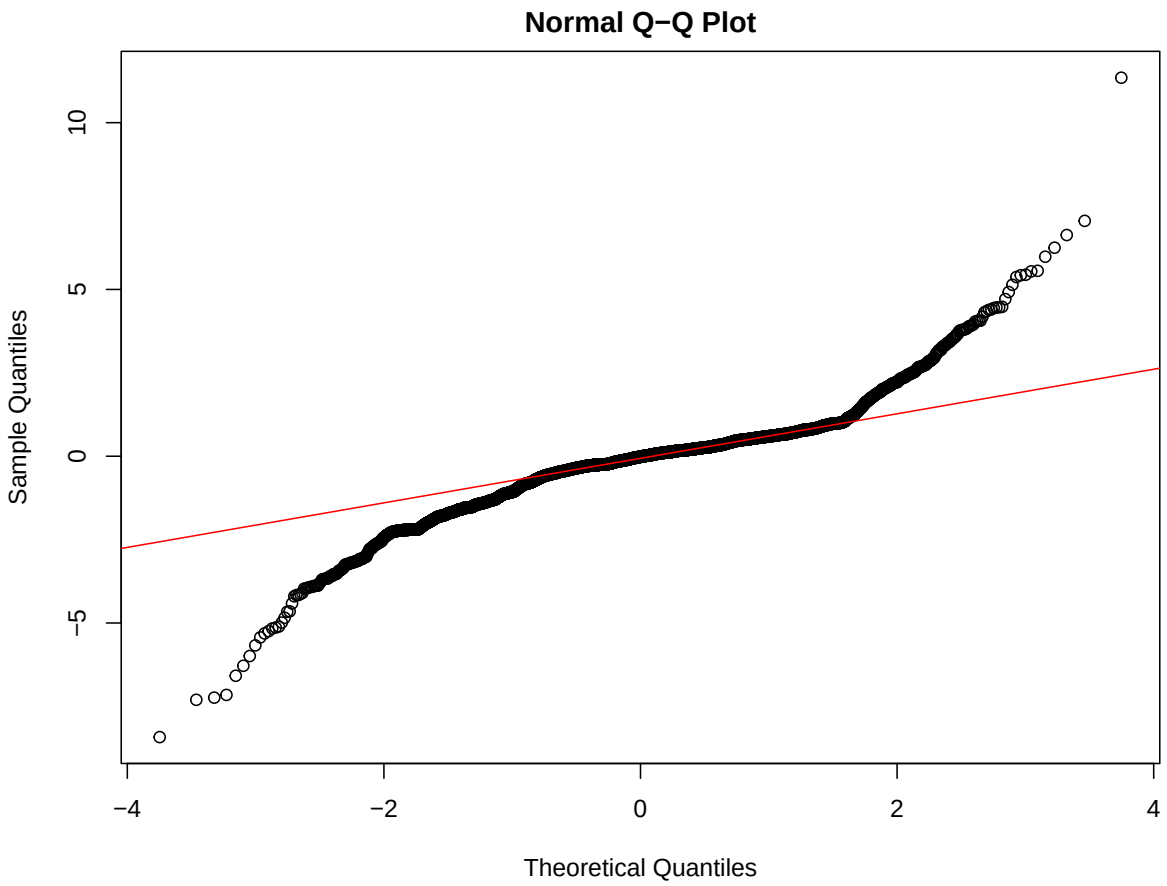

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
500
1000
1500
2000
Fitted values

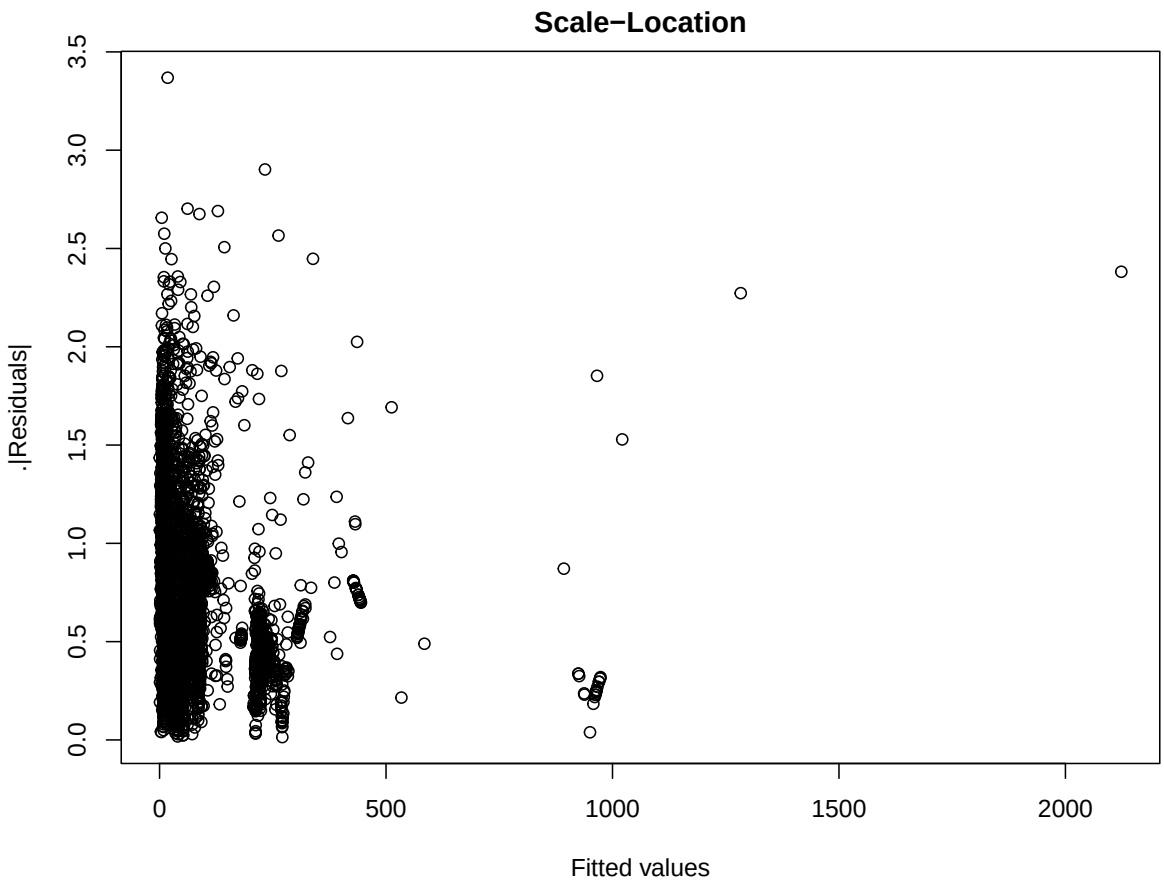


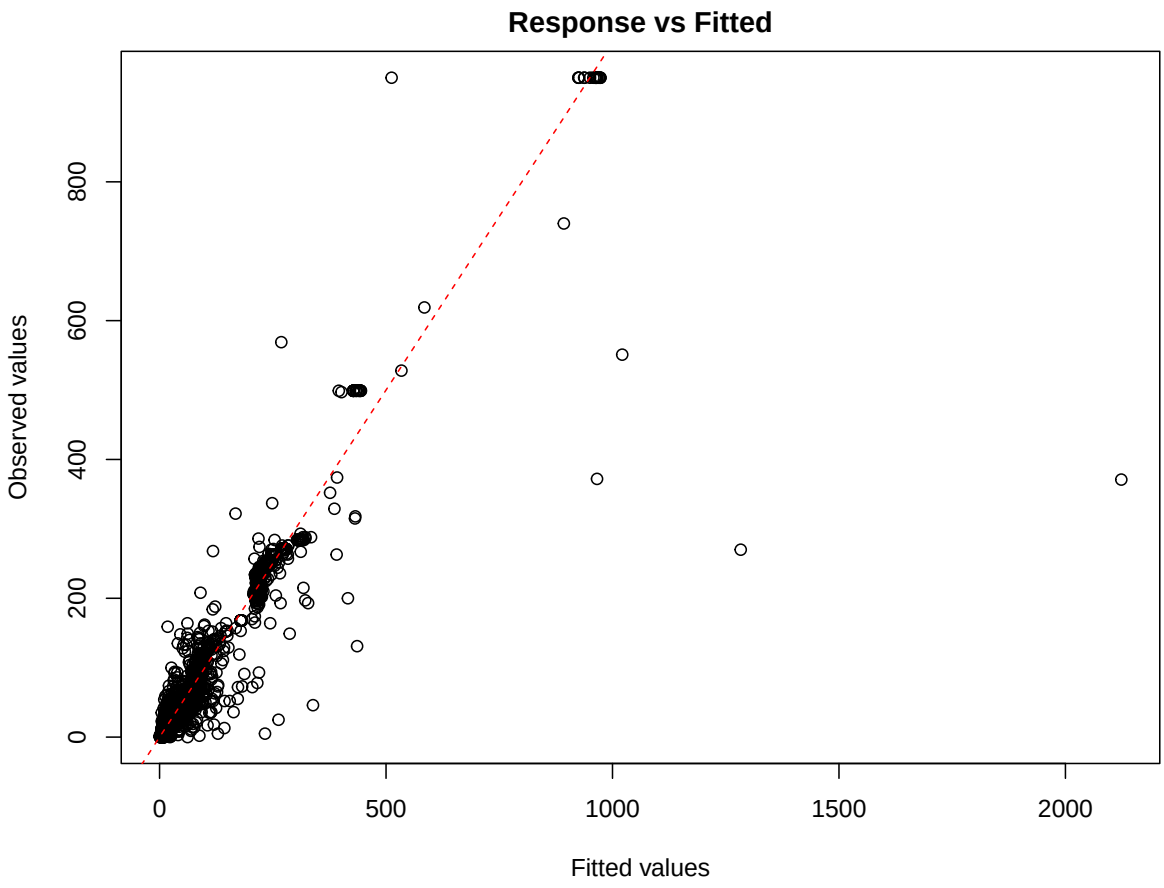

```
### Model 14 : 5c. Wind x Sun + Temp x Sun

Model data dimensions after removing NAs: 5601 x 6
Variables in model: total_butterflies, abundance_t_minus_1, mean_wind_speed, mean_temp, sunli
View IDs with observations: 9
Min/Max observations per view: 13 / 2941
GAM fitted successfully
Model Summary:

Family: Negative Binomial(17.806)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    sunlight_proportion) + ti(mean_temp, sunlight_proportion) +
```

```
    s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3403     0.1649   20.26   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                          edf Ref.df   Chi.sq p-value
s(abundance_t_minus_1)                  8.961  8.999 57751.72 < 2e-16 ***
s(mean_wind_speed)                      4.842  5.918    29.36 4.4e-05 ***
s(mean_temp)                            8.003  8.562    92.72 < 2e-16 ***
s(sunlight_proportion)                  7.974  8.704   102.09 < 2e-16 ***
ti(mean_wind_speed,sunlight_proportion) 7.490  9.463    25.29 0.00453 **
ti(mean_temp,sunlight_proportion)      10.474 12.454   113.47 < 2e-16 ***
s(view_id)                              7.776  8.000   124.17 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.874   Deviance explained = 93.5%
-REML =  21500  Scale est. = 1         n = 5601


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
10
5
0
−5
0
500
1000
1500
2000

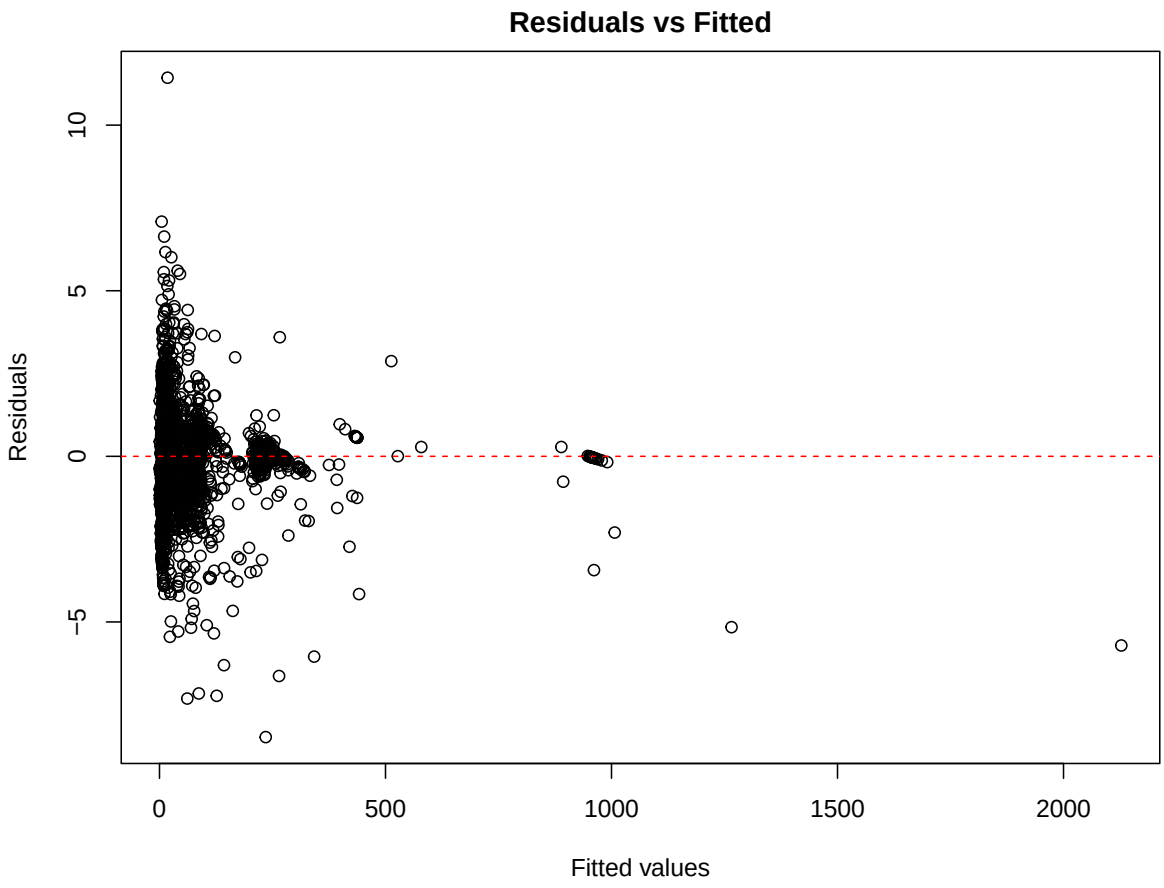

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4

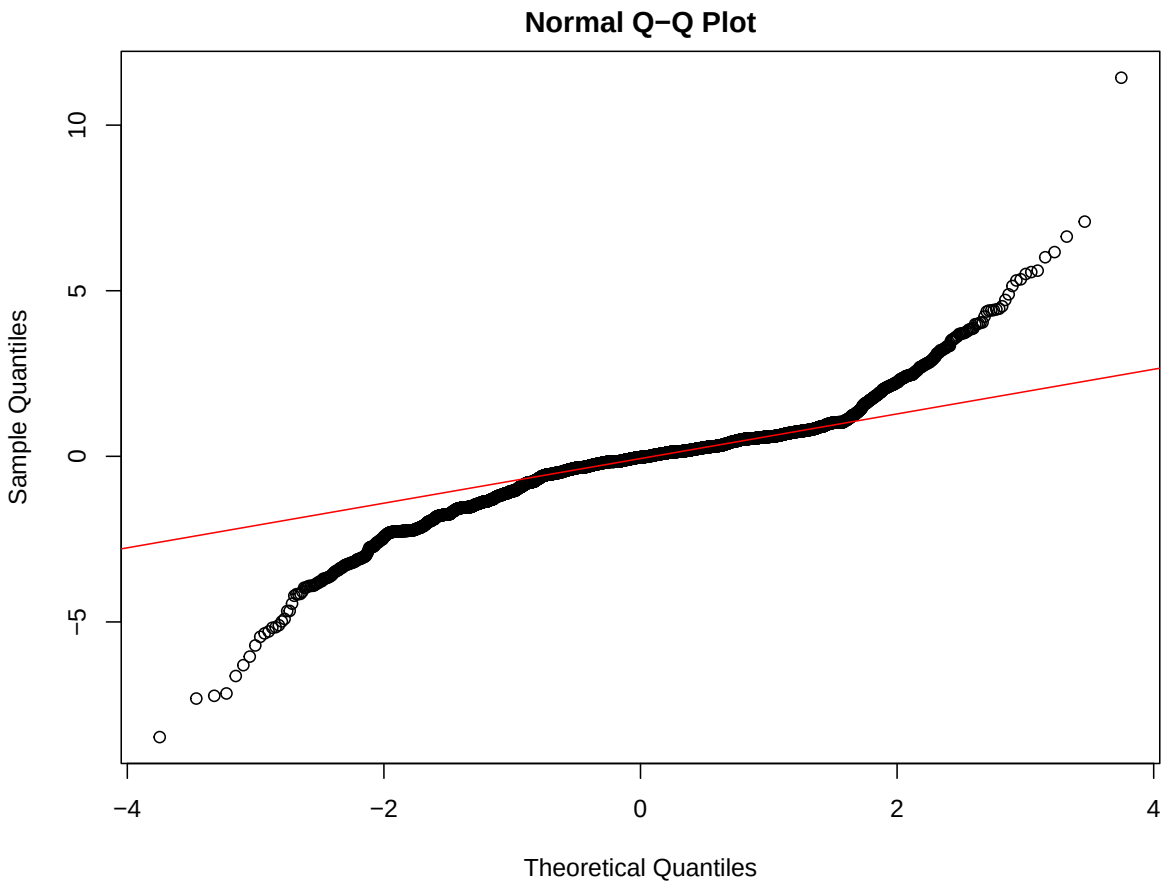

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
500
1000
1500
2000
Fitted values

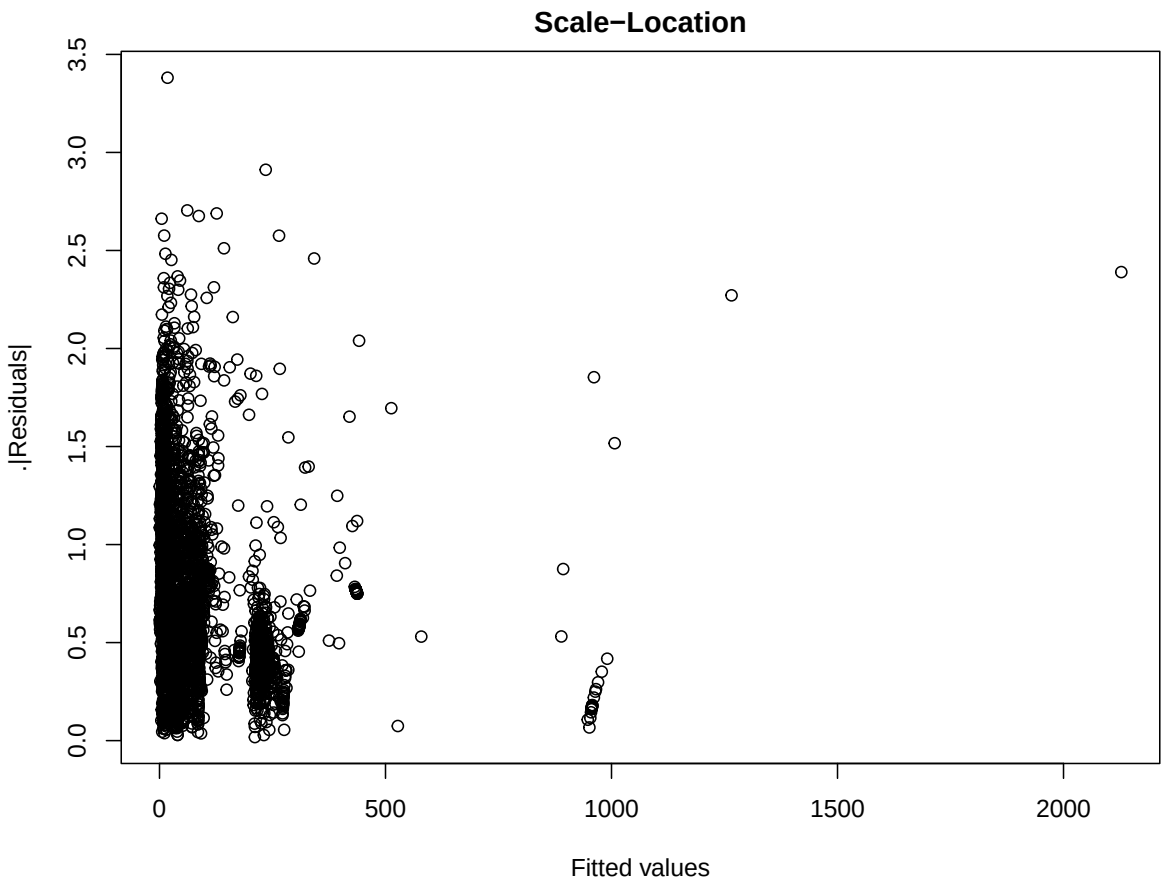


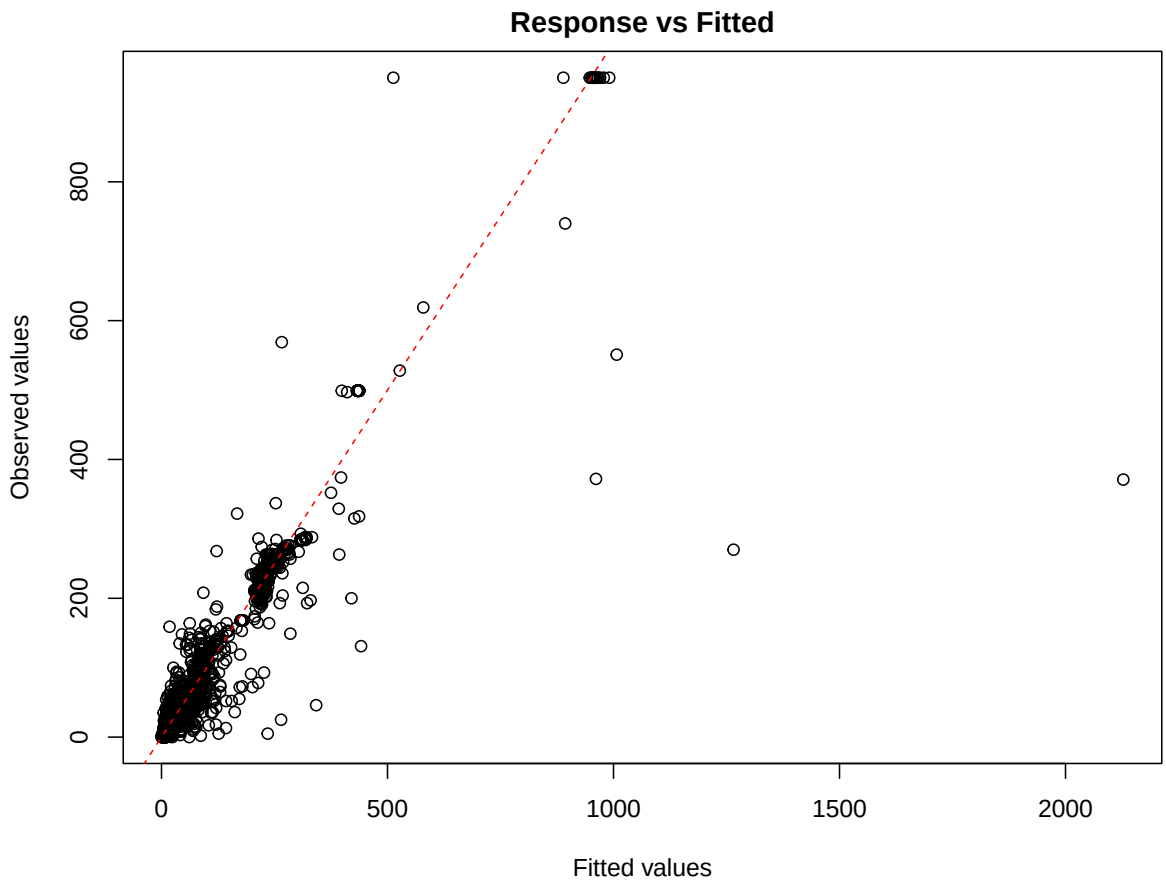

```
### Model 15 : 6. Global model

Model data dimensions after removing NAs: 5601 x 6
Variables in model: total_butterflies, abundance_t_minus_1, mean_wind_speed, mean_temp, sunli
View IDs with observations: 9
Min/Max observations per view: 13 / 2941
GAM fitted successfully
Model Summary:

Family: Negative Binomial(18.021)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + ti(mean_wind_speed, sunlight_proportion) + ti(mean_temp,
```

```
    sunlight_proportion) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3469     0.1651   20.27   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                          edf Ref.df   Chi.sq p-value
s(abundance_t_minus_1)                  8.962  8.999 55245.82  <2e-16 ***
s(mean_wind_speed)                      5.037  6.132    36.26  <2e-16 ***
s(mean_temp)                            8.136  8.656    97.65  <2e-16 ***
s(sunlight_proportion)                  7.928  8.679   100.95  <2e-16 ***
ti(mean_wind_speed,mean_temp)           6.658  8.038    41.42  <2e-16 ***
ti(mean_wind_speed,sunlight_proportion) 7.225  9.186    20.51  0.0194 *
ti(mean_temp,sunlight_proportion)      10.926 12.838   123.20  <2e-16 ***
s(view_id)                              7.775  8.000   123.64  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.874   Deviance explained = 93.6%
-REML =  21487  Scale est. = 1         n = 5601


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
10
5
0
−5
0
500
1000
1500
2000


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4


Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
500
1000
1500
2000
Fitted values




```r
# Model comparison table
comparison_30m <- data.frame(
  Model = names(aic_values_30m),
  AIC = round(aic_values_30m, 2)
) %>%
  arrange(AIC)

cat("\n### Model Comparison (30-minute lag)\n")
```

```
### Model Comparison (30-minute lag)
```

```r
kable(comparison_30m)
```

	Model	AIC
6. Global model	6. Global model	42783.35
5b. Wind x Temp + Temp x Sun	5b. Wind x Temp + Temp x Sun	42795.36
5c. Wind x Sun + Temp x Sun	5c. Wind x Sun + Temp x Sun	42820.34
4c. Temp x Sunlight interaction	4c. Temp x Sunlight interaction	42835.33
5a. Wind x Temp + Wind x Sun	5a. Wind x Temp + Wind x Sun	42895.82
4a. Wind x Temp interaction	4a. Wind x Temp interaction	42907.80
4b. Wind x Sunlight interaction	4b. Wind x Sunlight interaction	42919.99
3d. All additive	3d. All additive	42932.22
3c. Temp + Sunlight	3c. Temp + Sunlight	42955.46
3b. Wind + Sunlight	3b. Wind + Sunlight	43058.30
2c. Sunlight only	2c. Sunlight only	43076.43
3a. Wind + Temp	3a. Wind + Temp	43126.63
2b. Temperature only	2b. Temperature only	43137.36
2a. Wind only	2a. Wind only	43234.48
1. Baseline	1. Baseline	43244.82

1-Hour Lag Analysis

```r
data_60m <- prepare_lag_data(master_df, wind, lag_minutes = 60)
```

```
Preparing data for 60 minute lag...
Valid pairs after filtering: 6895
Calculating interval predictors efficiently...
Processing 6895 unique intervals...
Progress: 10 %
Progress: 20 %
Progress: 30 %
Progress: 40 %
Progress: 50 %
Progress: 60 %
Progress: 69.9 %
Progress: 79.9 %
Progress: 89.9 %
Progress: 99.9 %
Combining results...
Final dataset rows: 5618
Time delta range: 60 to 60 minutes
```

Exploratory Data Analysis (1 hour)

```r
cat("Data structure:\n")
```

```
Data structure:
```

```r
glimpse(data_60m)
```

```
Rows: 5,618
Columns: 30
$ deployment_id.x           <chr> "SC1", "SC1", "SC1", "SC1", "SC1", "SC1", "S~
$ image_filename            <chr> "SC1_20231117184501.JPG", "SC1_2023111718500~
$ total_butterflies         <dbl> 54, 54, 54, 54, 54, 54, 54, 54, 54, 54, 54, ~
$ butterflies_direct_sun    <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ datetime                  <dttm> 2023-11-17 18:45:01, 2023-11-17 18:50:01, 2~
$ view_id                   <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ wind_meter_name           <chr> "JazzPlay", "JazzPlay", "JazzPlay", "JazzPla~
$ deployment_id.y           <chr> "SC1", "SC1", "SC1", "SC1", "SC1", "SC1", "S~
$ timestamp                 <dbl> 2.023112e+13, 2.023112e+13, 2.023112e+13, 2.~
$ temperature               <dbl> 15, 16, 15, 16, 15, 16, 16, 15, 16, 15, 16, ~
$ confidence                <dbl> 0.4041873, 0.5208792, 0.6782328, 0.5311629, ~
$ extraction_status         <chr> "success", "success", "success", "success", ~
$ datetime_t_minus_1        <dttm> 2023-11-17 17:45:01, 2023-11-17 17:50:01, 2~
$ abundance_t_minus_1       <dbl> 54, 54, 54, 54, 54, 54, 54, 54, 54, 54, 54, ~
$ butterflies_sun_t_minus_1 <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ temperature_t_minus_1     <dbl> 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, ~
$ time_delta_mins           <dbl> 60, 60, 60, 60, 60, 60, 60, 60, 60, 60, 60, ~
$ mean_wind_speed           <dbl> 0.400000, 0.900000, 1.030769, 1.177778, 1.24~
$ max_wind_speed            <dbl> 0.4, 1.9, 1.9, 1.9, 1.9, 2.3, 2.7, 2.9, 2.9,~
$ sd_wind_speed             <dbl> 0.0000000, 0.6524678, 0.5297556, 0.5241863, ~
$ gust_factor               <dbl> 2.250000, 1.722222, 1.656716, 1.622642, 1.58~
$ minutes_above_2mps        <dbl> 0, 0, 0, 0, 0, 1, 3, 6, 7, 12, 16, 20, 25, 3~
$ mean_temp                 <dbl> 15.00000, 15.07692, 15.07692, 15.15385, 15.1~
$ max_temp                  <dbl> 15, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16, ~
$ min_temp                  <dbl> 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, ~
$ sd_temp                   <dbl> 0.0000000, 0.2773501, 0.2773501, 0.3755338, ~
$ sunlight_proportion       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ max_prop_sunlight         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ min_prop_sunlight         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ sd_prop_sunlight          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
```

```r
cat("\n\nData summary:\n")
```

```
Data summary:
```

```r
summary(data_60m)
```

```
 deployment_id.x    image_filename     total_butterflies butterflies_direct_sun
 Length:5618        Length:5618        Min.   :  0.00    Min.   :  0.000
 Class :character   Class :character   1st Qu.: 14.00    1st Qu.:  0.000
 Mode  :character   Mode  :character   Median : 35.00    Median :  0.000
                                       Mean   : 66.13    Mean   :  1.688
                                       3rd Qu.: 72.00    3rd Qu.:  0.000
                                       Max.   :950.00    Max.   :150.000

    datetime                          view_id     wind_meter_name
 Min.   :2023-11-17 18:45:01.00   4      :2931   Length:5618
 1st Qu.:2023-12-11 14:07:31.00   2      :1055   Class :character
 Median :2023-12-29 19:05:01.00   1      : 648   Mode  :character
 Mean   :2023-12-29 09:10:30.23   5      : 449
 3rd Qu.:2024-01-23 13:12:31.00   11     : 287
 Max.   :2024-02-04 10:50:01.00   9      : 175
                                  (Other):  73
 deployment_id.y      timestamp          temperature      confidence
 Length:5618        Min.   :2.023e+13   Min.   : 4.00   Min.   :0.1222
 Class :character   1st Qu.:2.023e+13   1st Qu.:11.00   1st Qu.:0.4982
 Mode  :character   Median :2.023e+13   Median :13.00   Median :0.5949
                    Mean   :2.024e+13   Mean   :13.22   Mean   :0.5971
                    3rd Qu.:2.024e+13   3rd Qu.:15.00   3rd Qu.:0.6962
                    Max.   :2.024e+13   Max.   :33.00   Max.   :1.0000

 extraction_status  datetime_t_minus_1               abundance_t_minus_1
 Length:5618        Min.   :2023-11-17 17:45:01.00   Min.   :  0.00
 Class :character   1st Qu.:2023-12-11 13:07:31.00   1st Qu.: 14.00
 Mode  :character   Median :2023-12-29 18:05:01.00   Median : 35.00
                    Mean   :2023-12-29 08:10:30.23   Mean   : 66.44
                    3rd Qu.:2024-01-23 12:12:31.00   3rd Qu.: 73.00
                    Max.   :2024-02-04 09:50:01.00   Max.   :950.00

 butterflies_sun_t_minus_1 temperature_t_minus_1 time_delta_mins
```

```
 Min.   :  0.000           Min.   : 4.00         Min.   :60
 1st Qu.:  0.000           1st Qu.:11.00         1st Qu.:60
 Median :  0.000           Median :13.00         Median :60
 Mean   :  1.712           Mean   :13.23         Mean   :60
 3rd Qu.:  0.000           3rd Qu.:15.00         3rd Qu.:60
 Max.   :150.000           Max.   :33.00         Max.   :60

 mean_wind_speed  max_wind_speed  sd_wind_speed     gust_factor
 Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :1.166
 1st Qu.:0.1871   1st Qu.:0.800   1st Qu.:0.1436   1st Qu.:1.352
 Median :0.7608   Median :1.300   Median :0.2701   Median :1.444
 Mean   :0.8873   Mean   :1.526   Mean   :0.2783   Mean   :  Inf
 3rd Qu.:1.3113   3rd Qu.:2.100   3rd Qu.:0.3822   3rd Qu.:1.586
 Max.   :5.2900   Max.   :6.700   Max.   :1.3876   Max.   :  Inf
                                  NA's   :1        NA's   :687
 minutes_above_2mps   mean_temp        max_temp        min_temp
 Min.   : 0.00      Min.   : 4.00   Min.   : 4.00   Min.   : 4.00
 1st Qu.: 0.00      1st Qu.:11.00   1st Qu.:11.00   1st Qu.:11.00
 Median : 0.00      Median :13.00   Median :13.00   Median :12.00
 Mean   : 5.86      Mean   :13.22   Mean   :13.77   Mean   :12.69
 3rd Qu.: 1.00      3rd Qu.:15.38   3rd Qu.:16.00   3rd Qu.:15.00
 Max.   :60.00      Max.   :32.14   Max.   :33.00   Max.   :30.00

    sd_temp       sunlight_proportion max_prop_sunlight min_prop_sunlight
 Min.   :0.0000   Min.   :0.000000    Min.   :0.00000   Min.   :0.0000
 1st Qu.:0.0000   1st Qu.:0.000000    1st Qu.:0.00000   1st Qu.:0.0000
 Median :0.5345   Median :0.000000    Median :0.00000   Median :0.0000
 Mean   :0.5121   Mean   :0.073082    Mean   :0.12364   Mean   :0.0415
 3rd Qu.:0.5774   3rd Qu.:0.004453    3rd Qu.:0.01923   3rd Qu.:0.0000
 Max.   :5.5076   Max.   :1.000000    Max.   :1.00000   Max.   :1.0000

 sd_prop_sunlight
 Min.   :0.00000
 1st Qu.:0.00000
 Median :0.00000
 Mean   :0.03808
 3rd Qu.:0.00000
 Max.   :0.70711
 NA's   :49
```

```r
# Check for data quality issues
cat("\n\nData quality checks:\n")
```

```
Data quality checks:
```

```r
cat("Total rows:", nrow(data_60m), "\n")
```

```
Total rows: 5618
```

```r
cat("Complete cases:", sum(complete.cases(data_60m)), "\n")
```

```
Complete cases: 4898
```

```r
cat("Rows with abundance_t_minus_1 > 0:", sum(data_60m$abundance_t_minus_1 > 0, na.rm = TRUE
```

```
Rows with abundance_t_minus_1 > 0: 5560
```

```r
cat("Rows with total_butterflies > 0:", sum(data_60m$total_butterflies > 0, na.rm = TRUE), "
```

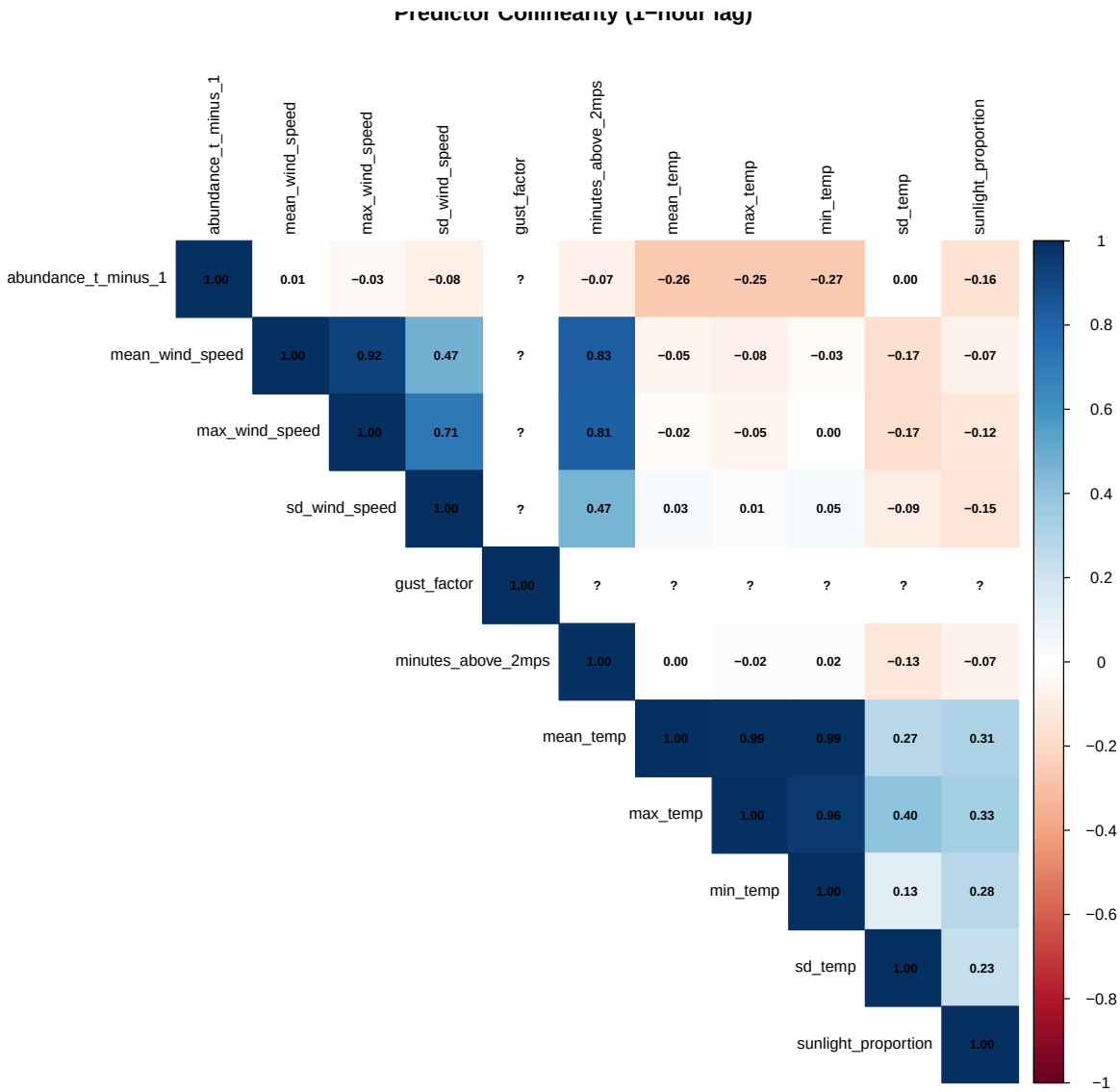
```
Rows with total_butterflies > 0: 5529
```

```r
cat("Unique view_ids:", length(unique(data_60m$view_id)), "\n")
```

```
Unique view_ids: 9
```

```r
# Select numeric predictors for correlation analysis
predictors_60m <- data_60m %>%
  select(abundance_t_minus_1, mean_wind_speed, max_wind_speed, sd_wind_speed,
         gust_factor, minutes_above_2mps, mean_temp, max_temp, min_temp, sd_temp,
         sunlight_proportion) %>%
  na.omit()

if (nrow(predictors_60m) > 1) {
  cor_matrix_60m <- cor(predictors_60m, use = "complete.obs")
  if (!any(is.na(cor_matrix_60m) | is.infinite(cor_matrix_60m))) {
    corrplot(cor_matrix_60m, method = "color", type = "upper",
             order = "hclust", tl.cex = 0.8, tl.col = "black",
             addCoef.col = "black", number.cex = 0.7,
             title = "Predictor Collinearity (1-hour lag)")
```

```r
  } else {
    corrplot(cor_matrix_60m, method = "color", type = "upper",
             order = "original", tl.cex = 0.8, tl.col = "black",
             addCoef.col = "black", number.cex = 0.7,
             title = "Predictor Collinearity (1-hour lag)")
  }
} else {
  cat("Not enough data for correlation analysis\n")
}
```



Model Fitting (1 hour)

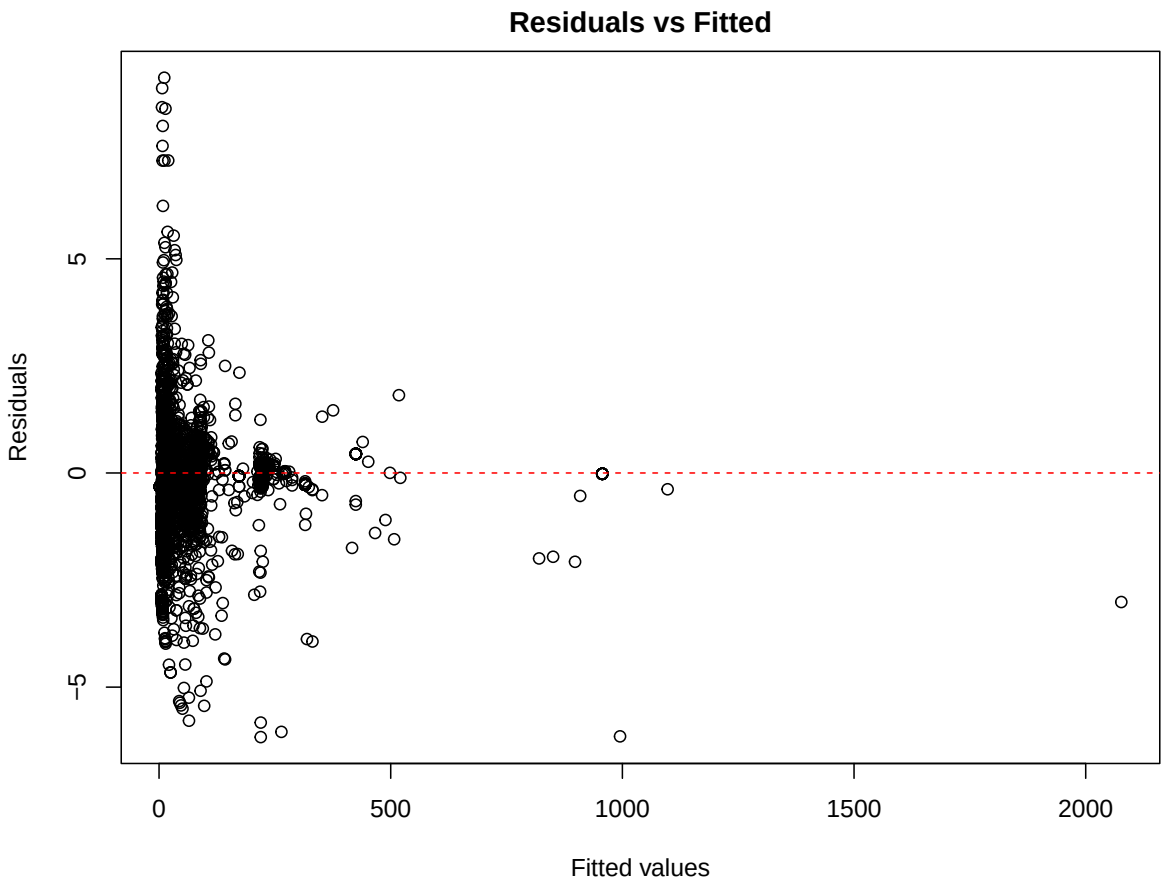
```
# Define comprehensive model formulas
model_formulas_60m <- list(
  # 1. Baseline Model
  "1. Baseline" = total_butterflies ~ s(abundance_t_minus_1) + s(view_id, bs = "re"),
```

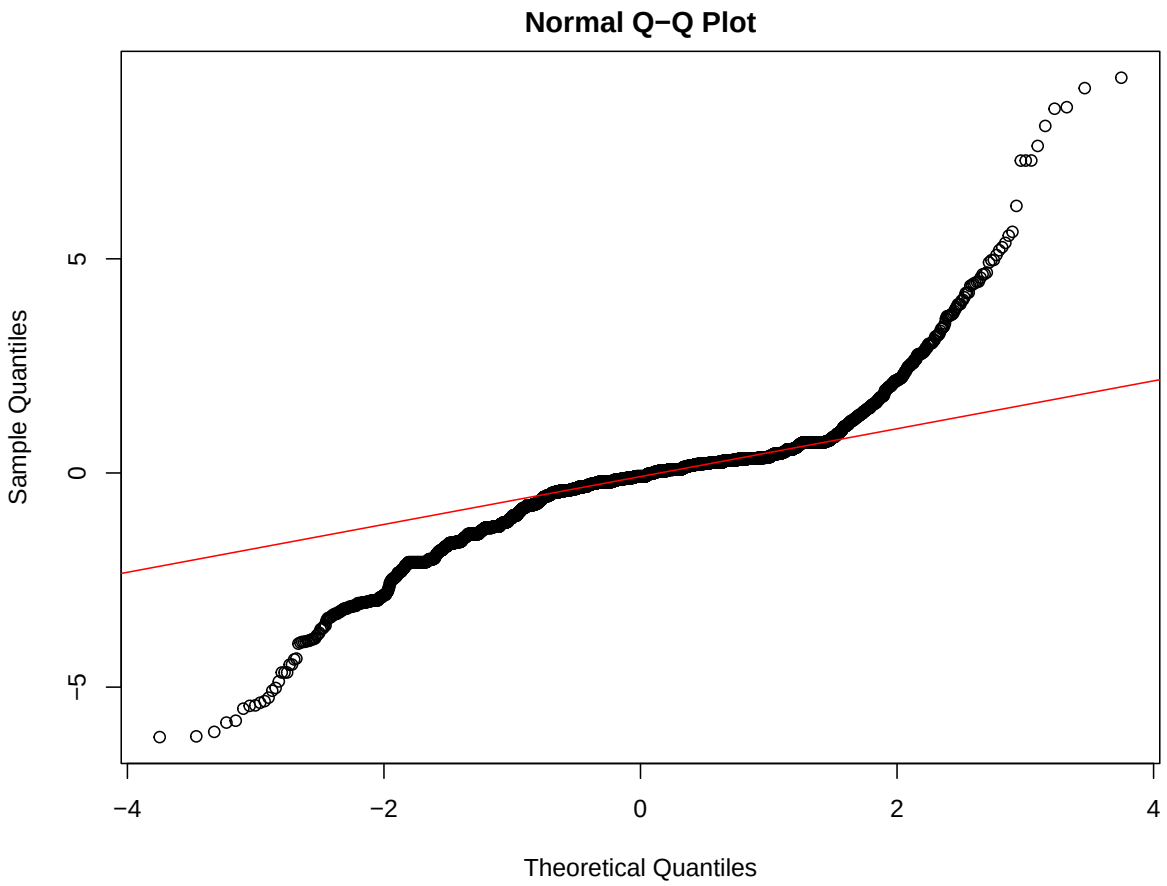
```
  # 2. Single Environmental Factor Models
  "2a. Wind only" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(view_
  "2b. Temperature only" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(vie
  "2c. Sunlight only" = total_butterflies ~ s(abundance_t_minus_1) + s(sunlight_proportion) +

  # 3. Additive Models (No Interactions)
  "3a. Wind + Temp" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(mea
  "3b. Wind + Sunlight" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s
  "3c. Temp + Sunlight" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(sunli
  "3d. All additive" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(me

  # 4. Single Interaction Models
  "4a. Wind x Temp interaction" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_sp
  "4b. Wind x Sunlight interaction" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wi
  "4c. Temp x Sunlight interaction" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wi

  # 5. Multiple Interaction Models
  "5a. Wind x Temp + Wind x Sun" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_s
  "5b. Wind x Temp + Temp x Sun" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_s
  "5c. Wind x Sun + Temp x Sun" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_sp

  # 6. Global Model with All Interactions
  "6. Global model" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(mea
)

# Storage for results
models_60m <- list()
aic_values_60m <- numeric()

# Fit each model
for (i in seq_along(model_formulas_60m)) {
  model_name <- names(model_formulas_60m)[i]
  formula <- model_formulas_60m[[i]]

  cat("\n### Model", i, ":", model_name, "\n\n")

  tryCatch({
    # Check data before fitting
    model_vars <- all.vars(formula)
    model_data <- data_60m %>%
      select(all_of(model_vars), view_id) %>%
      na.omit()
```

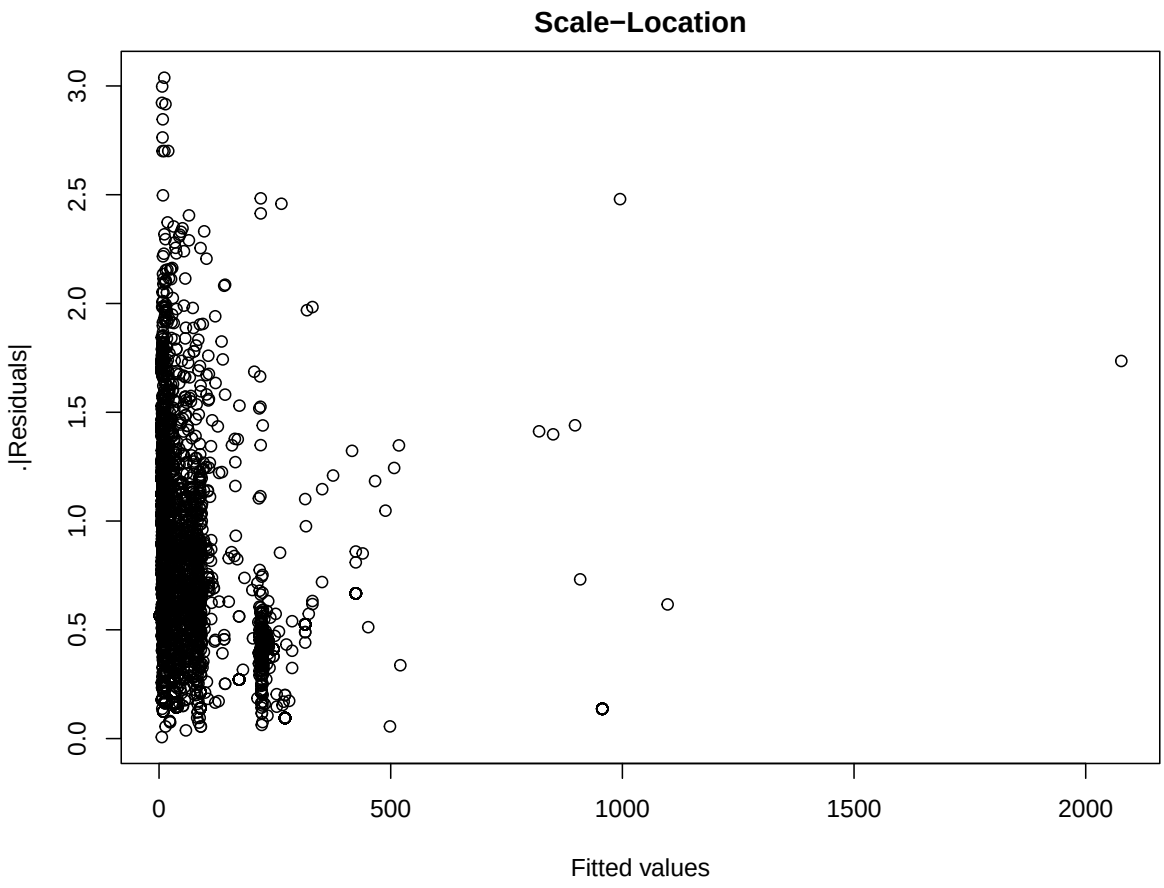
```r
cat("Model data dimensions after removing NAs:", nrow(model_data), "x", ncol(model_data)

# Fit GAM
gam_model <- gam(
  formula,
  family = nb(),
  data = model_data
)

cat("GAM fitted successfully\n")

# Store model
models_60m[[model_name]] <- gam_model
aic_values_60m[model_name] <- AIC(gam_model)

# Print summary
cat("Model Summary:\n")
print(summary(gam_model))

# Diagnostic plots - individual plots for better visibility in PDF
cat("\n\nDiagnostic plots:\n")

# Set up for individual plots
par(mfrow = c(1, 1), mar = c(4, 4, 2, 1))

# Plot 1: Residuals vs Fitted
plot(fitted(gam_model), residuals(gam_model),
     xlab = "Fitted values", ylab = "Residuals",
     main = "Residuals vs Fitted")
abline(h = 0, col = "red", lty = 2)

# Plot 2: Q-Q plot
qqnorm(residuals(gam_model))
qqline(residuals(gam_model), col = "red")

# Plot 3: Scale-Location (if appropriate)
plot(fitted(gam_model), sqrt(abs(residuals(gam_model))),
     xlab = "Fitted values", ylab = "√|Residuals|",
     main = "Scale-Location")

# Plot 4: Response vs Fitted
plot(fitted(gam_model), model_data$total_butterflies,
```

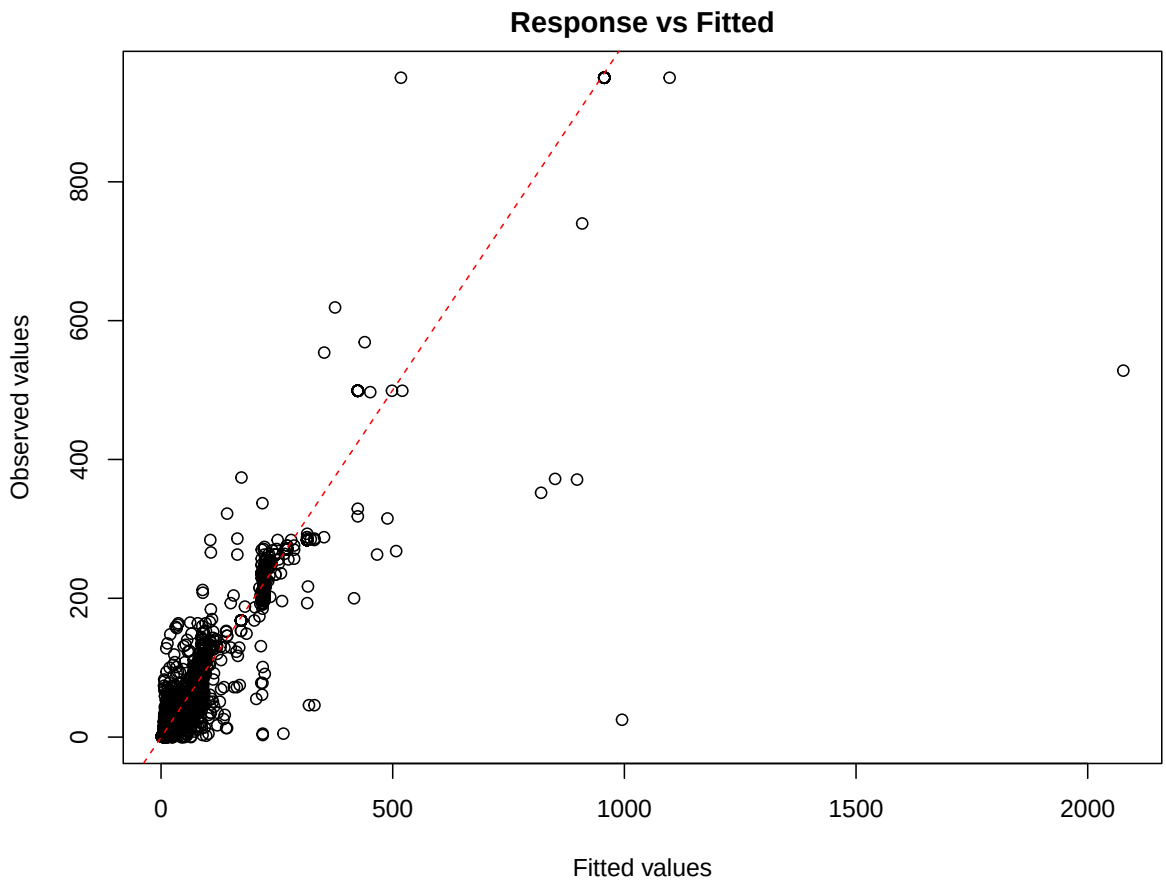
```r
         xlab = "Fitted values", ylab = "Observed values",
         main = "Response vs Fitted")
    abline(a = 0, b = 1, col = "red", lty = 2)

  }, error = function(e) {
    cat("Error fitting model:", e$message, "\n")
    aic_values_60m[model_name] <- NA
  })
}
```

```
### Model 1 : 1. Baseline

Model data dimensions after removing NAs: 5618 x 3
GAM fitted successfully
Model Summary:

Family: Negative Binomial(7.321)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3853     0.1913   17.69   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                         edf Ref.df Chi.sq p-value
s(abundance_t_minus_1) 8.915  8.997  37036  <2e-16 ***
s(view_id)             7.776  8.000    131  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.873   Deviance explained = 87.5%
-REML =  23014  Scale est. = 1         n = 5618


Diagnostic plots:
```


Residuals vs Fitted
Residuals
5
0
−5
0
500
1000
1500
2000
Fitted values


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−5
0
5
−4
−2
0
2
4

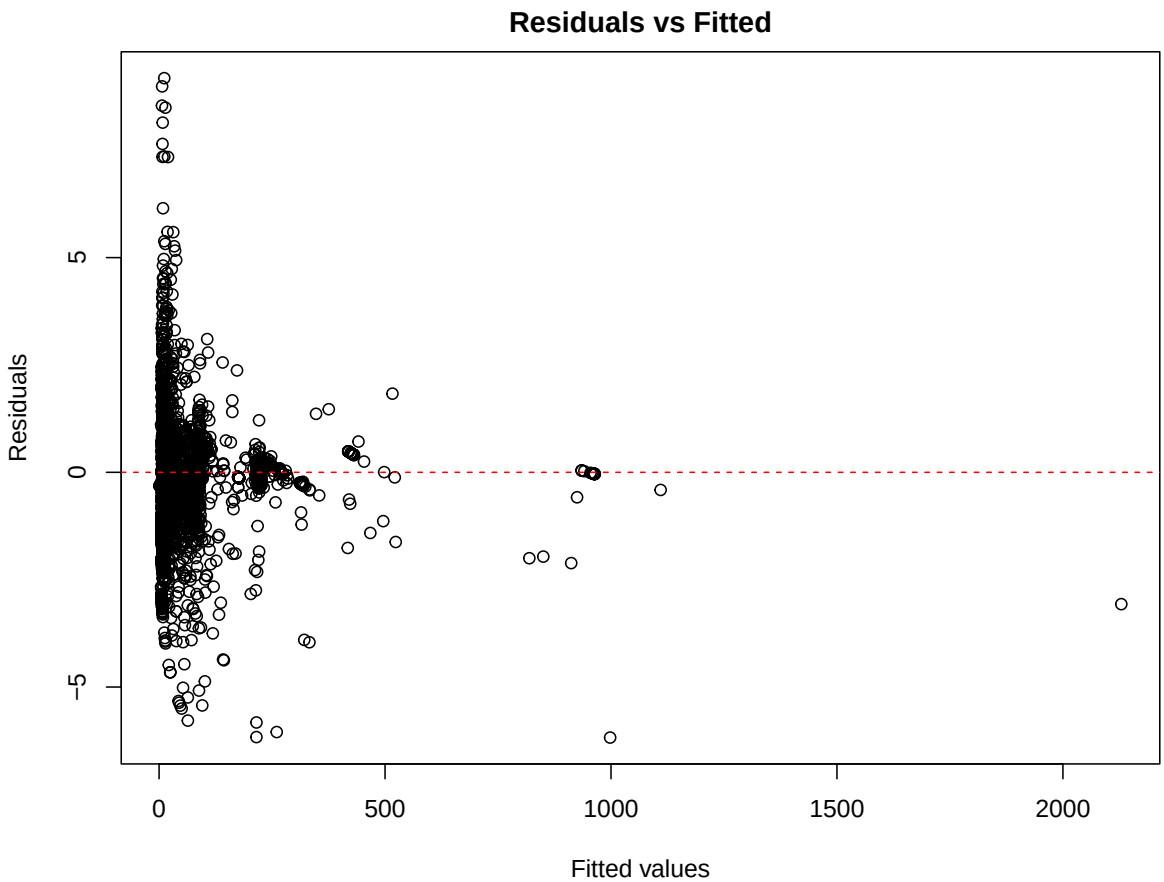

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
500
1000
1500
2000
Fitted values

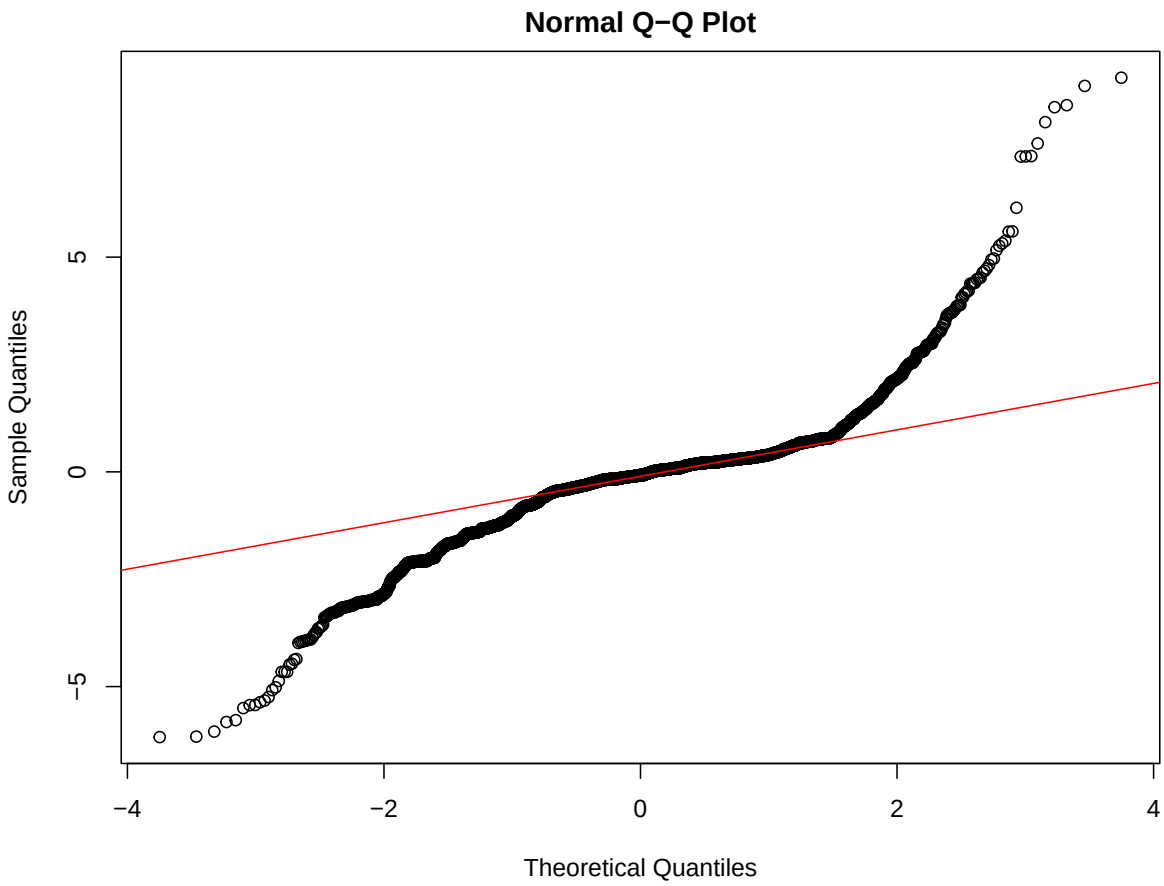


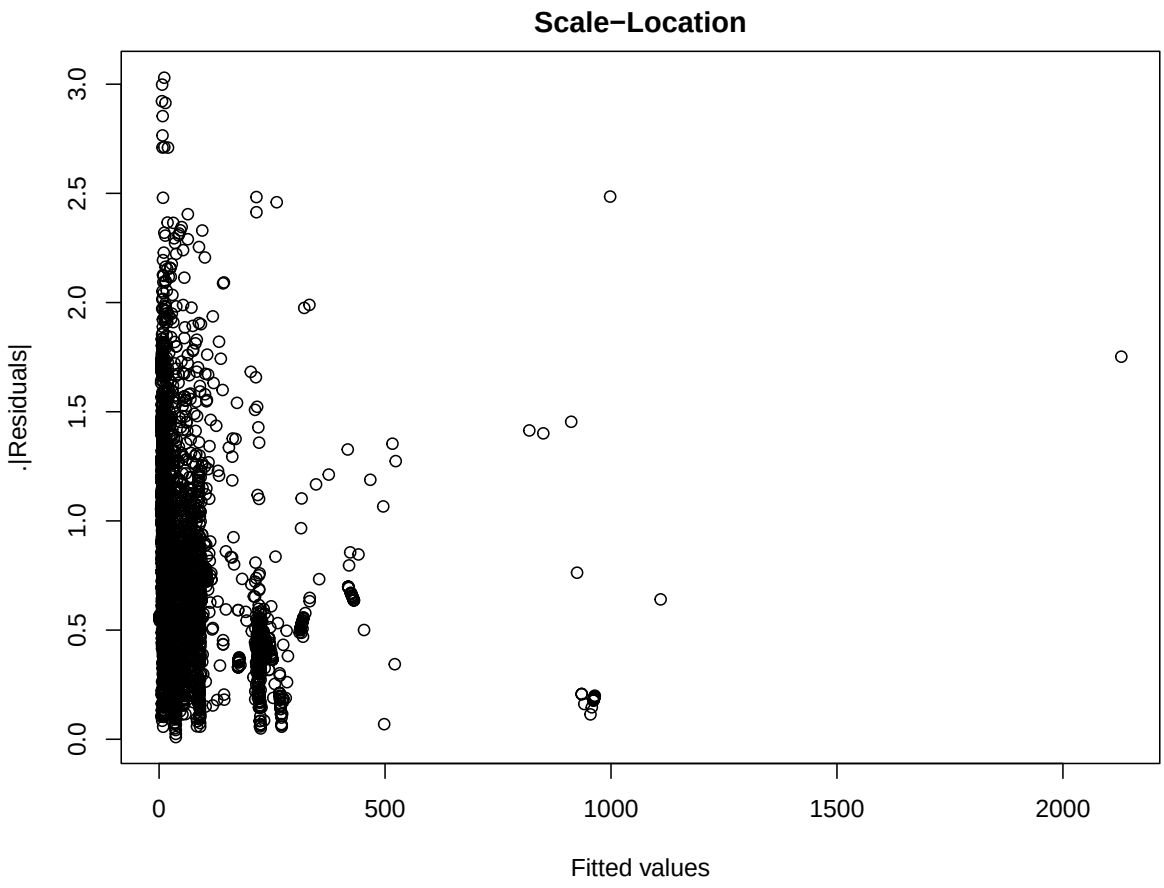

```
### Model 2 : 2a. Wind only

Model data dimensions after removing NAs: 5618 x 4
GAM fitted successfully
Model Summary:

Family: Negative Binomial(7.381)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3888     0.1942   17.45   <2e-16 ***
```

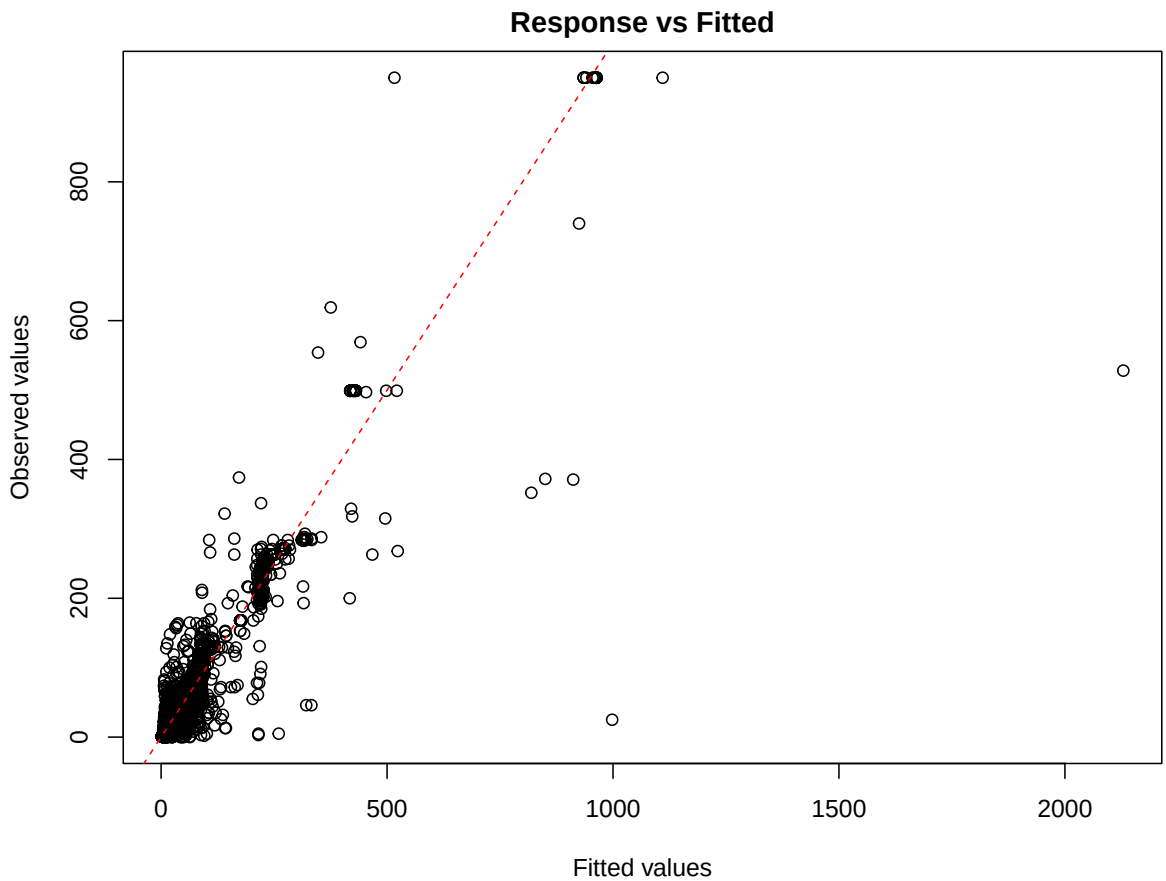
```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df   Chi.sq  p-value
s(abundance_t_minus_1) 8.916  8.997 36183.40  < 2e-16 ***
s(mean_wind_speed)     5.099  6.217    27.23 0.000129 ***
s(view_id)             7.781  8.000   132.44  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.869   Deviance explained = 87.6%
-REML =  23009  Scale est. = 1         n = 5618
```

Diagnostic plots:



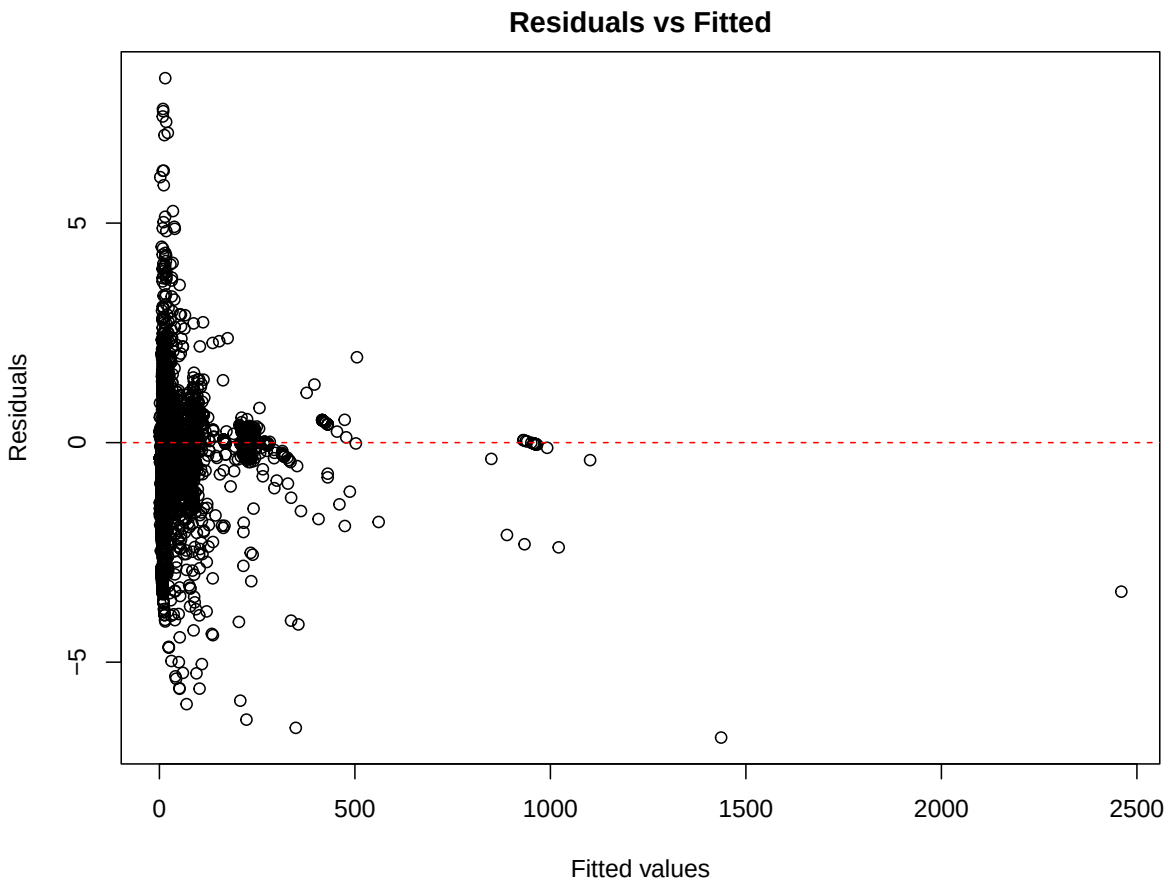

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−5
0
5
−4
−2
0
2
4

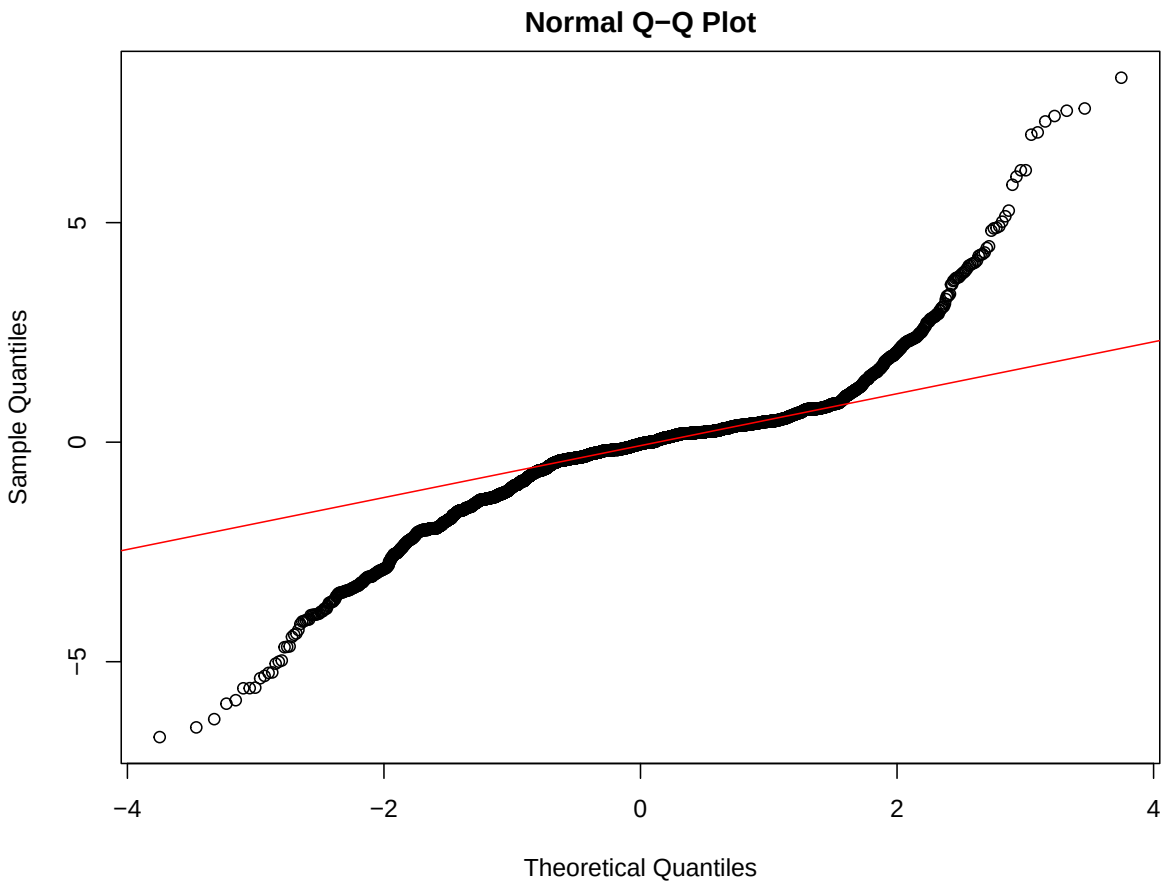

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
500
1000
1500
2000
Fitted values

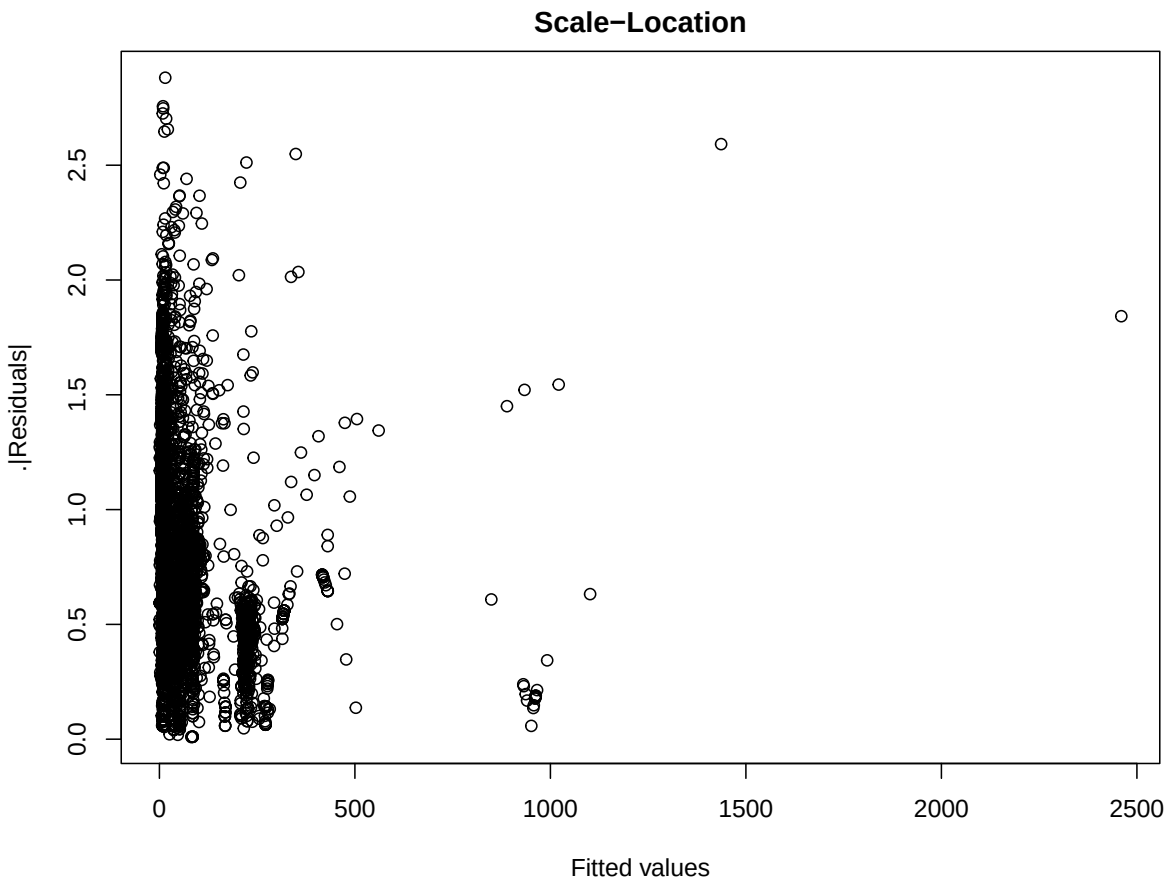


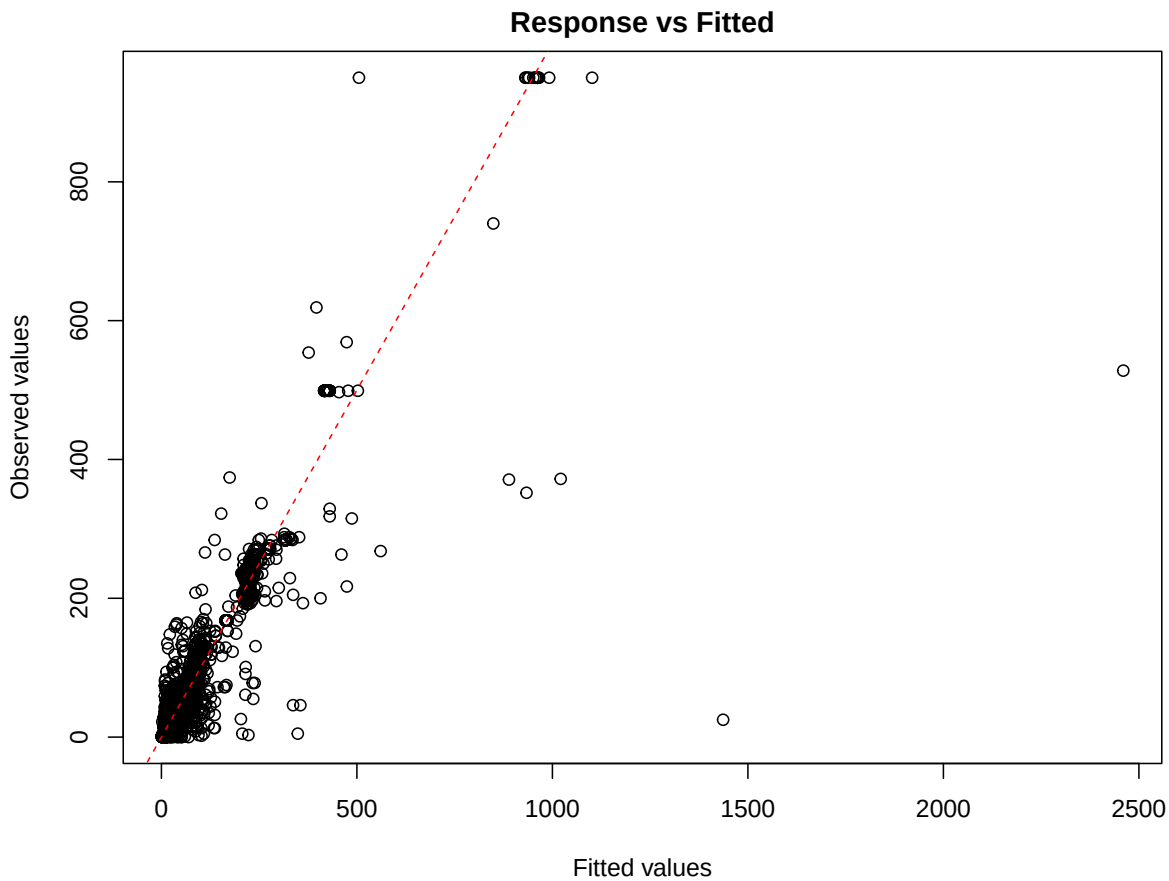

```
### Model 3 : 2b. Temperature only

Model data dimensions after removing NAs: 5618 x 4
GAM fitted successfully
Model Summary:

Family: Negative Binomial(7.69)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(view_id,
    bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    3.365      0.174   19.34   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df  Chi.sq p-value
s(abundance_t_minus_1) 8.925  8.998 36406.3  <2e-16 ***
s(mean_temp)           8.084  8.651   255.9  <2e-16 ***
s(view_id)             7.741  8.000   165.5  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.816   Deviance explained = 88.1%
-REML =  22894  Scale est. = 1         n = 5618


Diagnostic plots:
```



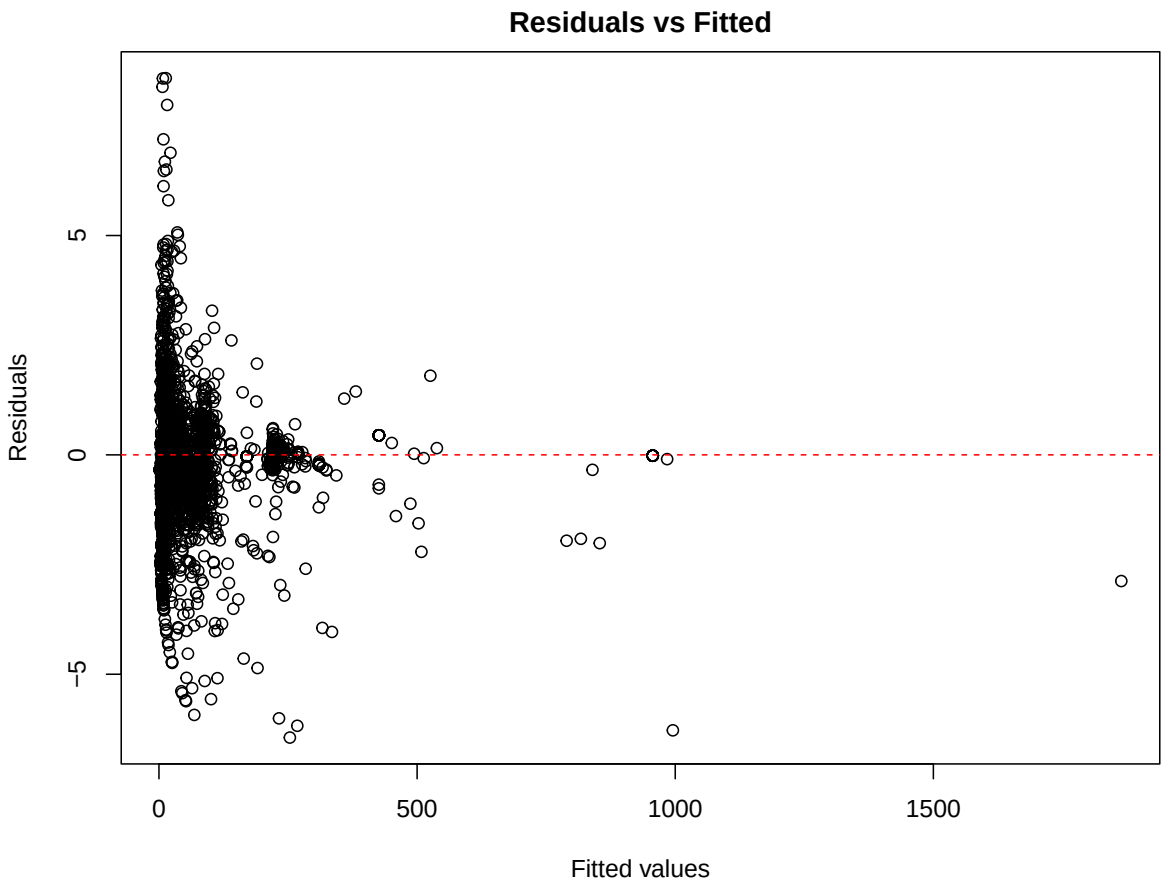

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−5
0
5
−4
−2
0
2
4

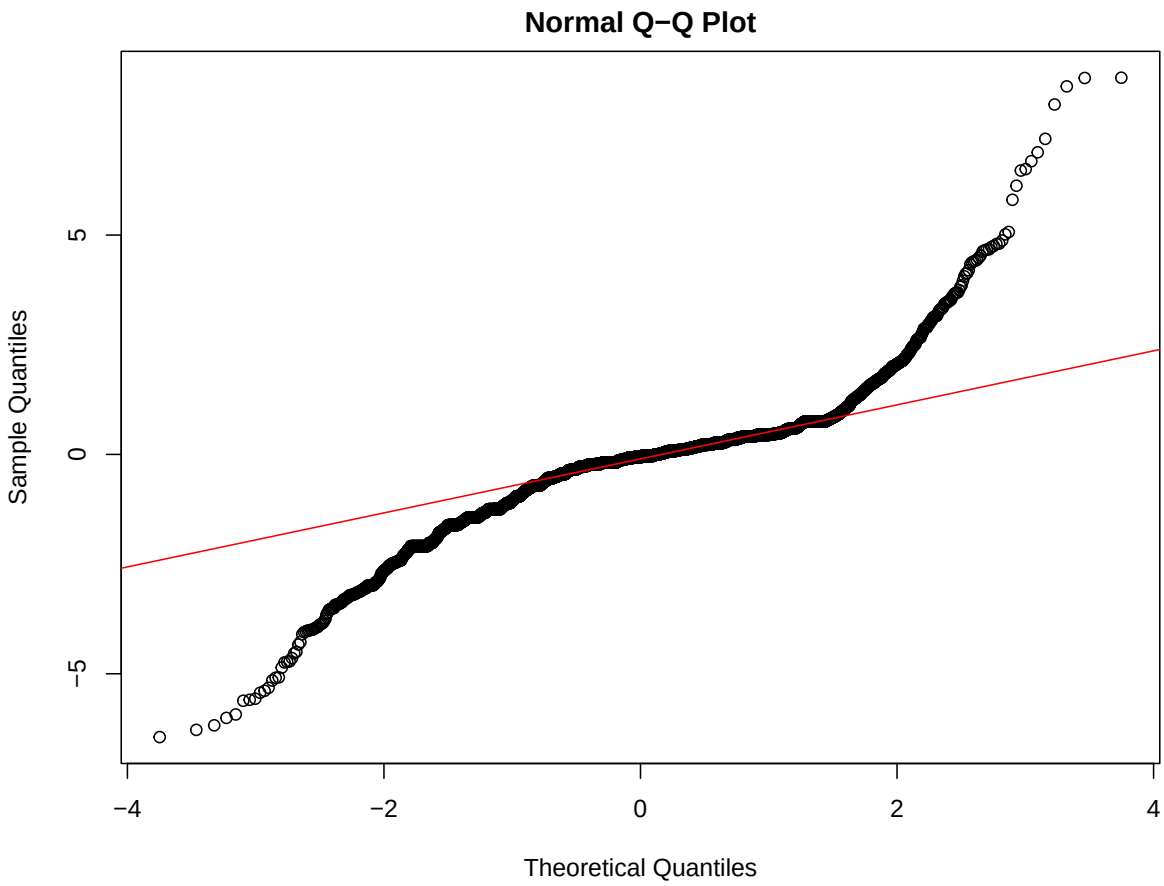

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
2500
Fitted values

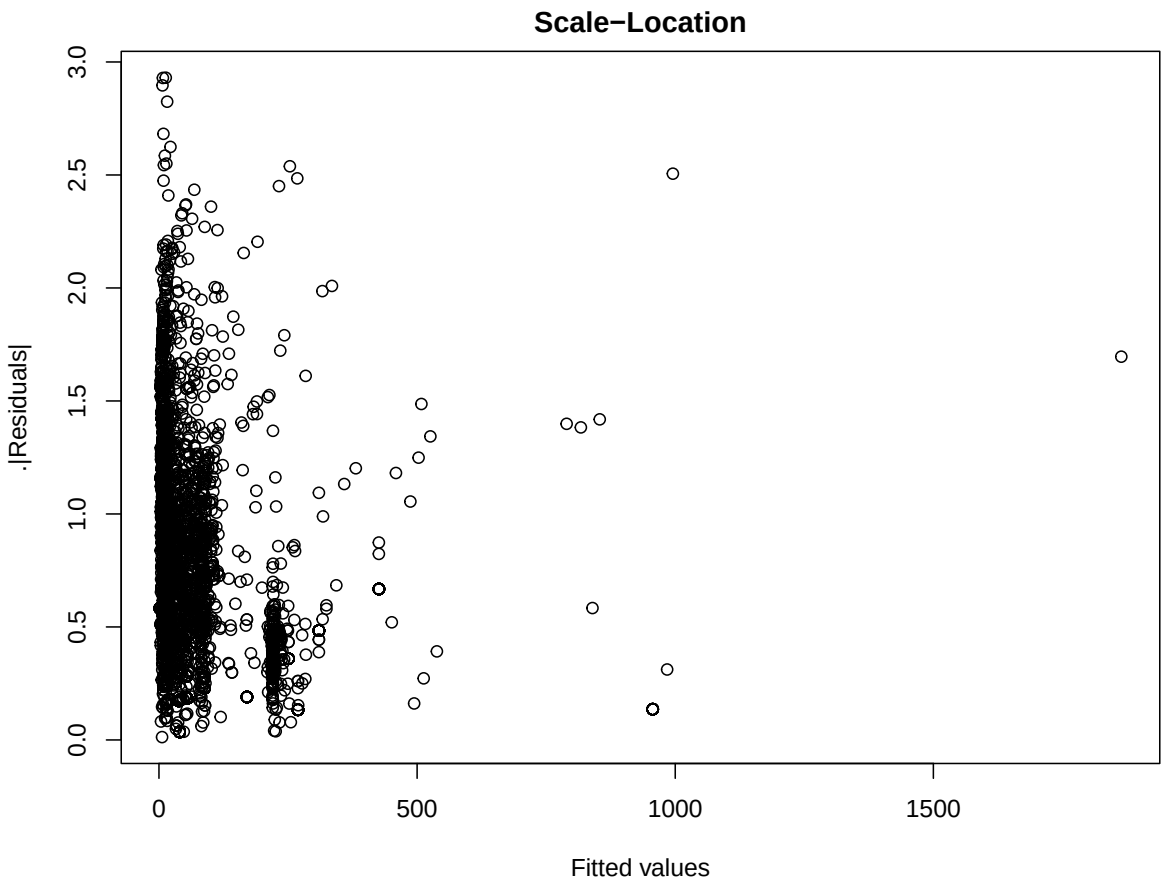


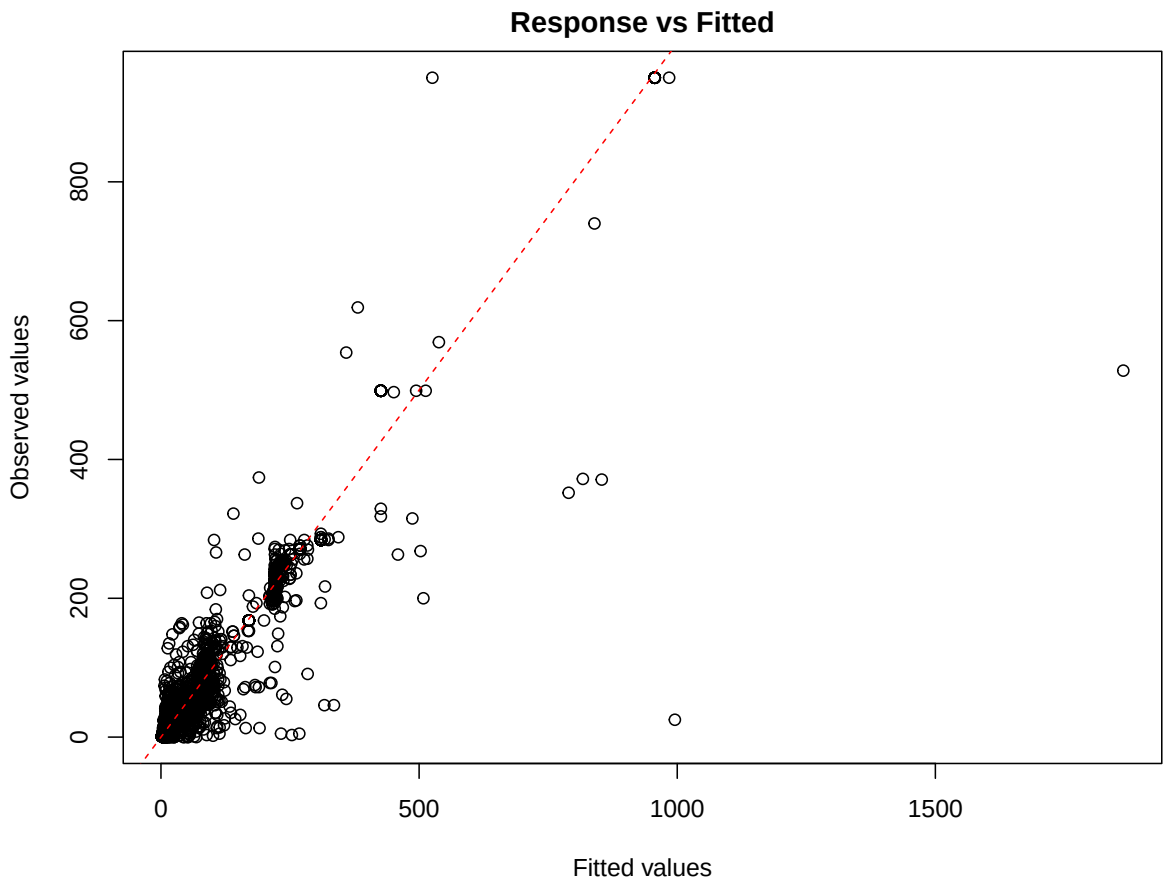

```
### Model 4 : 2c. Sunlight only

Model data dimensions after removing NAs: 5618 x 4
GAM fitted successfully
Model Summary:

Family: Negative Binomial(7.647)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(sunlight_proportion) +
    s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3789     0.1825   18.52   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df  Chi.sq p-value
s(abundance_t_minus_1) 8.912  8.997 35711.1  <2e-16 ***
s(sunlight_proportion) 8.321  8.868   223.8  <2e-16 ***
s(view_id)             7.761  8.000   103.9  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.885   Deviance explained =   88%
-REML =  22920  Scale est. = 1         n = 5618


Diagnostic plots:
```



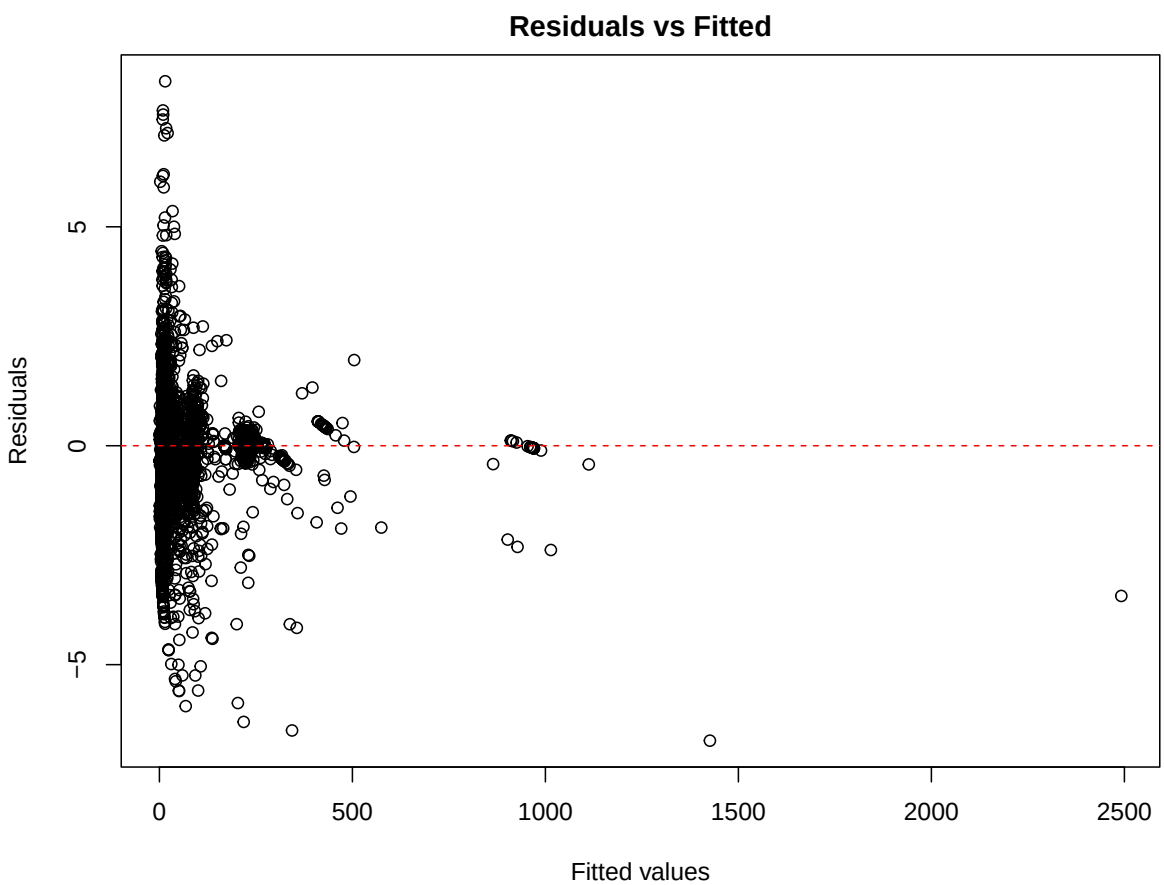

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−5
0
5
−4
−2
0
2
4

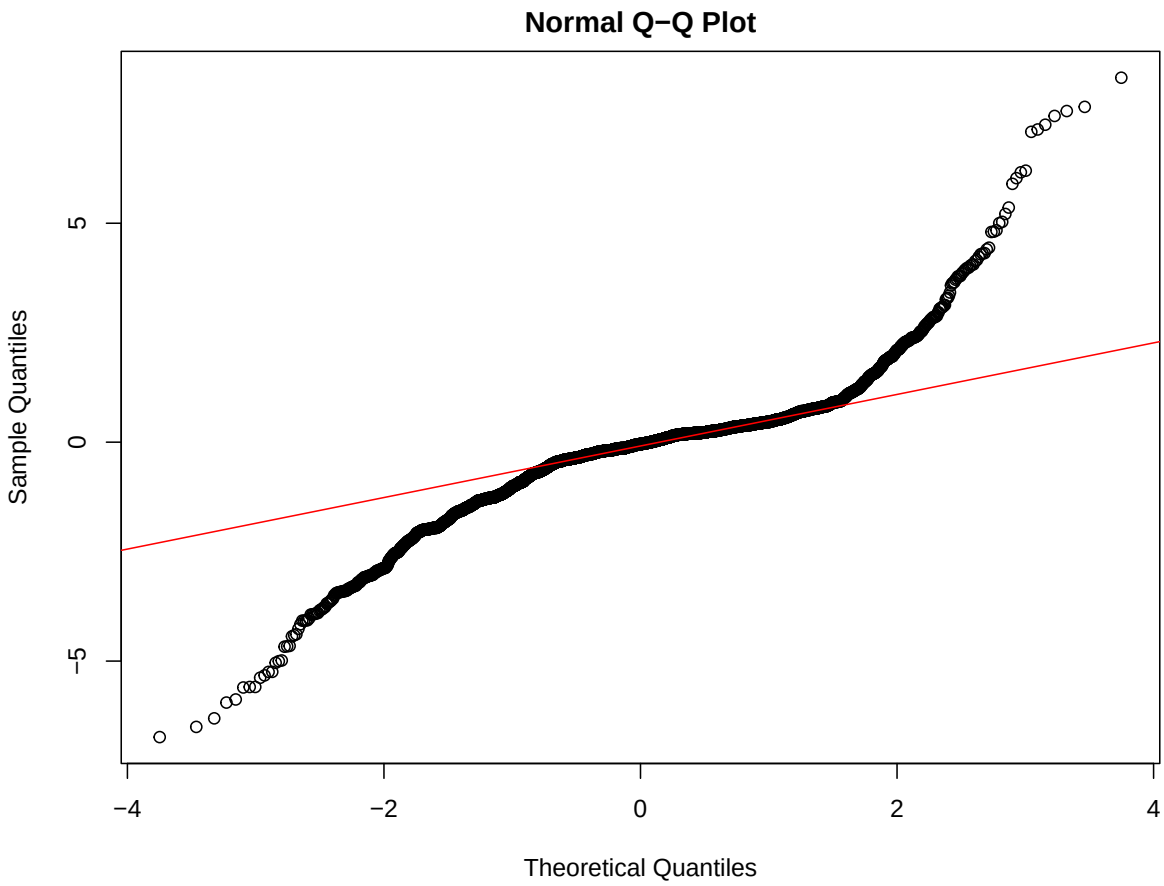

Scale−Location
.|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
500
1000
1500
Fitted values

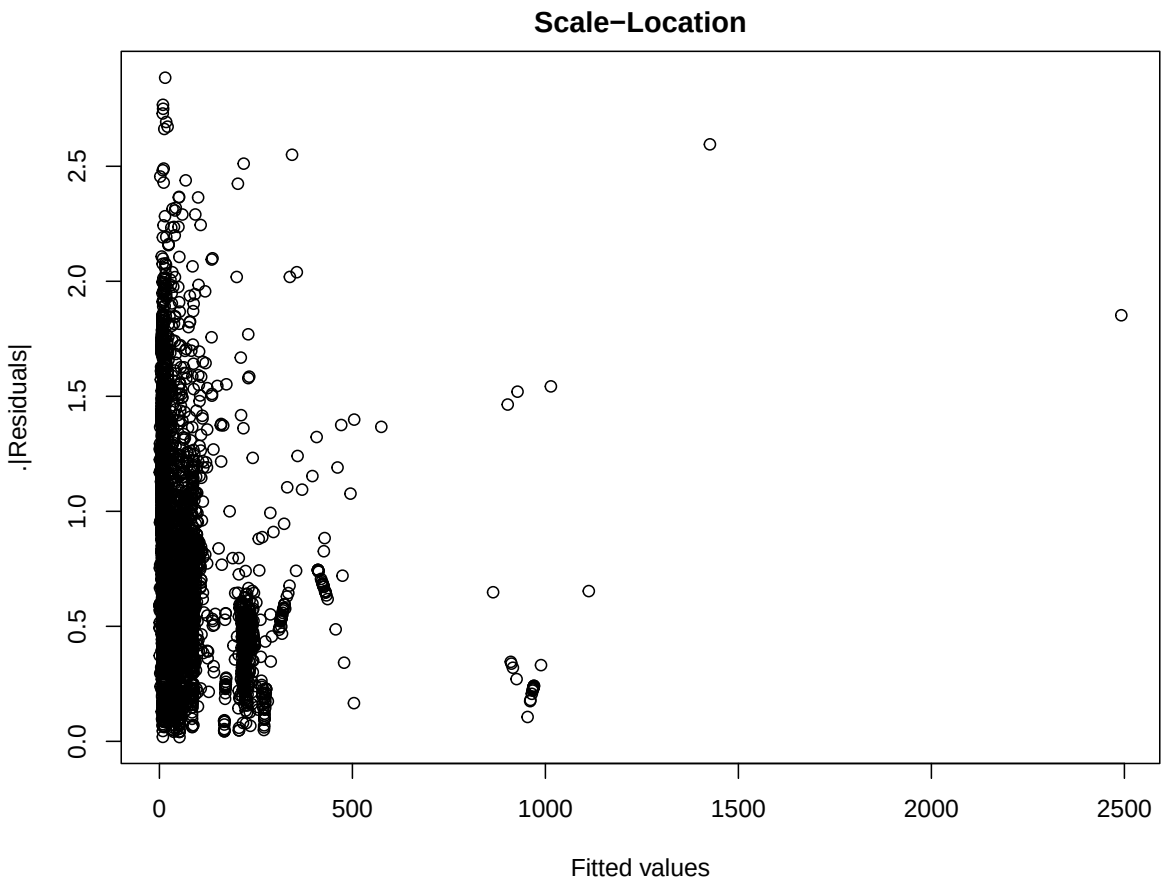


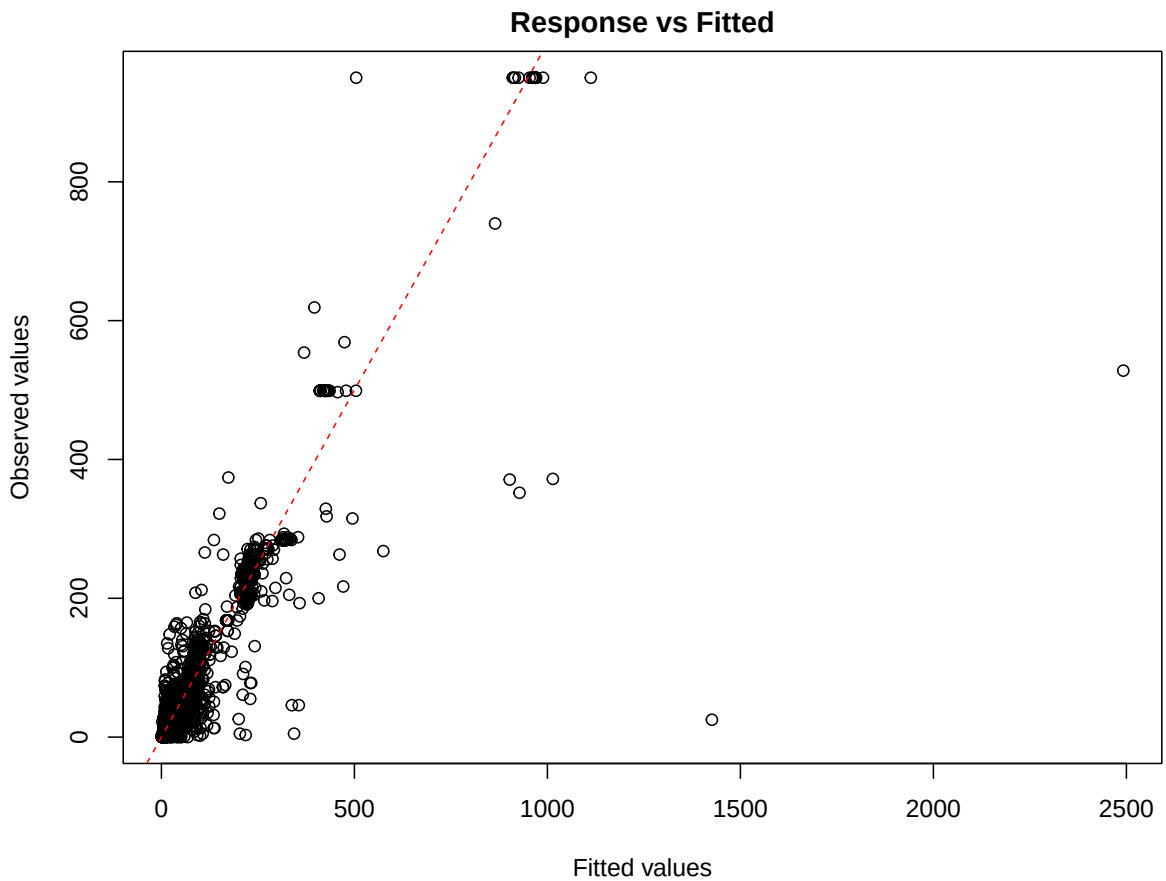

```
### Model 5 : 3a. Wind + Temp

Model data dimensions after removing NAs: 5618 x 5
GAM fitted successfully
Model Summary:

Family: Negative Binomial(7.754)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    3.369      0.177   19.03   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df   Chi.sq p-value
s(abundance_t_minus_1) 8.926  8.998 35560.58 < 2e-16 ***
s(mean_wind_speed)     5.806  6.965    27.06 0.00026 ***
s(mean_temp)           8.091  8.656   252.18 < 2e-16 ***
s(view_id)             7.747  8.000   157.72 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.815   Deviance explained = 88.2%
-REML =  22891  Scale est. = 1         n = 5618


Diagnostic plots:
```



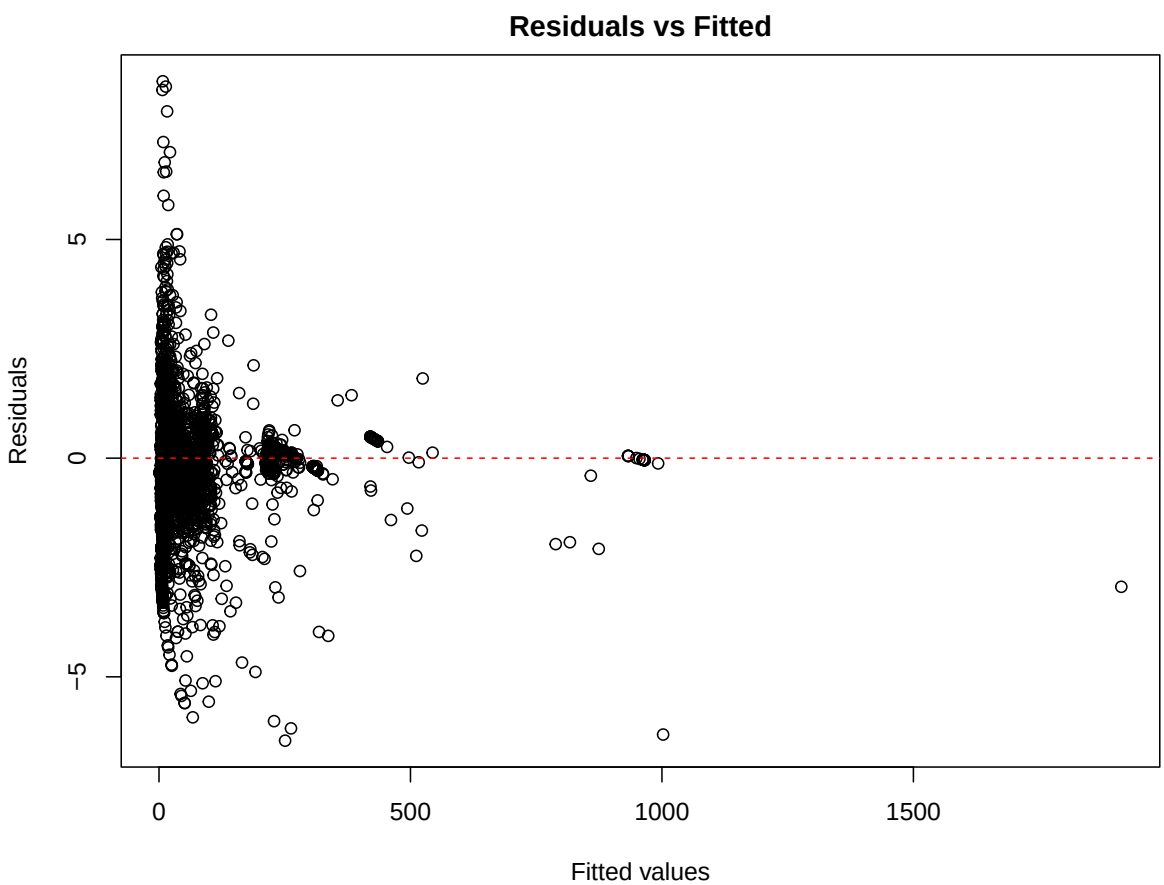

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−5
0
5
−4
−2
0
2
4

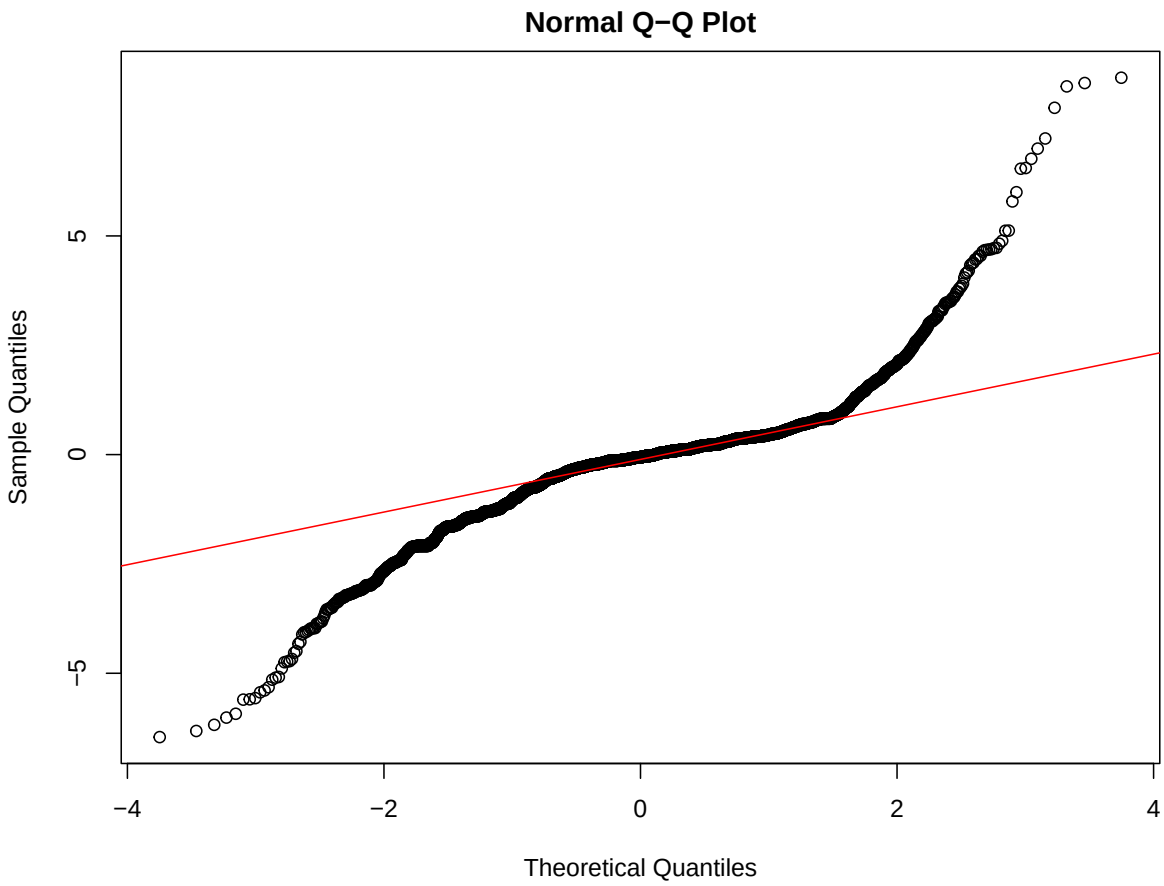

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
2500
Fitted values

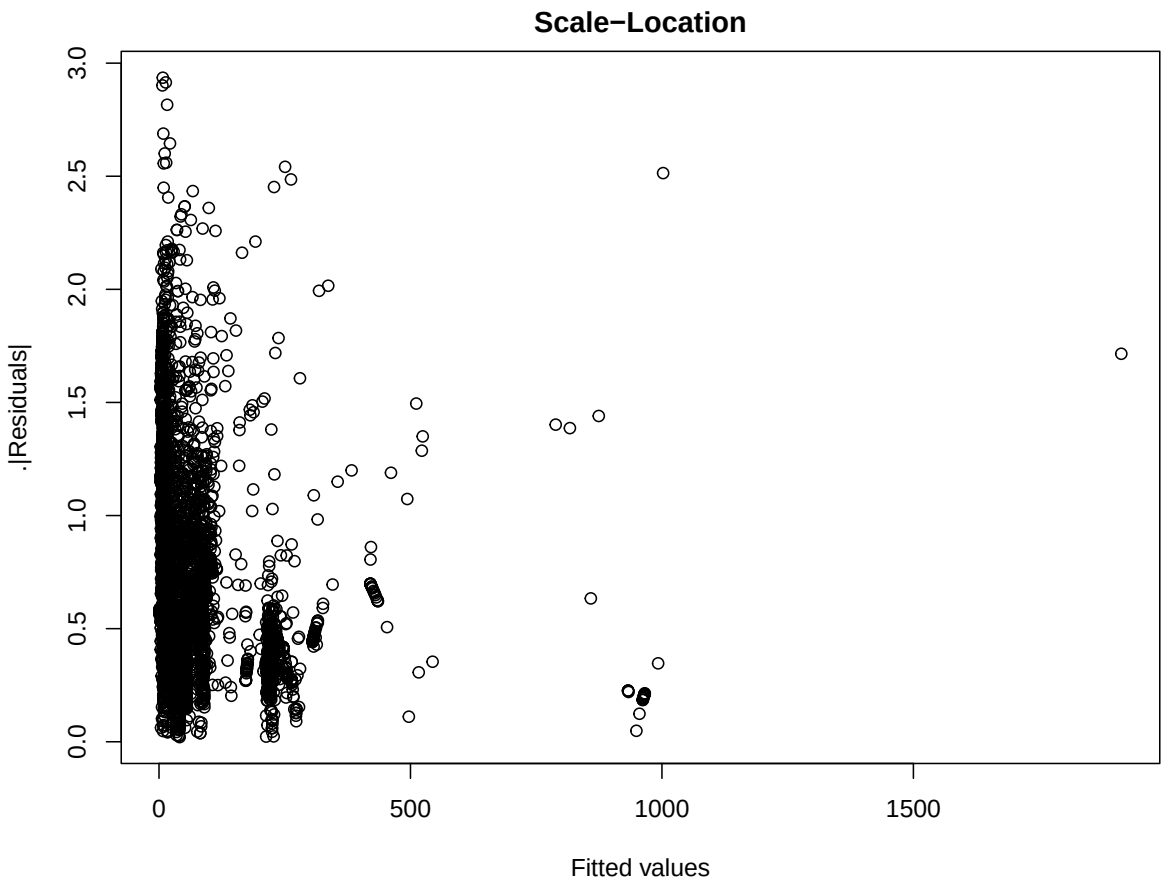


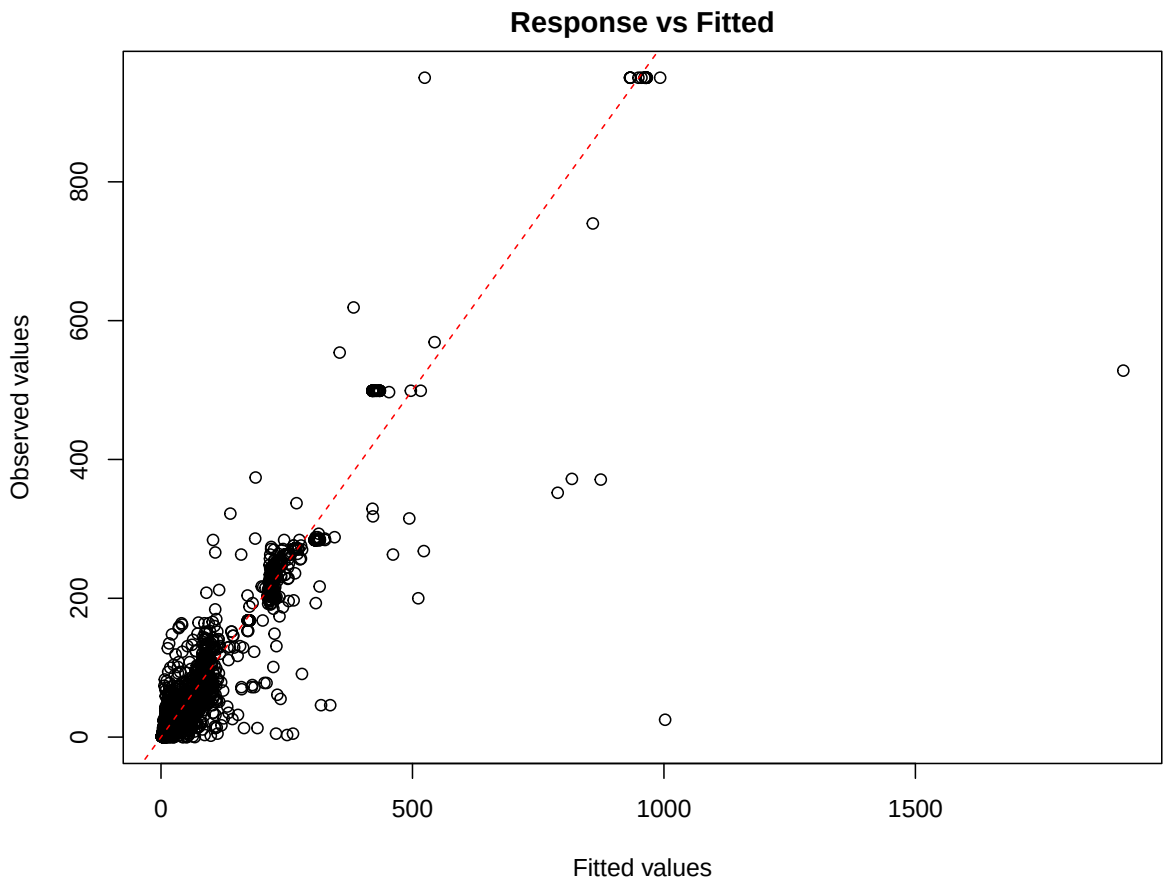

```
### Model 6 : 3b. Wind + Sunlight

Model data dimensions after removing NAs: 5618 x 5
GAM fitted successfully
Model Summary:

Family: Negative Binomial(7.731)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(sunlight_proportion) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3826     0.1853   18.25   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df   Chi.sq p-value
s(abundance_t_minus_1) 8.913  8.997 34742.36  <2e-16 ***
s(mean_wind_speed)     6.334  7.478    42.82  <2e-16 ***
s(sunlight_proportion) 8.304  8.862   230.56  <2e-16 ***
s(view_id)             7.767  8.000   110.29  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.882   Deviance explained = 88.1%
-REML =  22912  Scale est. = 1         n = 5618


Diagnostic plots:
```



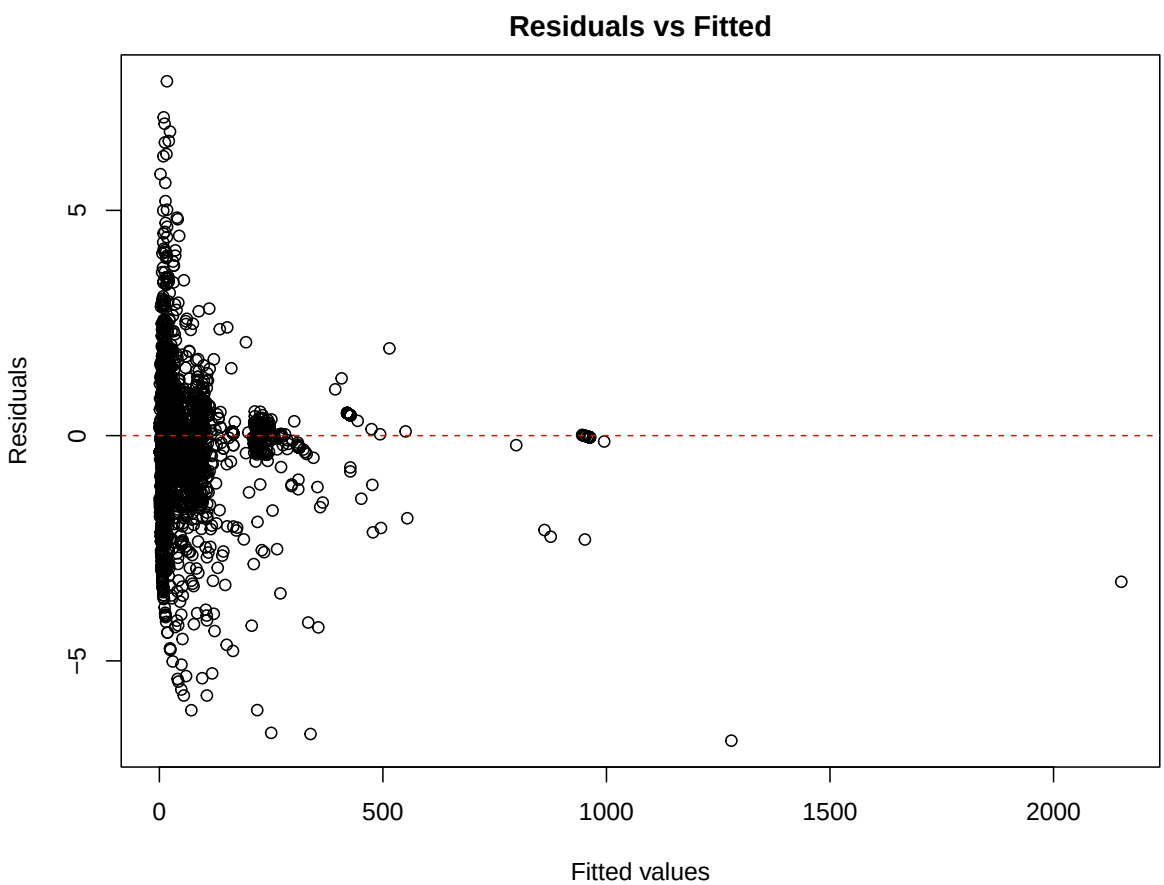

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
5
0
−5
−4
−2
0
2
4

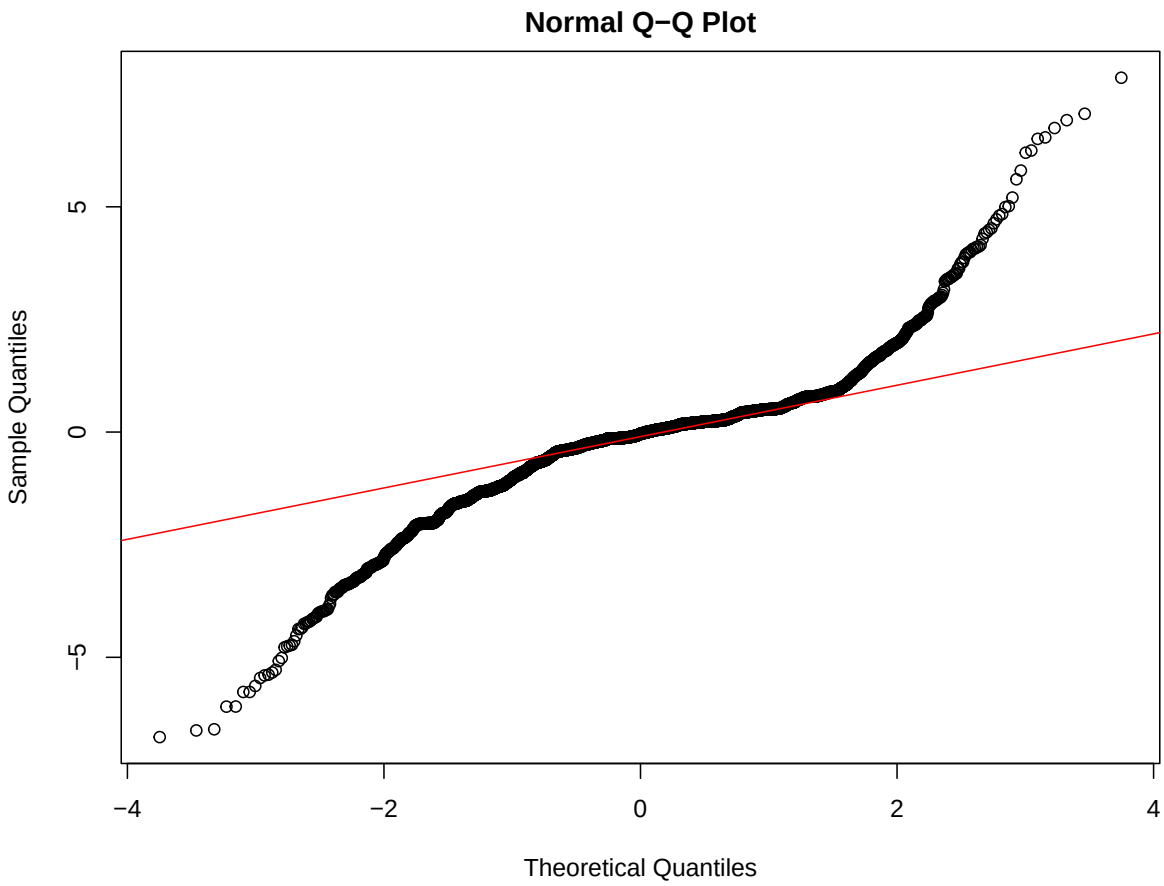

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
500
1000
1500
Fitted values

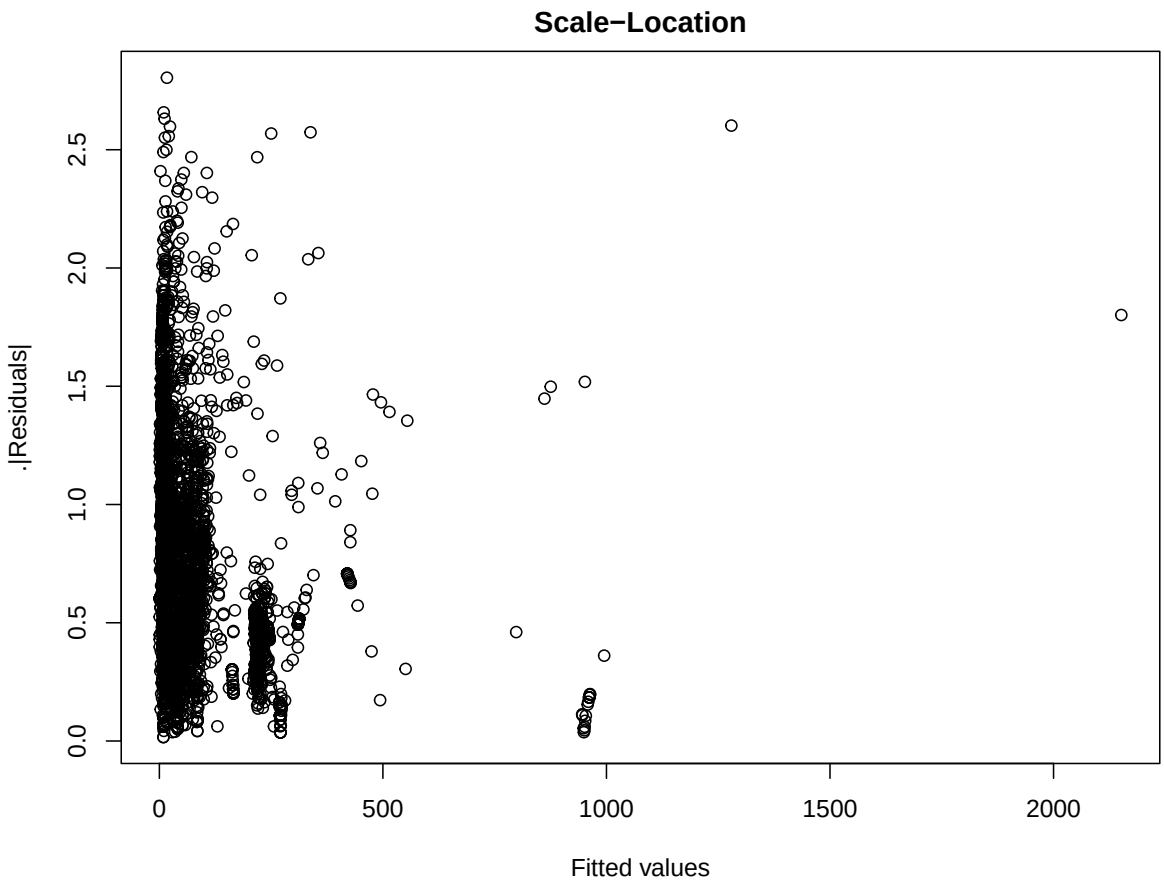


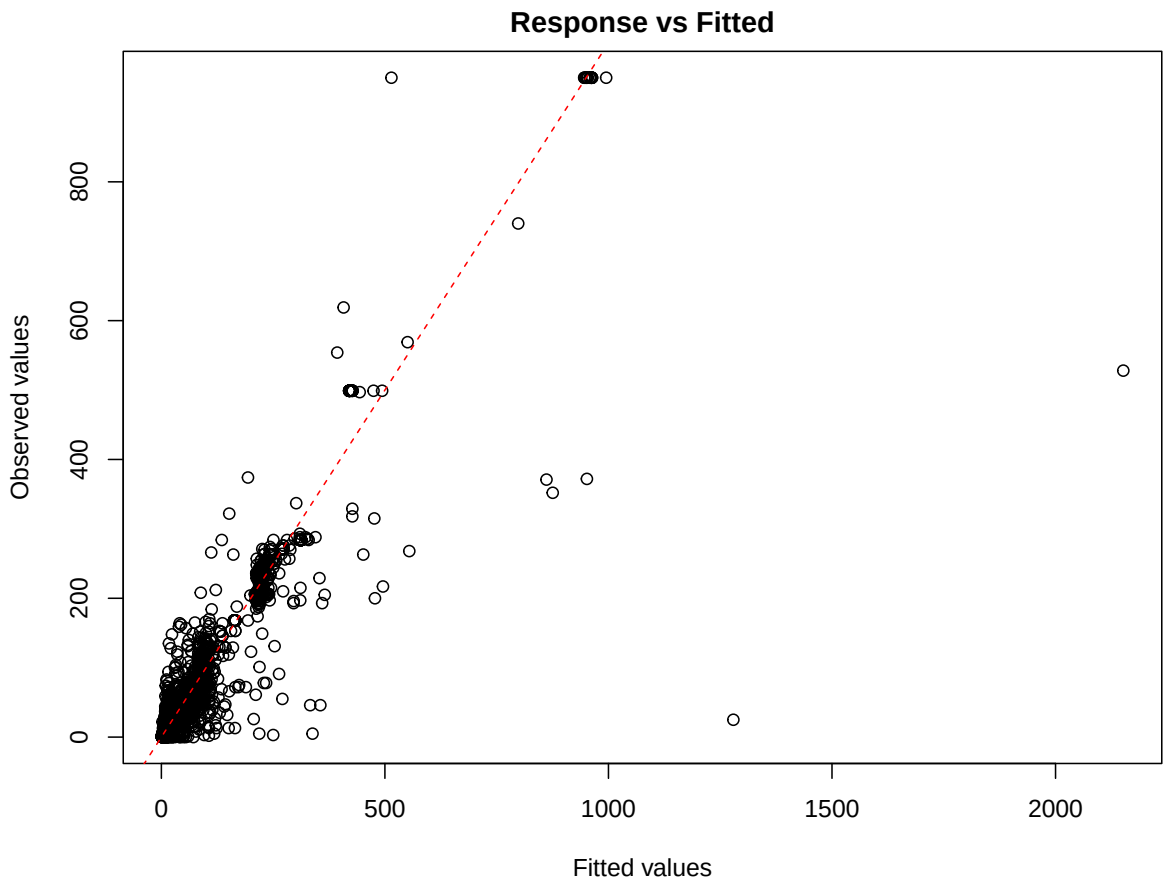

```
### Model 7 : 3c. Temp + Sunlight

Model data dimensions after removing NAs: 5618 x 5
GAM fitted successfully
Model Summary:

Family: Negative Binomial(8.148)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(sunlight_proportion) +
    s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3562     0.1696   19.79   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df  Chi.sq p-value
s(abundance_t_minus_1) 8.921  8.997 36326.3  <2e-16 ***
s(mean_temp)           8.264  8.758   280.8  <2e-16 ***
s(sunlight_proportion) 8.435  8.907   249.2  <2e-16 ***
s(view_id)             7.735  8.000   155.2  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.848   Deviance explained = 88.7%
-REML =  22787  Scale est. = 1         n = 5618


Diagnostic plots:
```



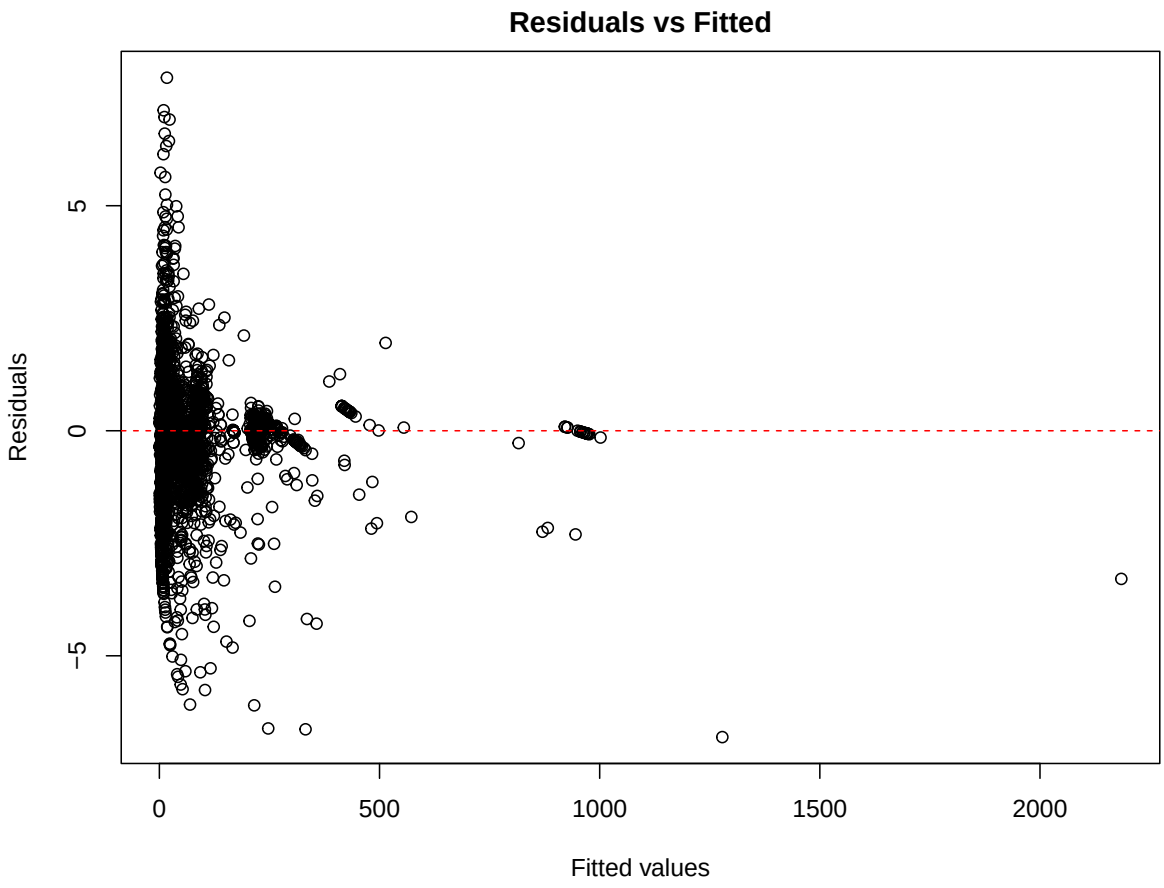

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−5
0
5
−4
−2
0
2
4

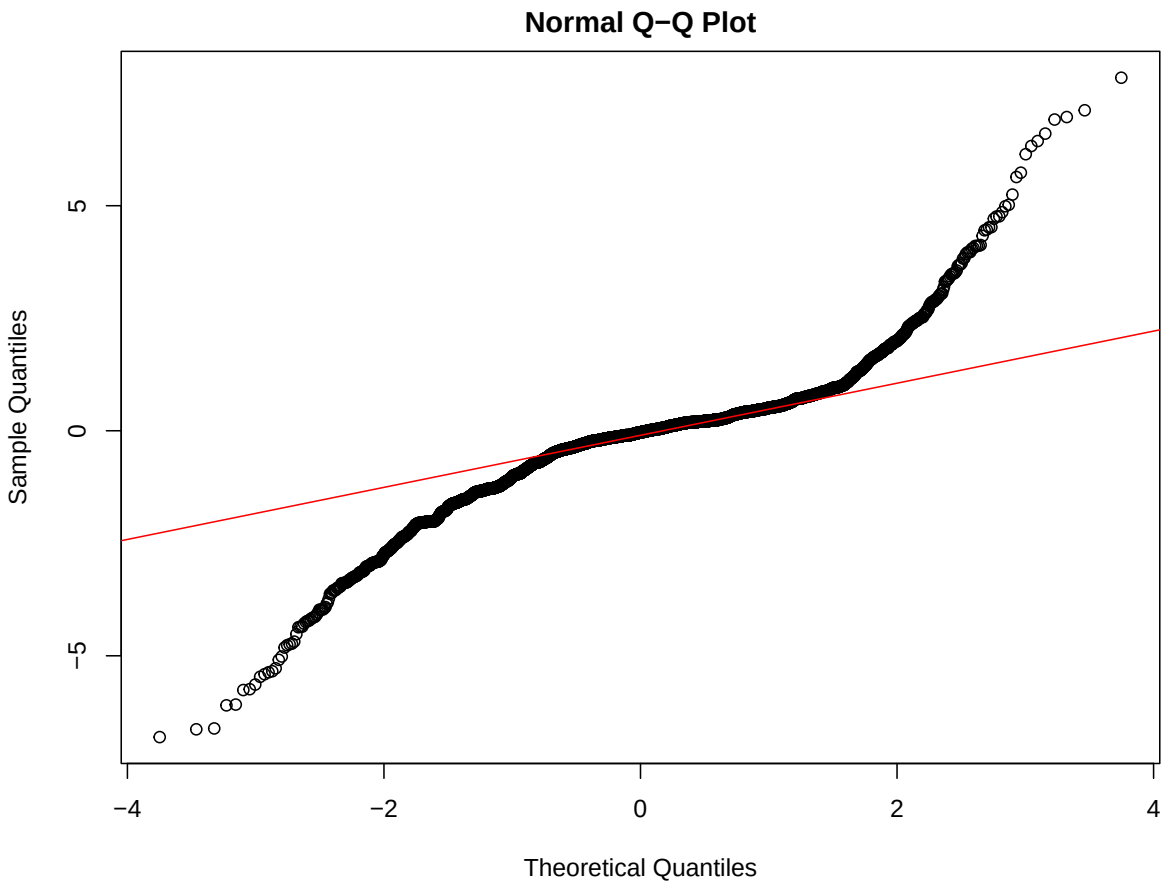

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
Fitted values

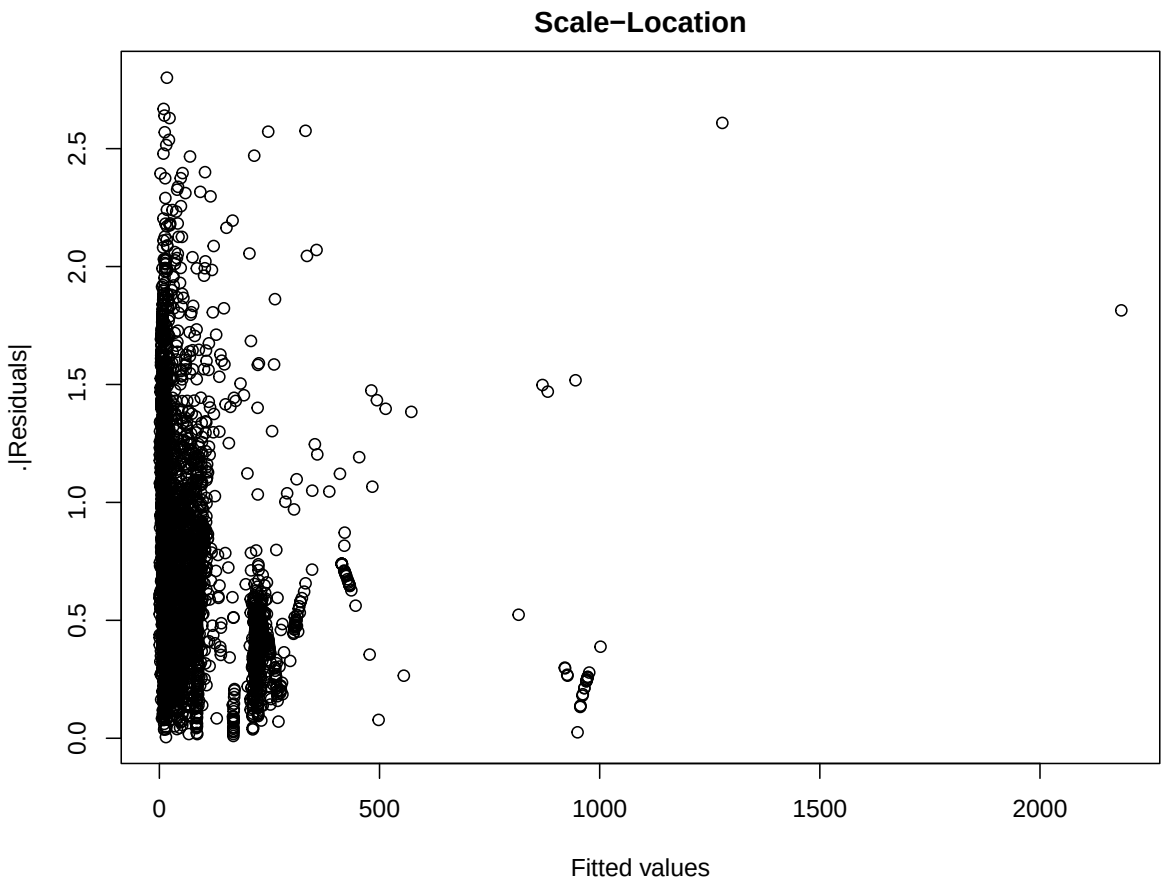


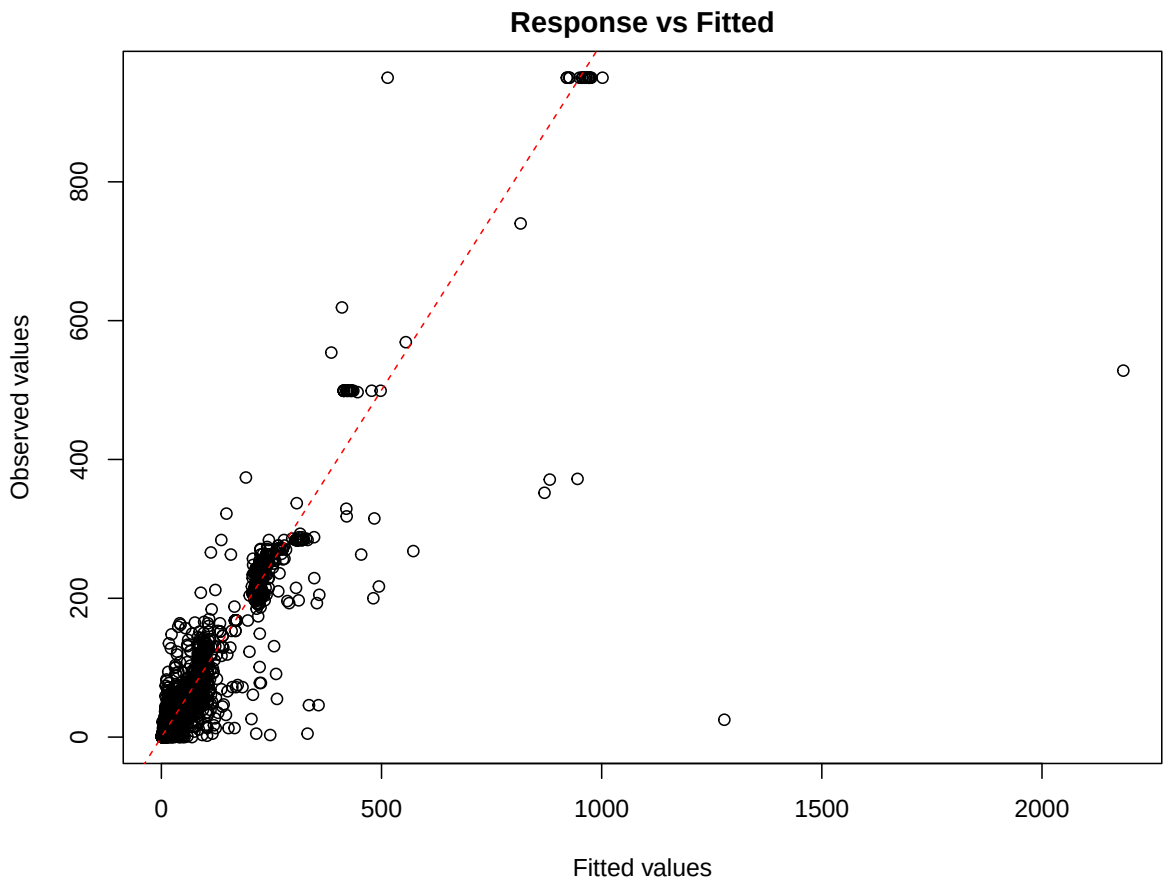

```
### Model 8 : 3d. All additive

Model data dimensions after removing NAs: 5618 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(8.24)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3613     0.1729   19.44   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df   Chi.sq p-value
s(abundance_t_minus_1) 8.922  8.997 35407.36  <2e-16 ***
s(mean_wind_speed)     6.698  7.806    44.16  <2e-16 ***
s(mean_temp)           8.257  8.755   275.74  <2e-16 ***
s(sunlight_proportion) 8.411  8.899   257.69  <2e-16 ***
s(view_id)             7.743  8.000   152.54  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.846   Deviance explained = 88.8%
-REML =  22779  Scale est. = 1         n = 5618


Diagnostic plots:
```


Residuals vs Fitted
Residuals
5
0
−5
0
500
1000
1500
2000
Fitted values

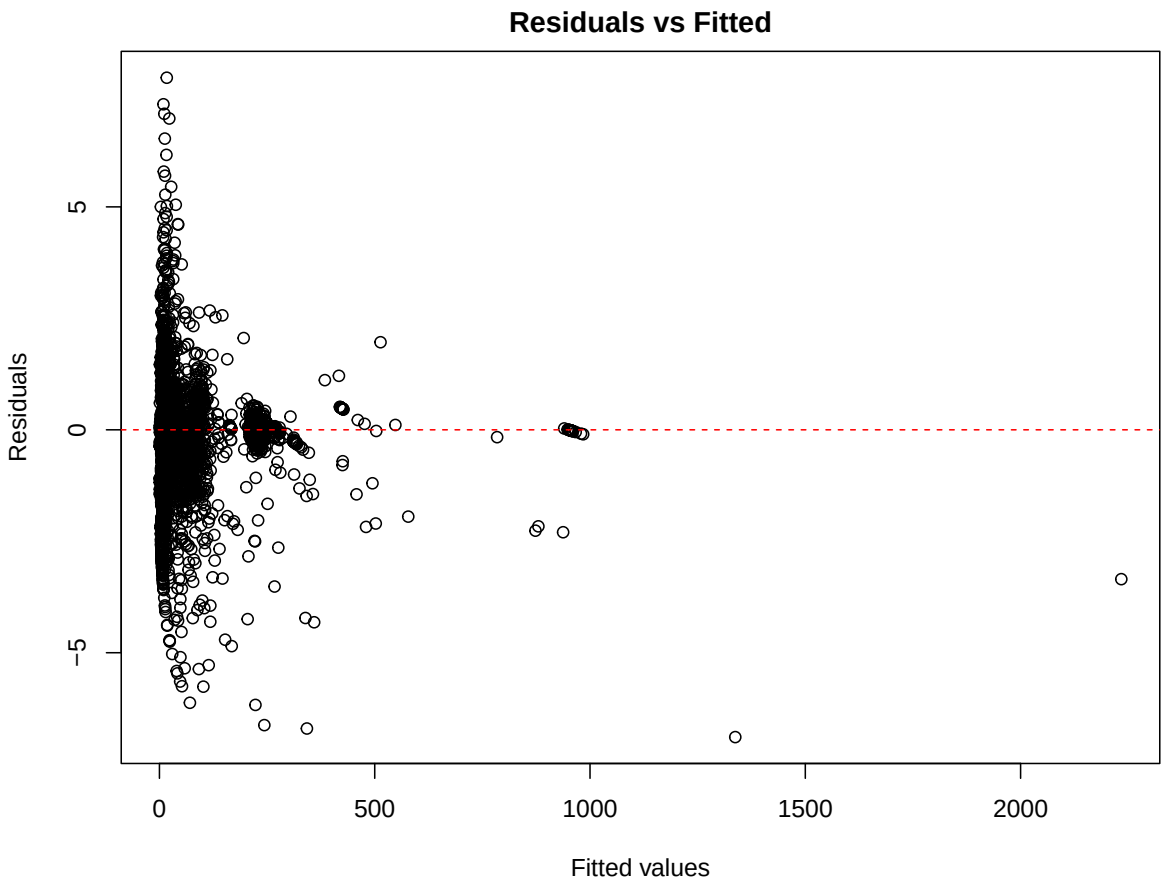

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−5
0
5
−4
−2
0
2
4

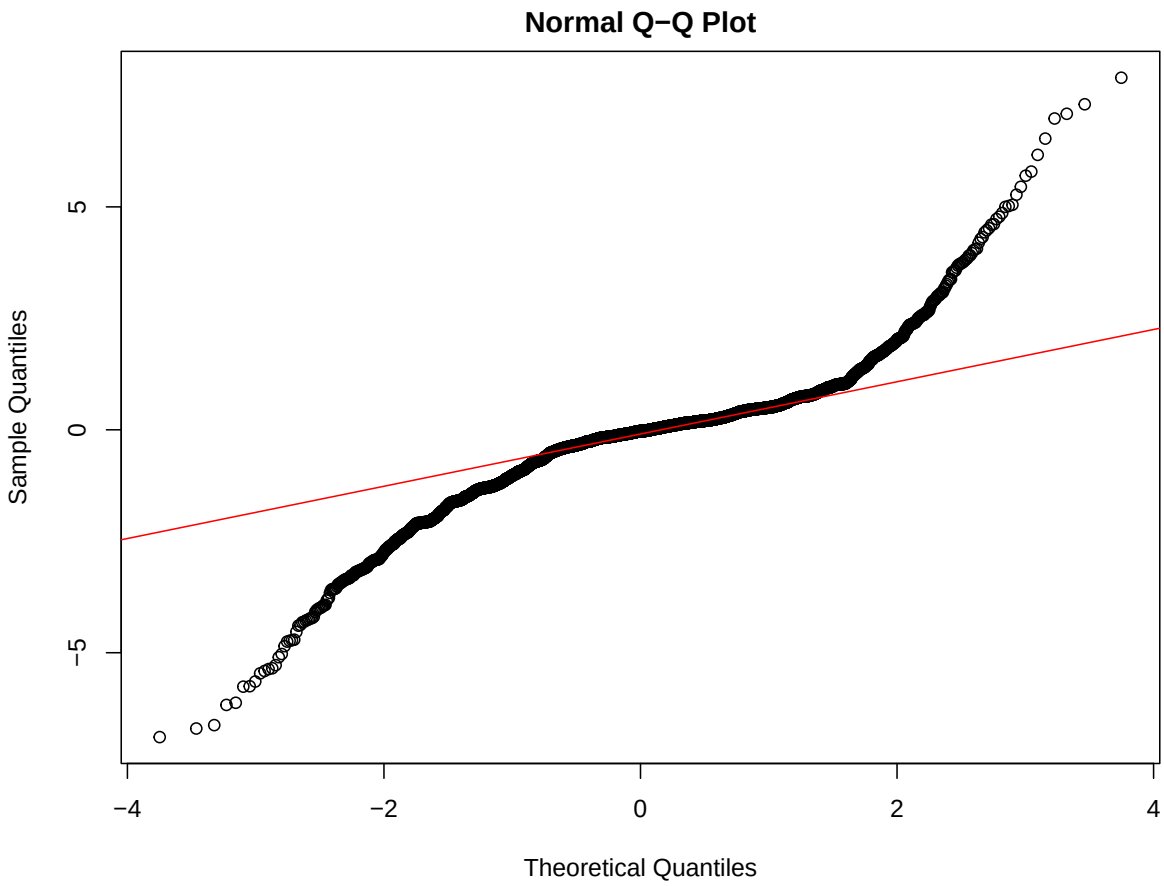

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
Fitted values

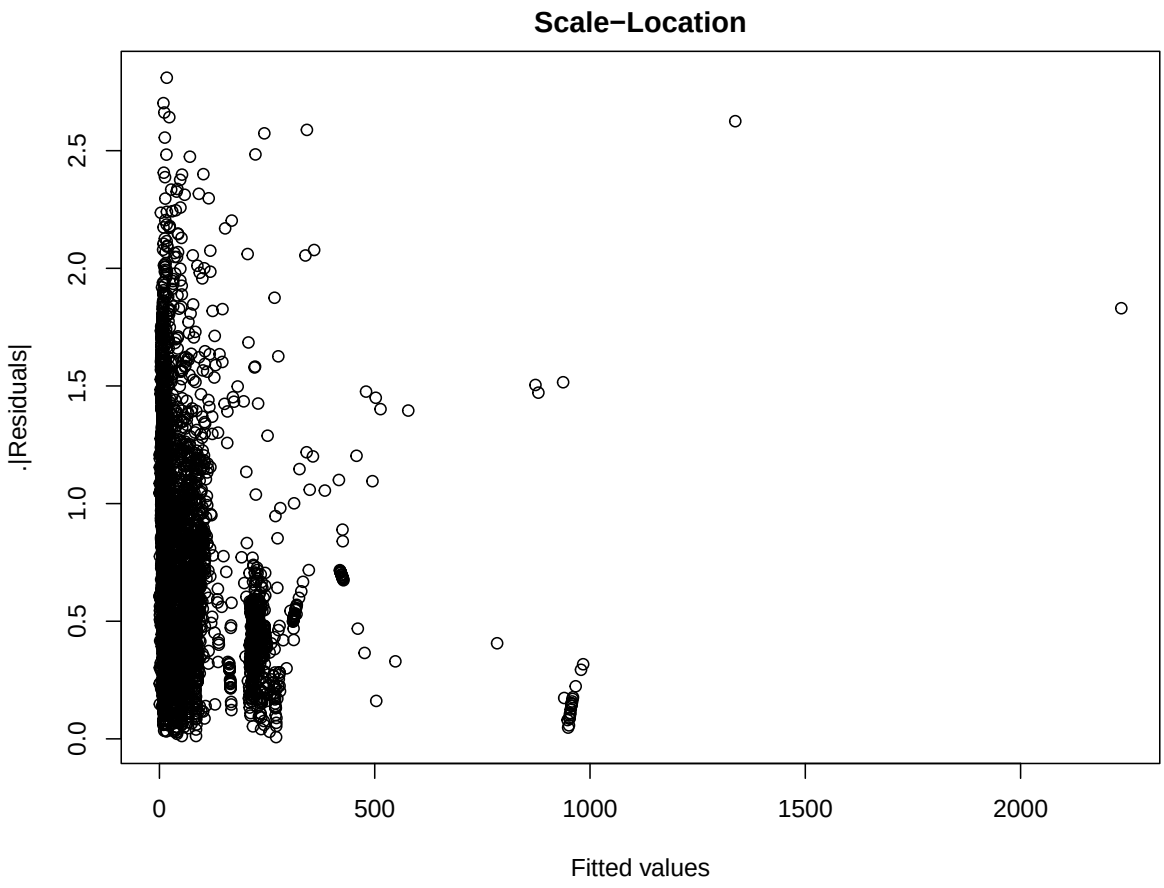


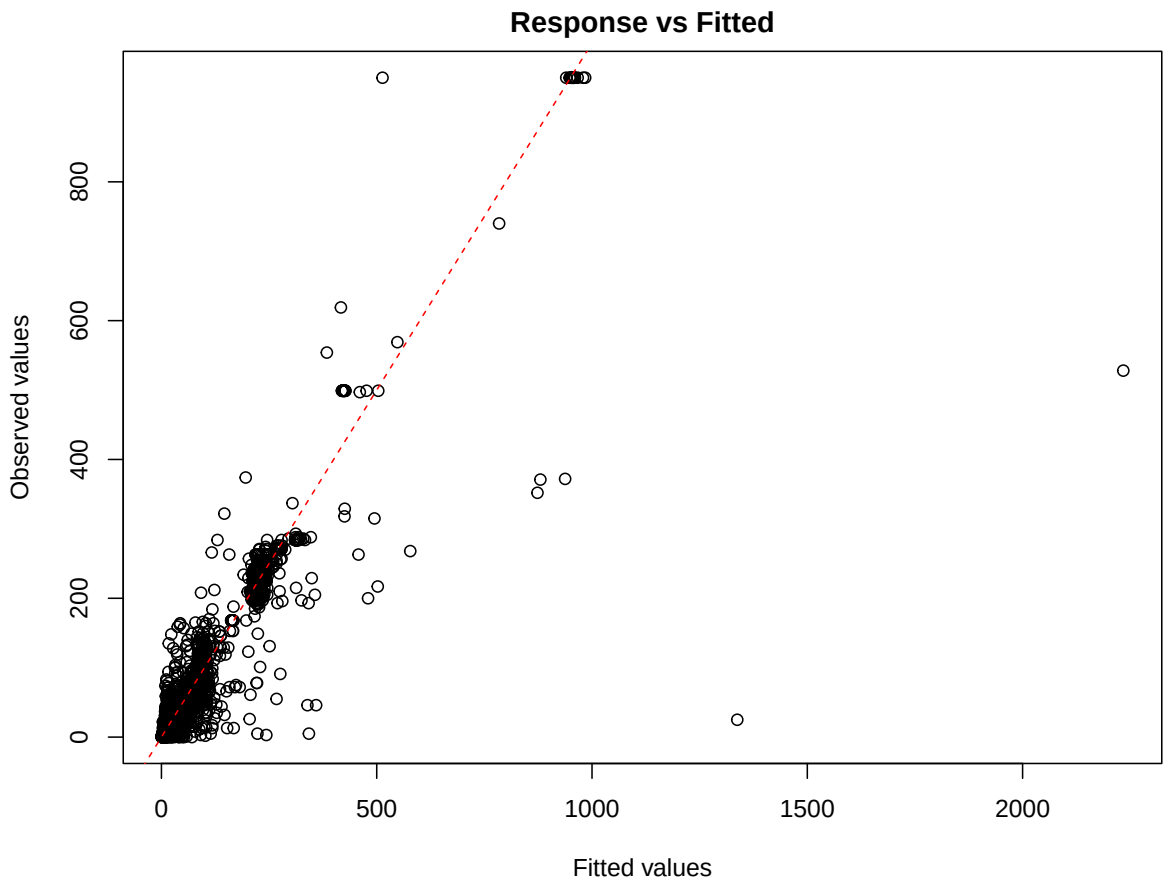

```
### Model 9 : 4a. Wind x Temp interaction

Model data dimensions after removing NAs: 5618 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(8.323)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)    3.366      0.172   19.57   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                edf Ref.df   Chi.sq  p-value
s(abundance_t_minus_1)        8.923  8.997 33440.29  < 2e-16 ***
s(mean_wind_speed)            6.553  7.677    39.26 1.99e-05 ***
s(mean_temp)                  8.322  8.802   254.72  < 2e-16 ***
s(sunlight_proportion)        8.427  8.904   261.48  < 2e-16 ***
ti(mean_wind_speed,mean_temp) 7.921  9.393    48.49  < 2e-16 ***
s(view_id)                    7.739  8.000   148.71  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.839   Deviance explained = 88.9%
-REML =  22766  Scale est. = 1         n = 5618


Diagnostic plots:
```


Residuals vs Fitted
Residuals
5
0
−5
0
500
1000
1500
2000
Fitted values

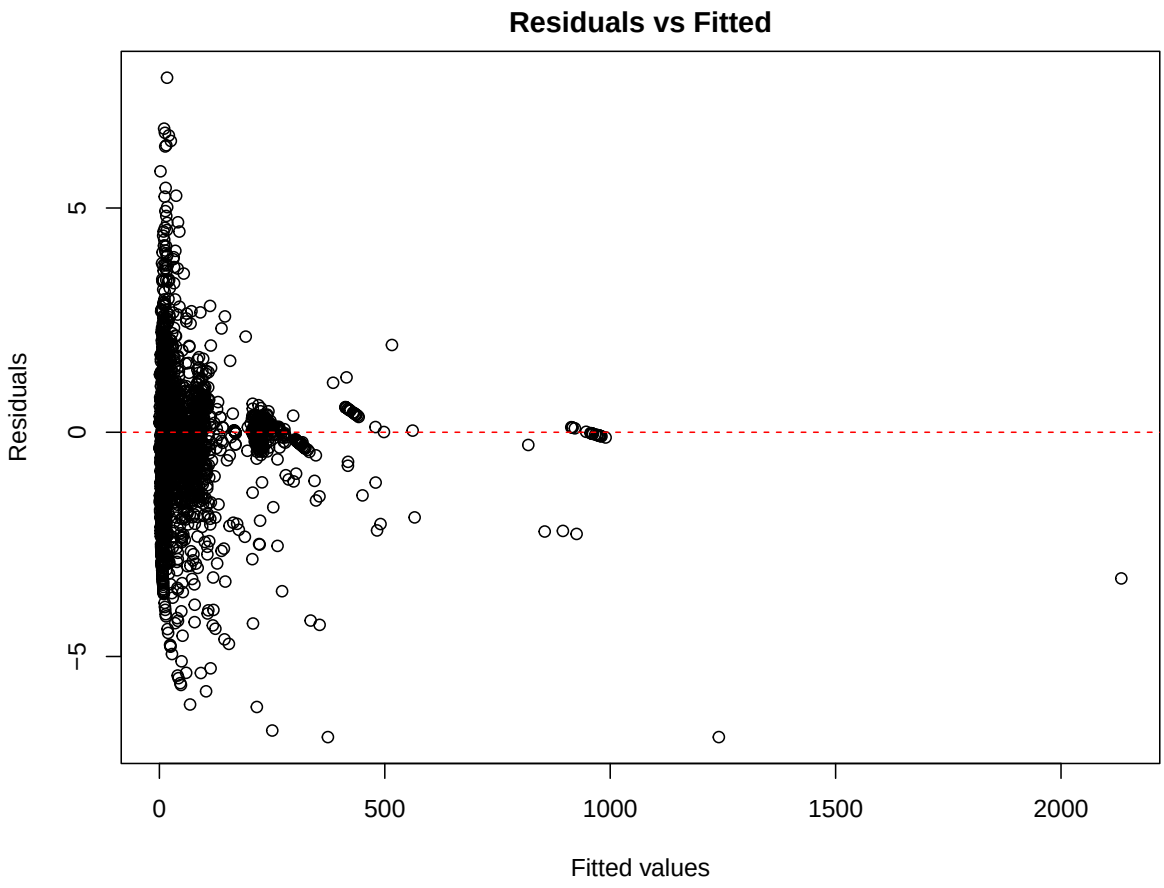

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
5
0
−5
−4
−2
0
2
4

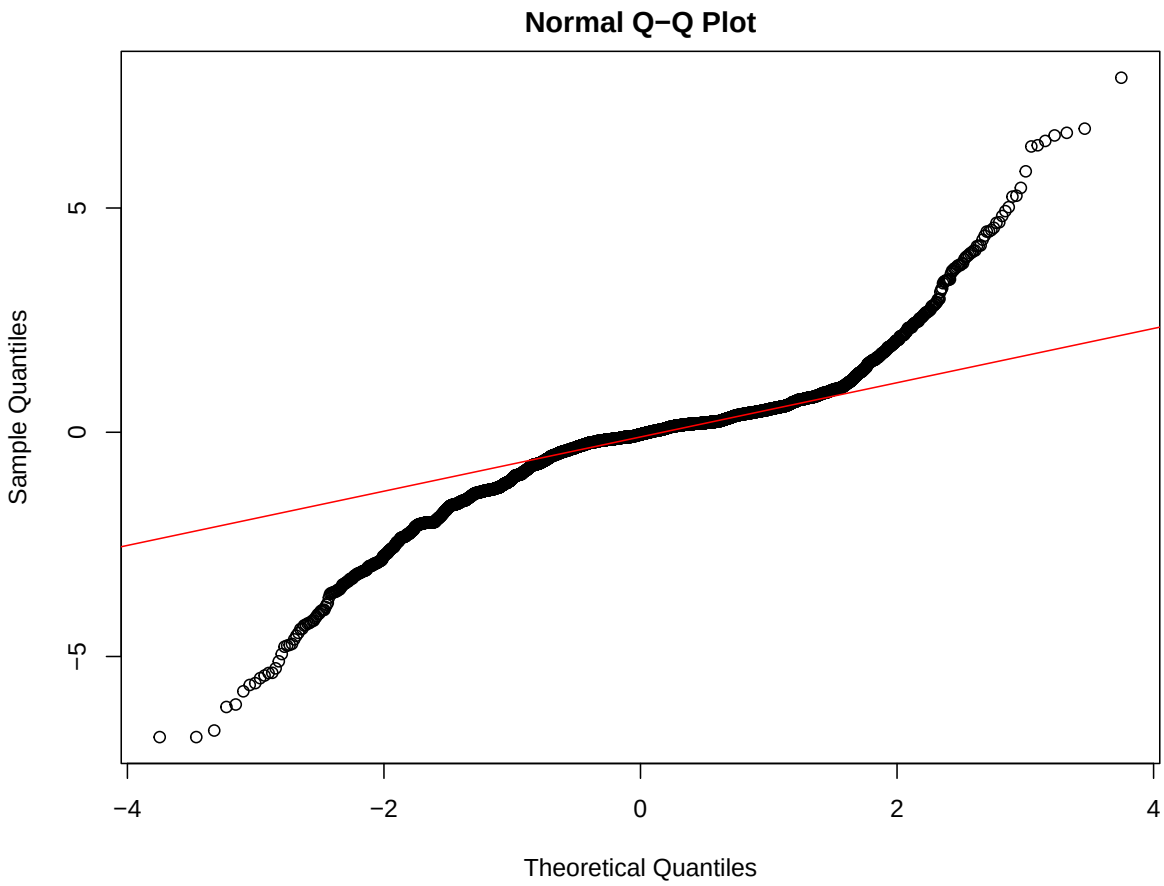

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
Fitted values

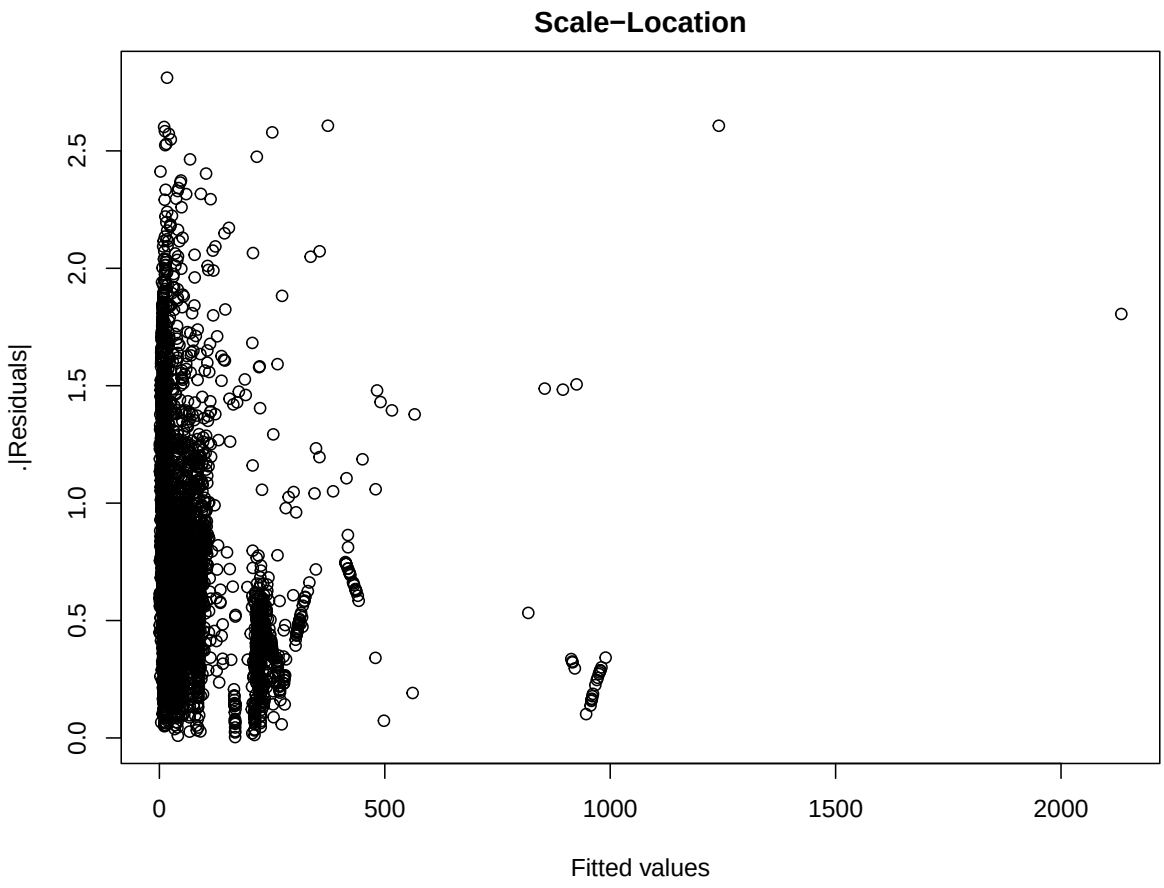


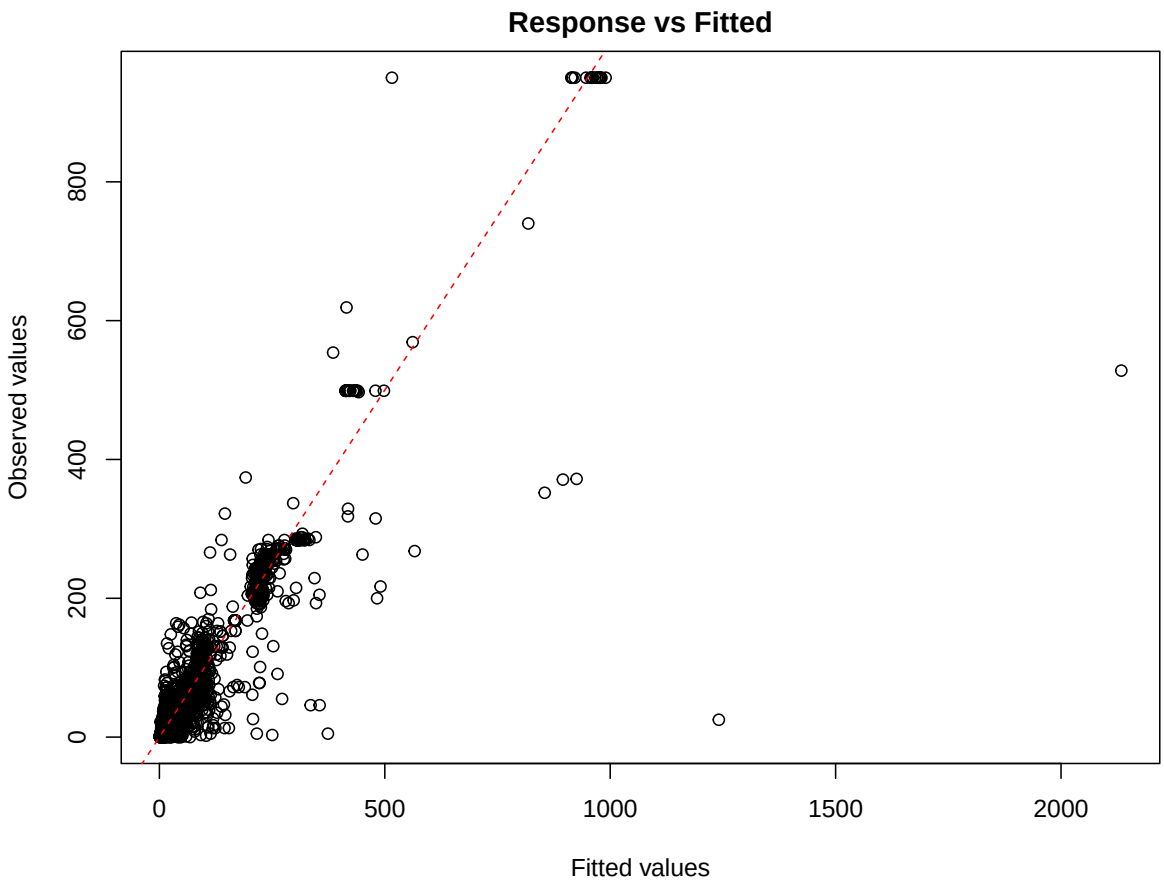

```
### Model 10 : 4b. Wind x Sunlight interaction

Model data dimensions after removing NAs: 5618 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(8.305)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    sunlight_proportion) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)   3.3576     0.1728   19.43   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                          edf Ref.df   Chi.sq  p-value
s(abundance_t_minus_1)                  8.922  8.997 35353.08  < 2e-16 ***
s(mean_wind_speed)                      6.705  7.810    35.98 5.04e-06 ***
s(mean_temp)                            8.249  8.749   253.25  < 2e-16 ***
s(sunlight_proportion)                  8.360  8.880   237.50  < 2e-16 ***
ti(mean_wind_speed,sunlight_proportion) 9.777 11.834    40.40 6.40e-05 ***
s(view_id)                              7.742  8.000   151.44  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.851   Deviance explained = 88.9%
-REML =  22769  Scale est. = 1         n = 5618


Diagnostic plots:
```


Residuals vs Fitted
Residuals
5
0
−5
0
500
1000
1500
2000
Fitted values

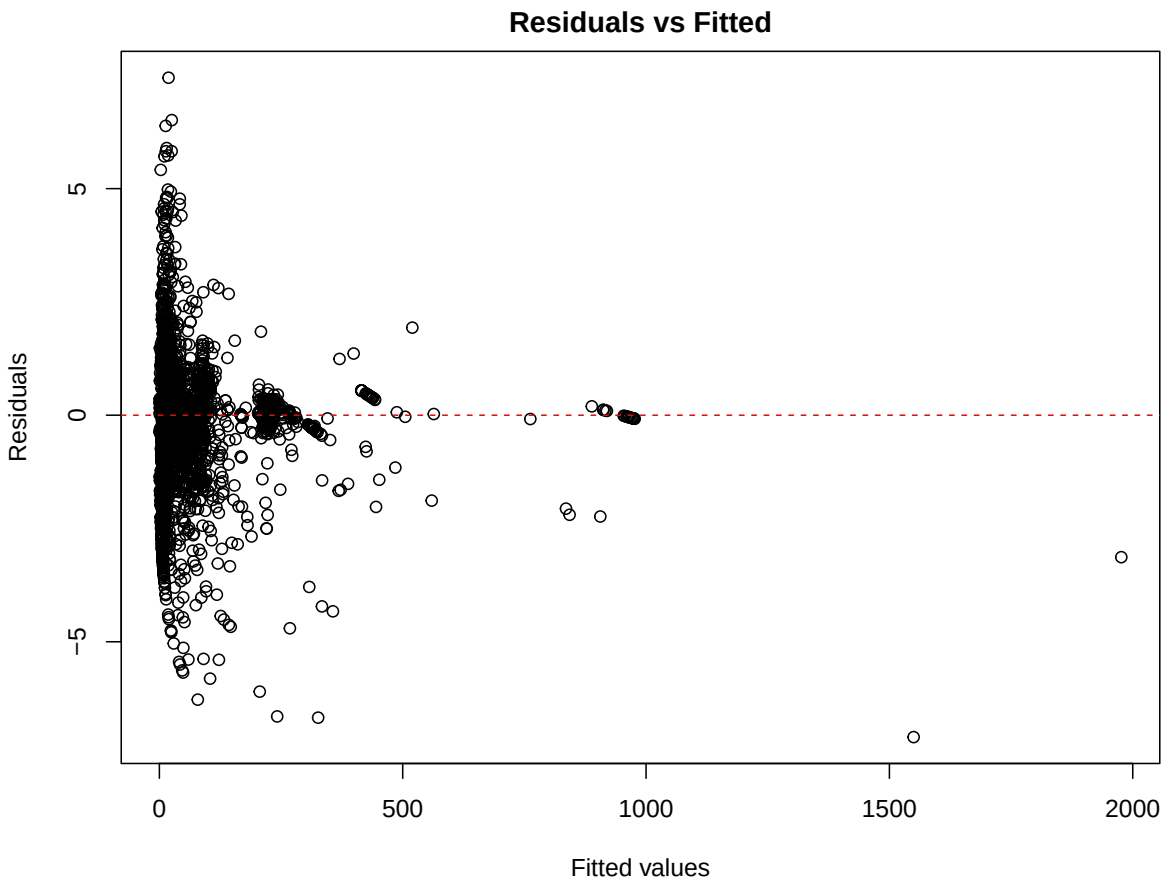

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−5
0
5
−4
−2
0
2
4

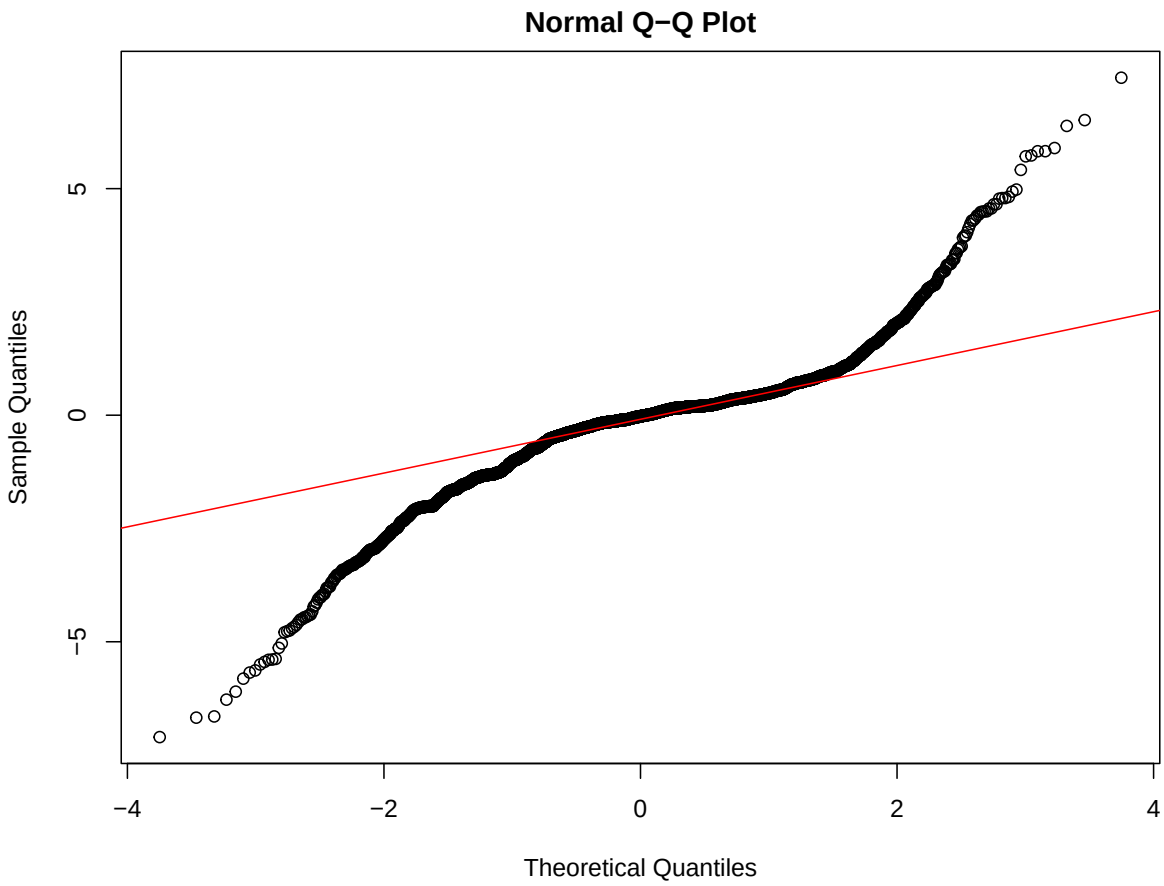

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
Fitted values

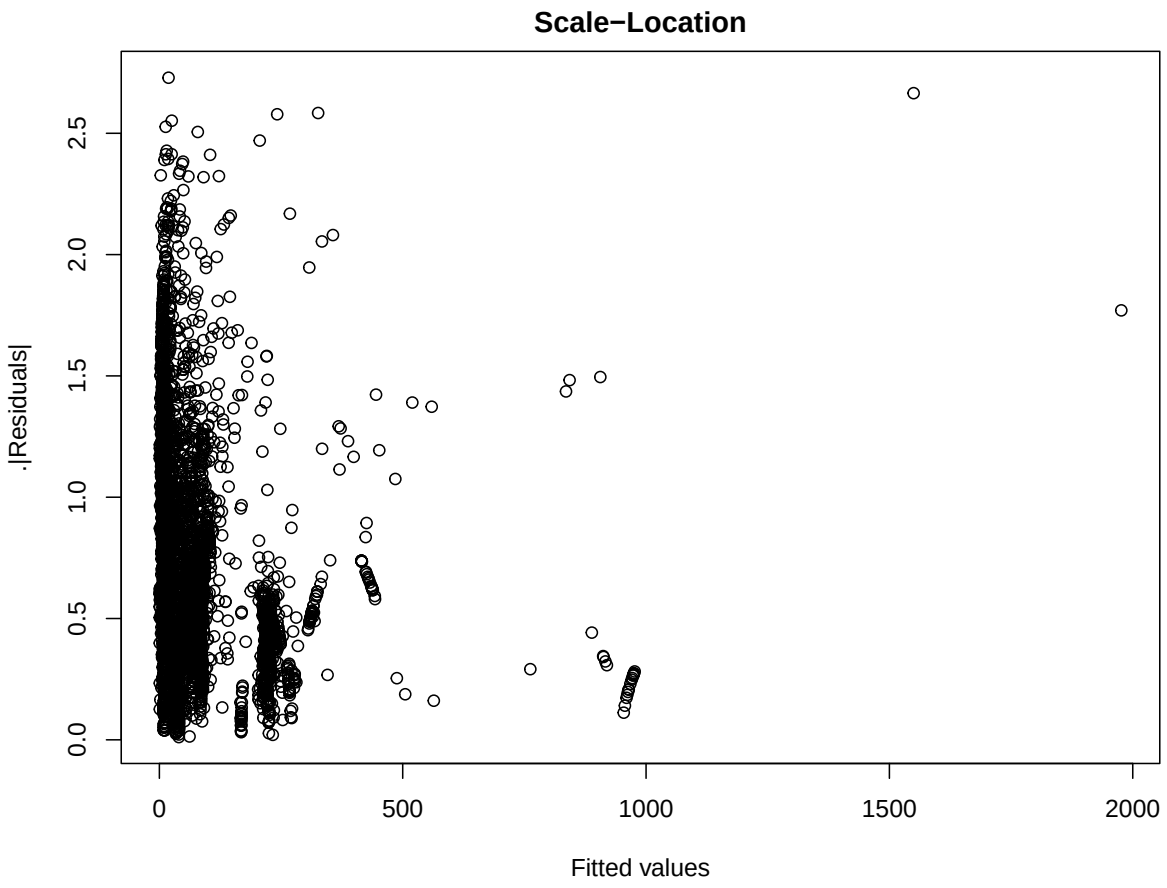


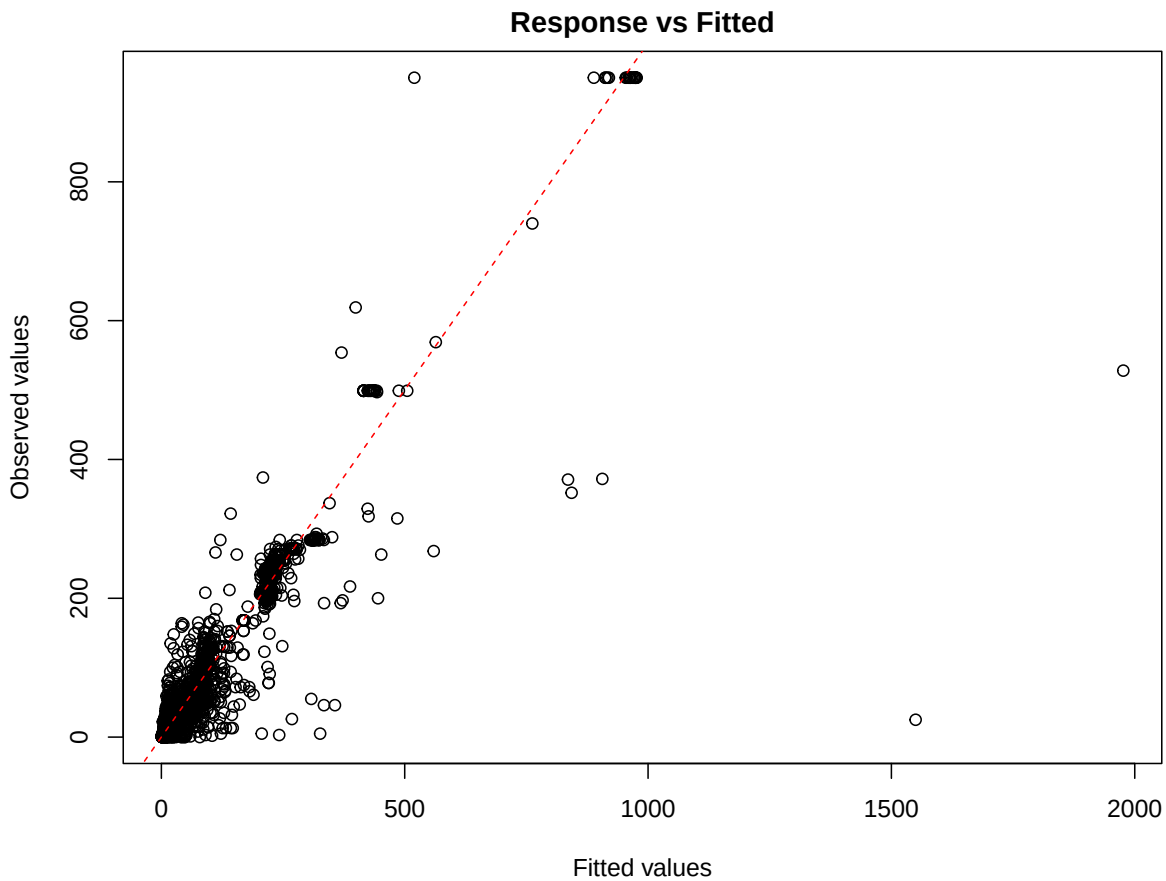

```
### Model 11 : 4c. Temp x Sunlight interaction

Model data dimensions after removing NAs: 5618 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(8.422)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_temp, sunlight_proportion) +
    s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)   3.3342     0.1683   19.82   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                    edf Ref.df   Chi.sq p-value
s(abundance_t_minus_1)            8.922  8.997 35197.13  <2e-16 ***
s(mean_wind_speed)                6.701  7.809    40.09  <2e-16 ***
s(mean_temp)                      8.195  8.709   187.57  <2e-16 ***
s(sunlight_proportion)            8.007  8.728    84.61  <2e-16 ***
ti(mean_temp,sunlight_proportion) 8.950 10.999   111.80  <2e-16 ***
s(view_id)                        7.733  8.000   154.15  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.847   Deviance explained = 89.1%
-REML =  22734  Scale est. = 1         n = 5618


Diagnostic plots:
```


Residuals vs Fitted
Residuals
5
0
−5
0
500
1000
1500
2000
Fitted values


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−5
0
5
−4
−2
0
2
4

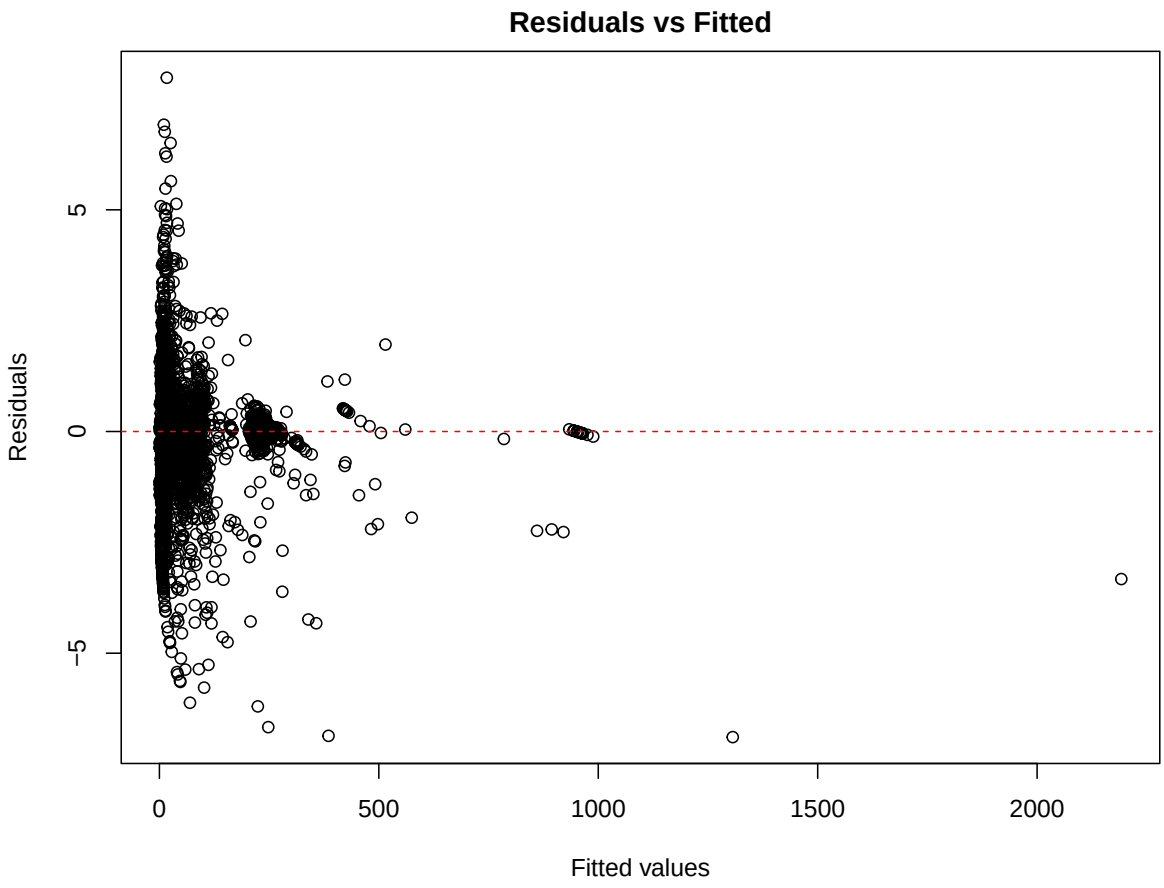

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
Fitted values

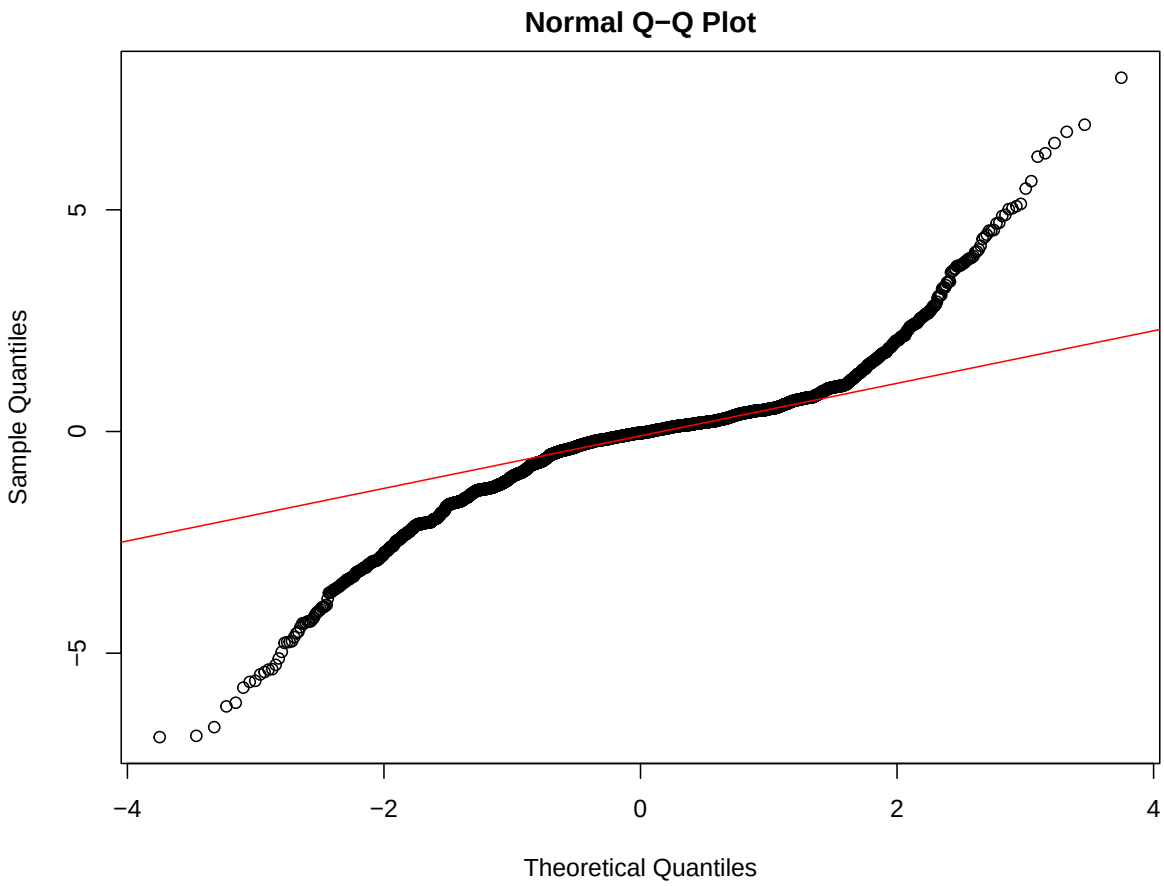


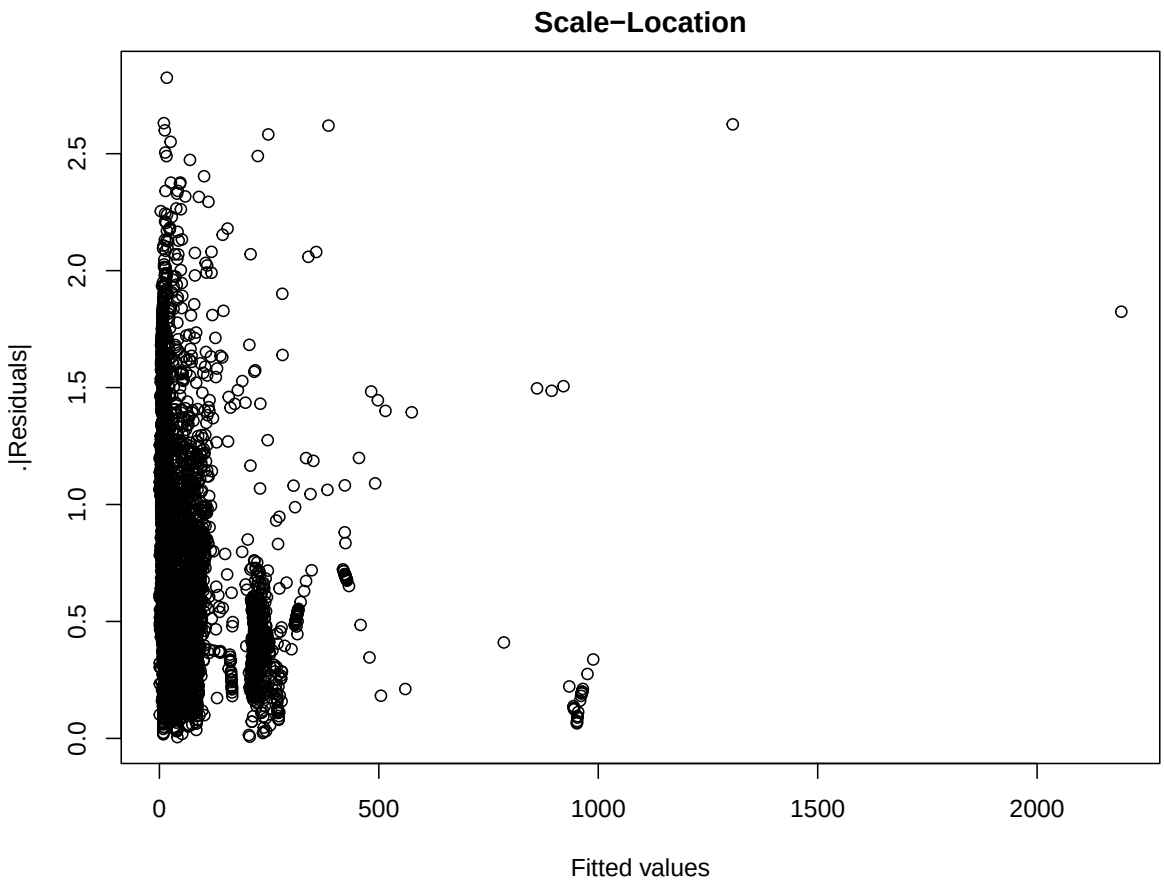

```
### Model 12 : 5a. Wind x Temp + Wind x Sun

Model data dimensions after removing NAs: 5618 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(8.394)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + ti(mean_wind_speed, sunlight_proportion) + s(view_id,
    bs = "re")

Parametric coefficients:
```

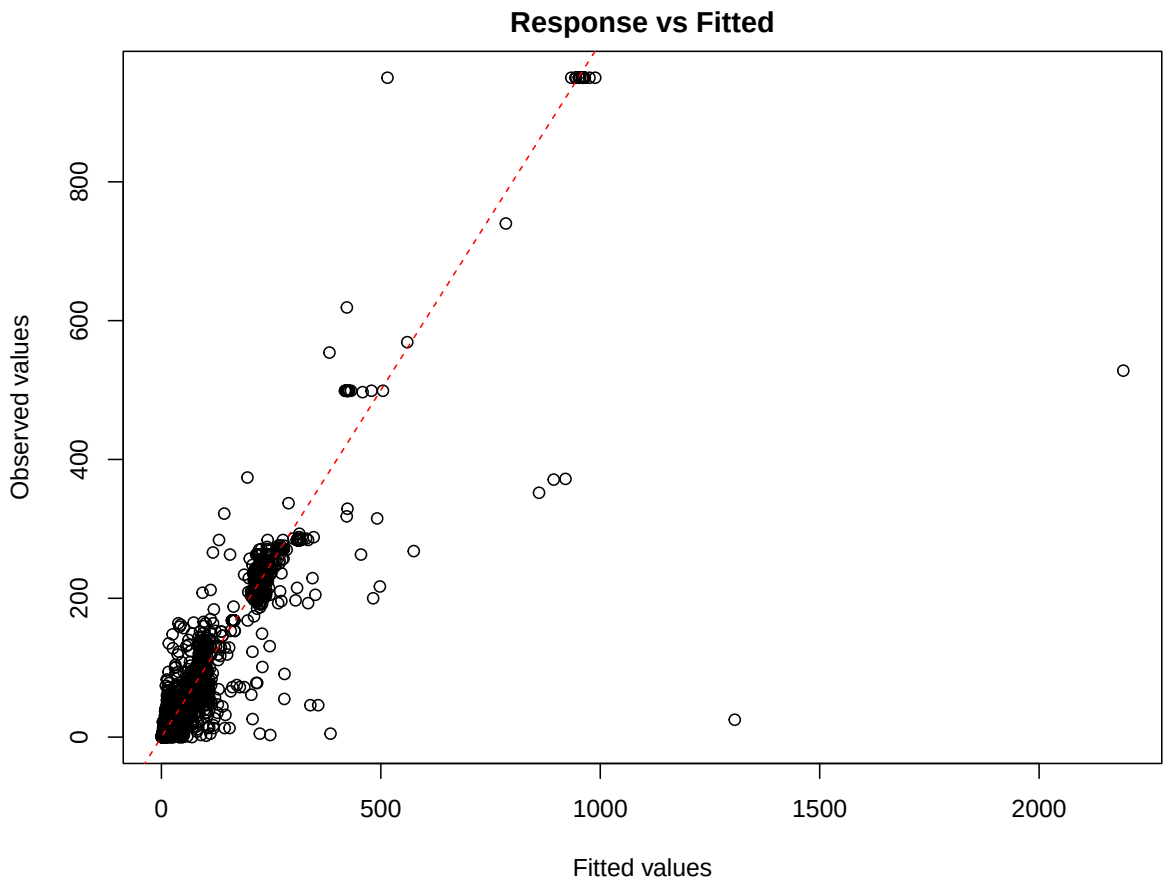
```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3612     0.1718   19.56   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                          edf Ref.df   Chi.sq  p-value
s(abundance_t_minus_1)                  8.923  8.997 33202.71  < 2e-16 ***
s(mean_wind_speed)                      6.591  7.707    36.99 4.32e-07 ***
s(mean_temp)                            8.313  8.796   239.55  < 2e-16 ***
s(sunlight_proportion)                  8.381  8.887   241.24  < 2e-16 ***
ti(mean_wind_speed,mean_temp)           8.204  9.650    52.01  < 2e-16 ***
ti(mean_wind_speed,sunlight_proportion) 9.893 11.926    42.80 2.92e-05 ***
s(view_id)                              7.738  8.000   148.65  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.843   Deviance explained =   89%
-REML =  22754  Scale est. = 1         n = 5618
```

Diagnostic plots:


Residuals vs Fitted
Residuals
5
0
−5
0
500
1000
1500
2000
Fitted values

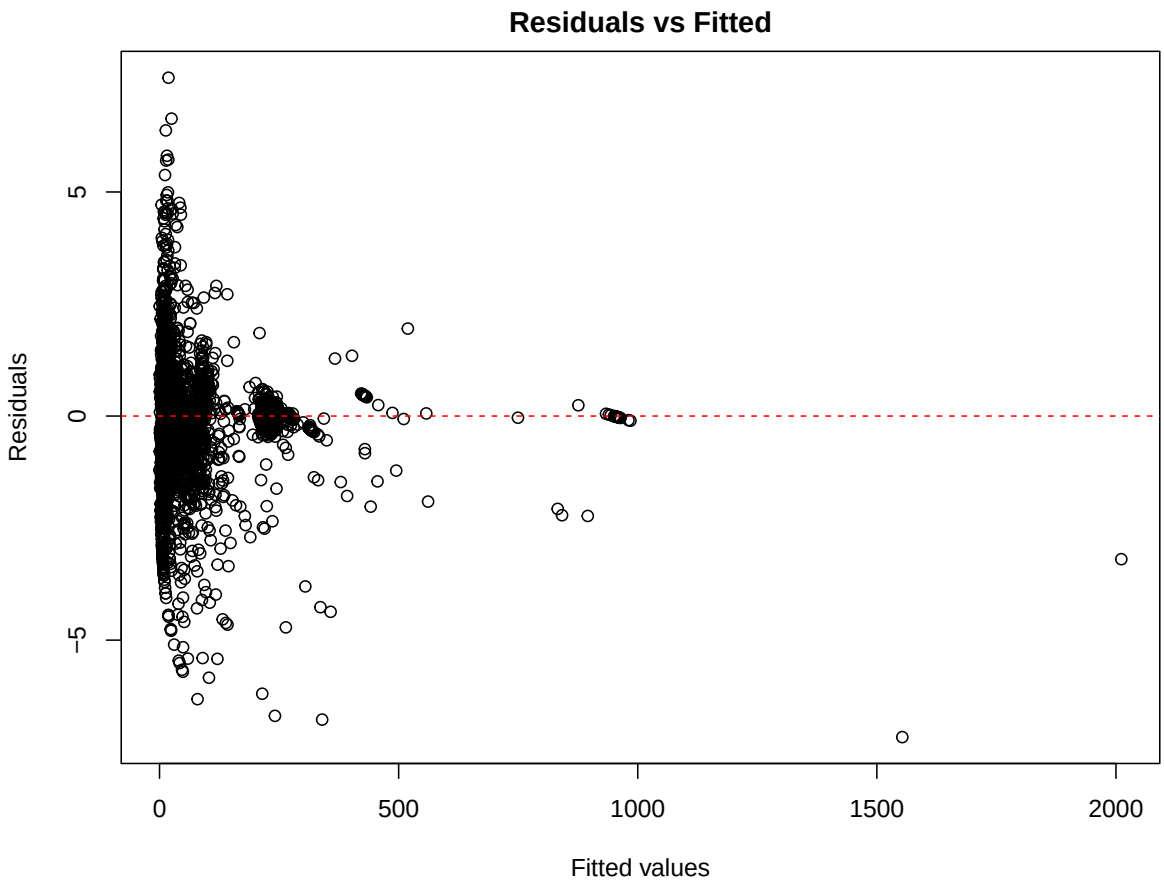

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−5
0
5
−4
−2
0
2
4

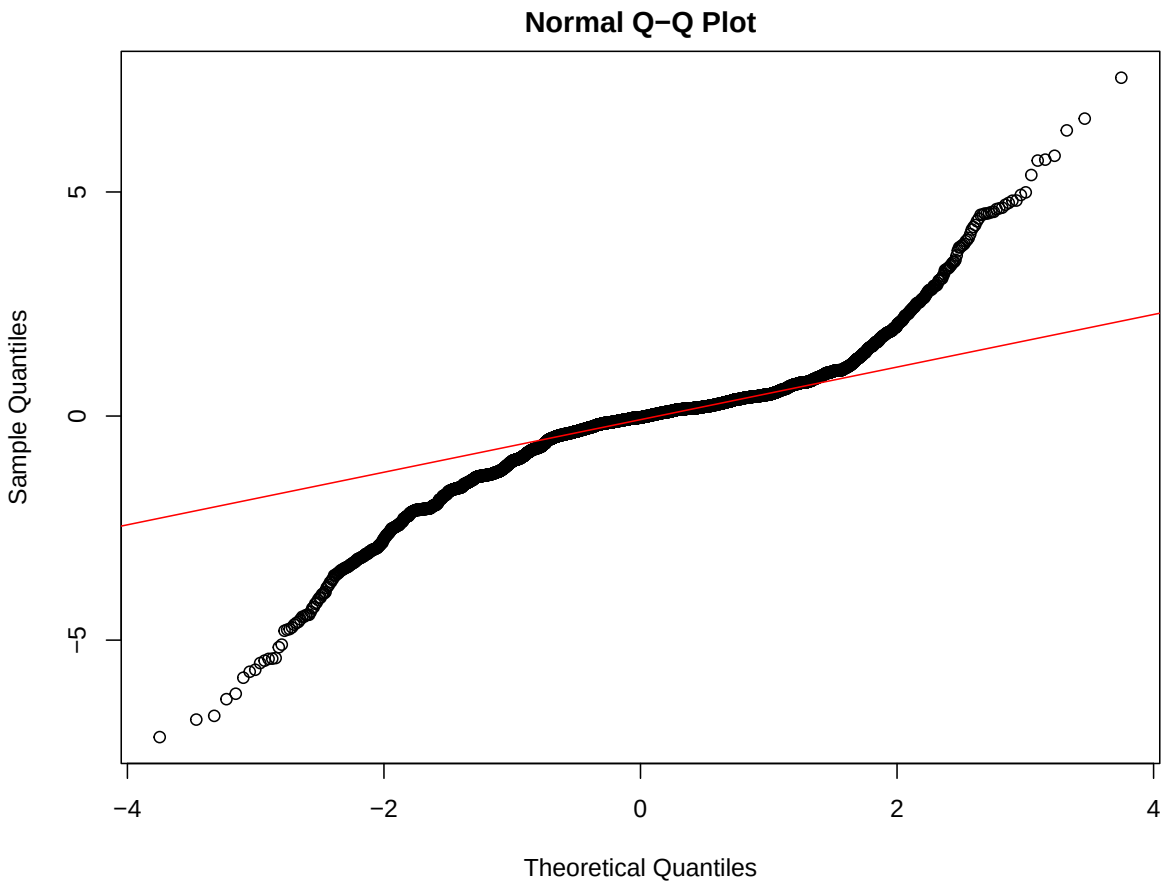

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
Fitted values

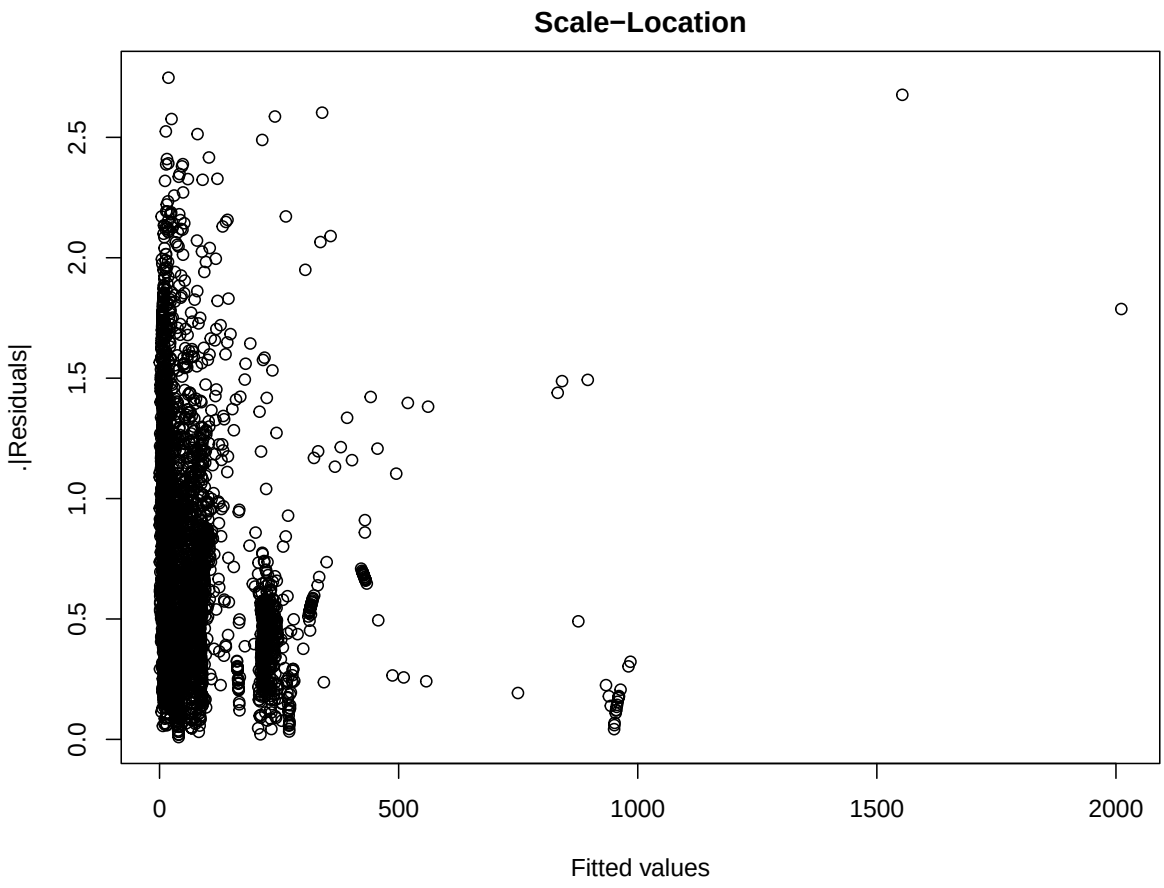


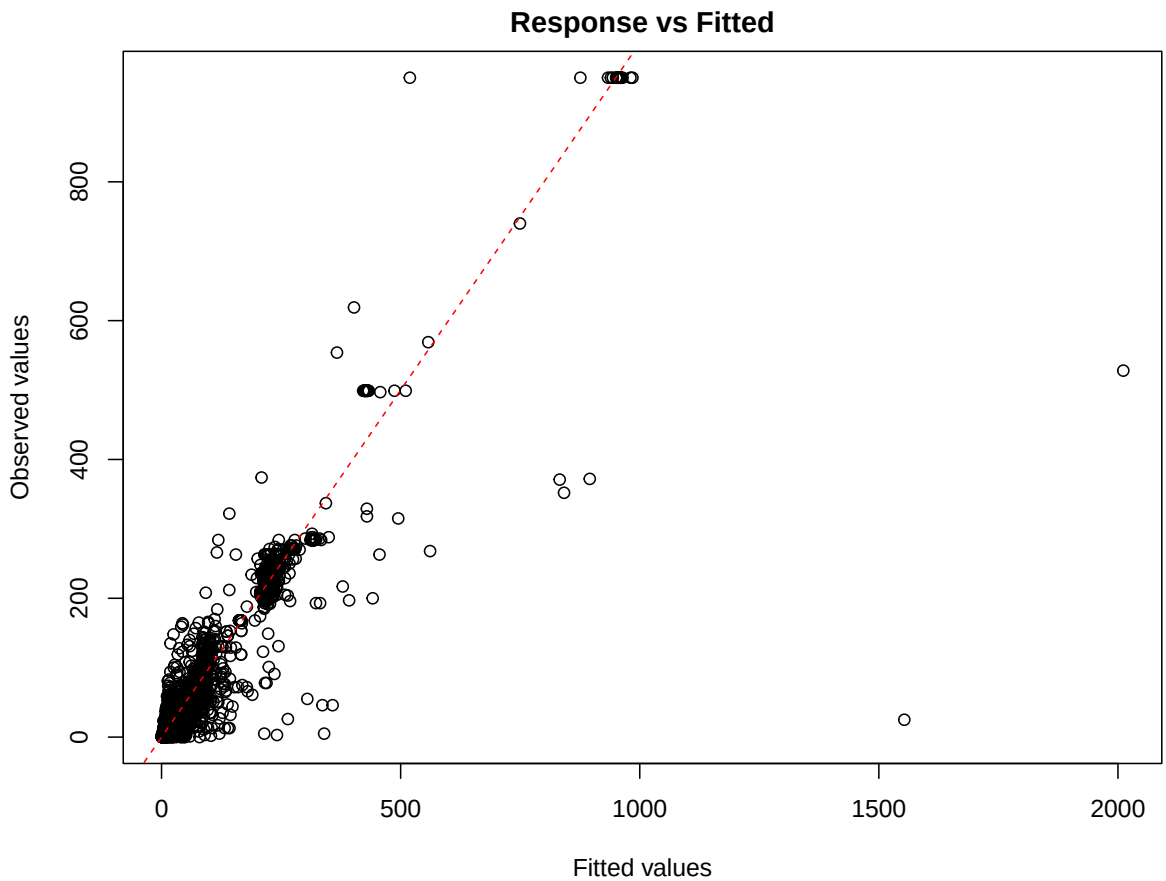

```
### Model 13 : 5b. Wind x Temp + Temp x Sun

Model data dimensions after removing NAs: 5618 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(8.565)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + ti(mean_temp, sunlight_proportion) + s(view_id,
    bs = "re")

Parametric coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3449     0.1681    19.9   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                     edf Ref.df   Chi.sq p-value
s(abundance_t_minus_1)             8.923  8.997 33407.39  <2e-16 ***
s(mean_wind_speed)                 6.525  7.653    48.73  <2e-16 ***
s(mean_temp)                       8.363  8.810   189.11  <2e-16 ***
s(sunlight_proportion)             7.946  8.697    81.85  <2e-16 ***
ti(mean_wind_speed,mean_temp)      7.963  9.406    62.90  <2e-16 ***
ti(mean_temp,sunlight_proportion) 10.197 12.158   128.26  <2e-16 ***
s(view_id)                         7.731  8.000   151.74  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.845   Deviance explained = 89.2%
-REML =  22713  Scale est. = 1         n = 5618


Diagnostic plots:
```


Residuals vs Fitted
Residuals
5
0
−5
0
500
1000
1500
2000
Fitted values

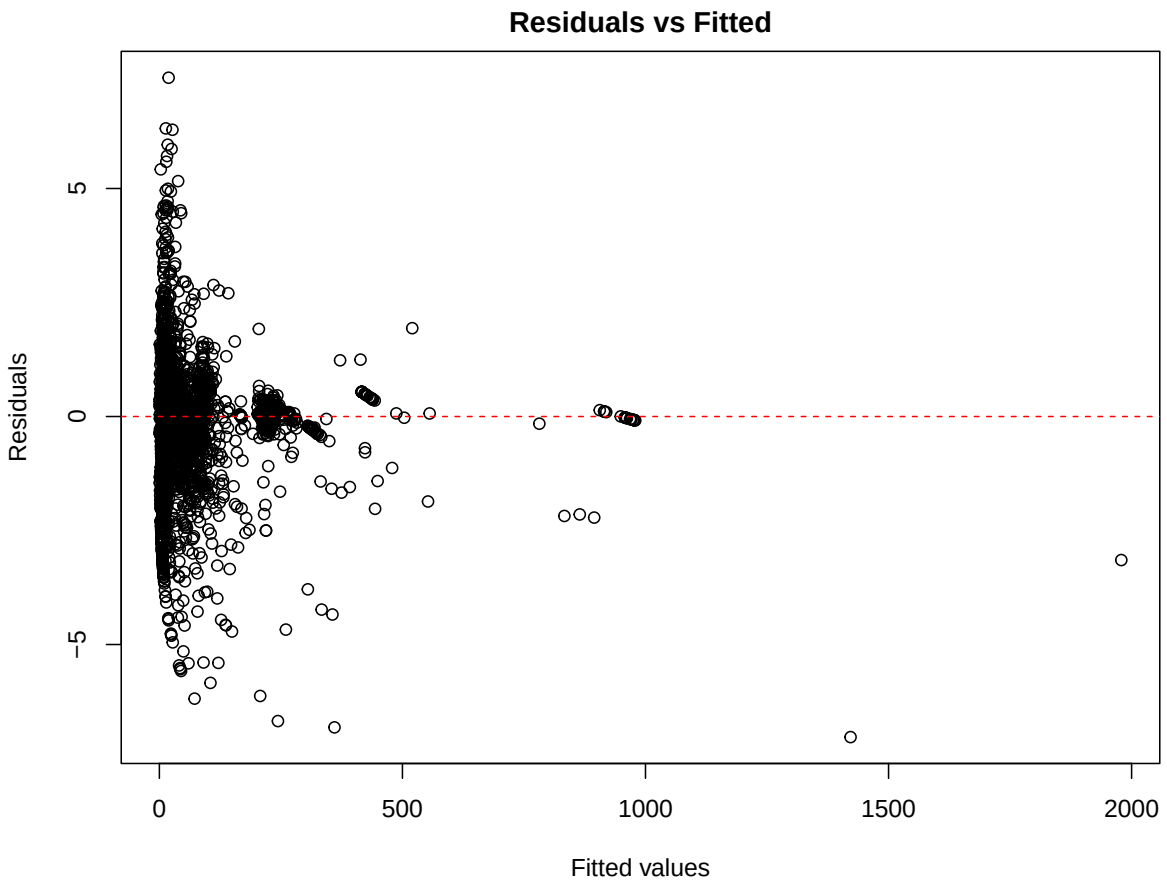

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
5
0
−5
−4
−2
0
2
4

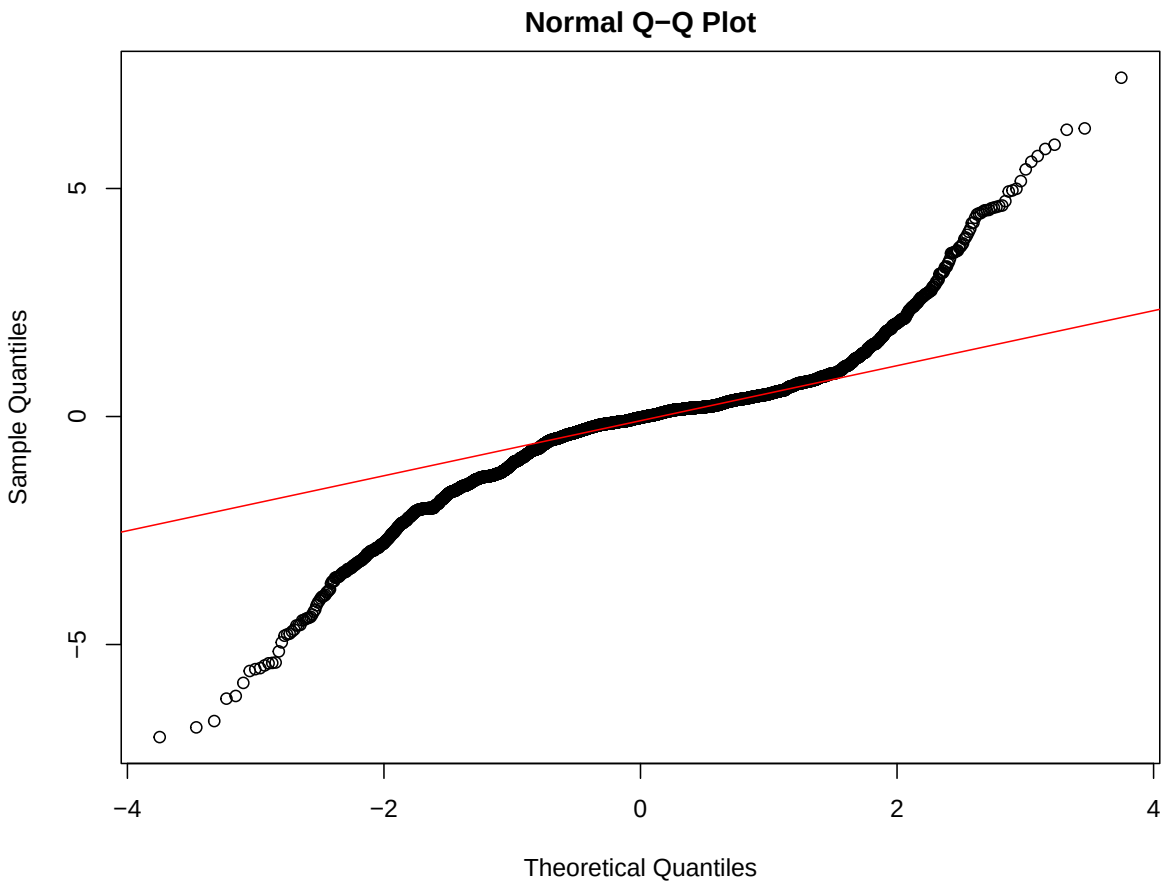

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
Fitted values

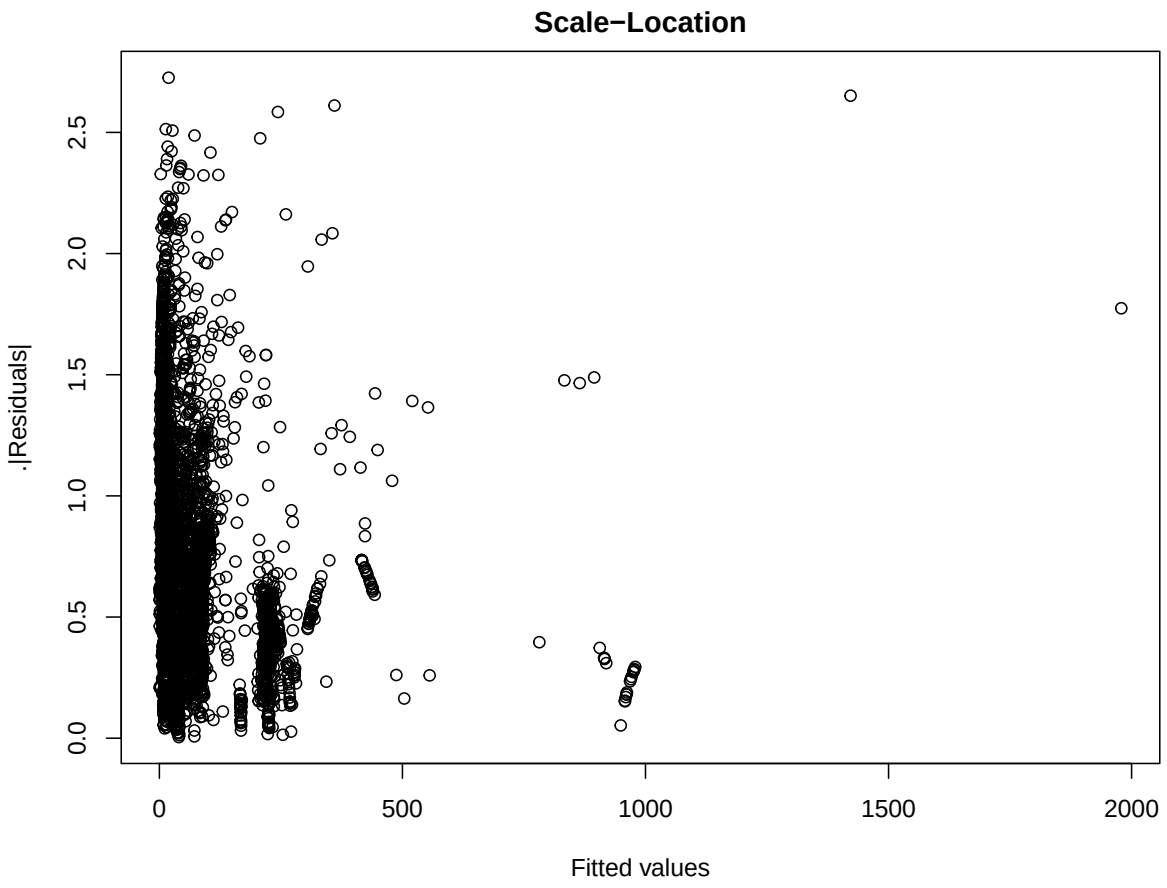


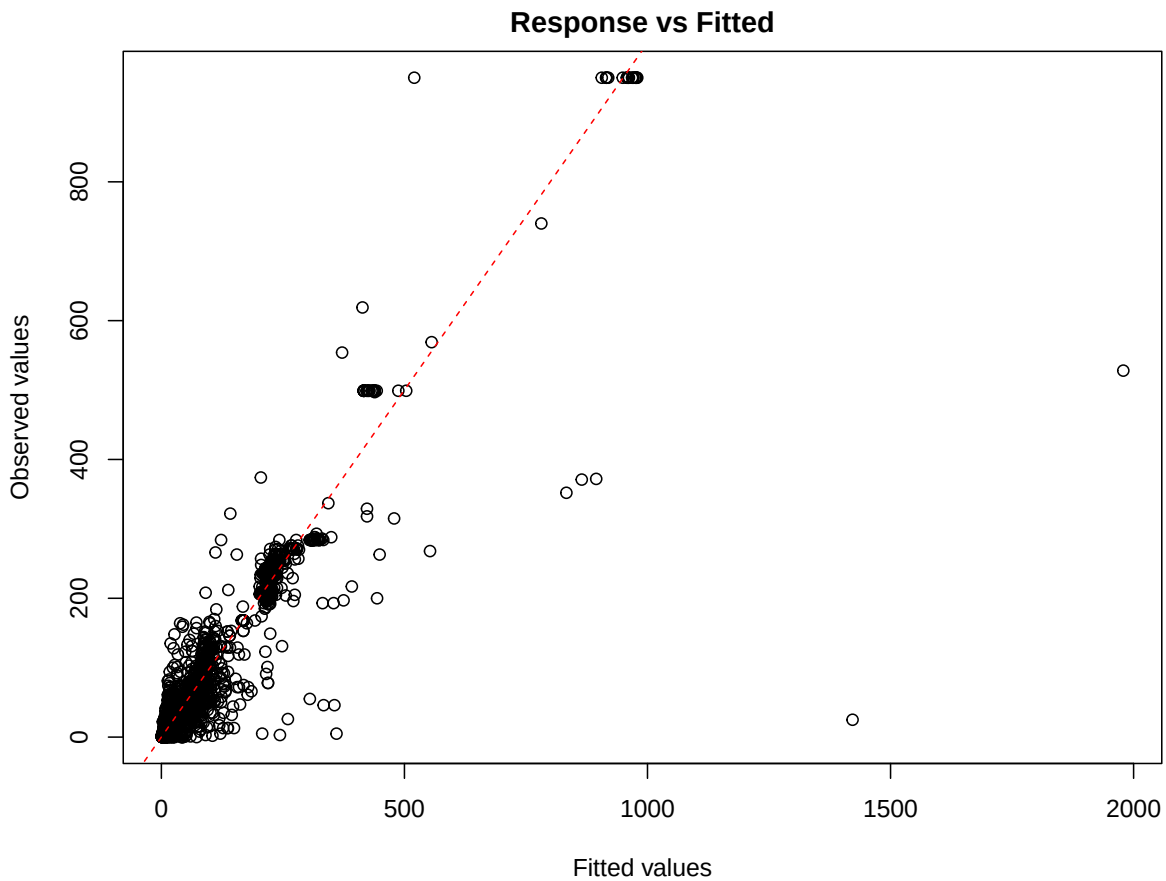

```
### Model 14 : 5c. Wind x Sun + Temp x Sun

Model data dimensions after removing NAs: 5618 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(8.489)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    sunlight_proportion) + ti(mean_temp, sunlight_proportion) +
    s(view_id, bs = "re")

Parametric coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3341     0.1688   19.75   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                          edf Ref.df   Chi.sq  p-value
s(abundance_t_minus_1)                  8.922  8.997 35269.12  < 2e-16 ***
s(mean_wind_speed)                      6.658  7.768    36.38 2.78e-06 ***
s(mean_temp)                            8.190  8.705   182.68  < 2e-16 ***
s(sunlight_proportion)                  7.895  8.667    78.91  < 2e-16 ***
ti(mean_wind_speed,sunlight_proportion) 9.687 11.749    36.08 0.000406 ***
ti(mean_temp,sunlight_proportion)       9.018 11.050   105.25  < 2e-16 ***
s(view_id)                              7.734  8.000   156.63  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.854   Deviance explained = 89.1%
-REML =  22728  Scale est. = 1         n = 5618


Diagnostic plots:
```


Residuals vs Fitted
Residuals
5
0
−5
0
500
1000
1500
2000
Fitted values

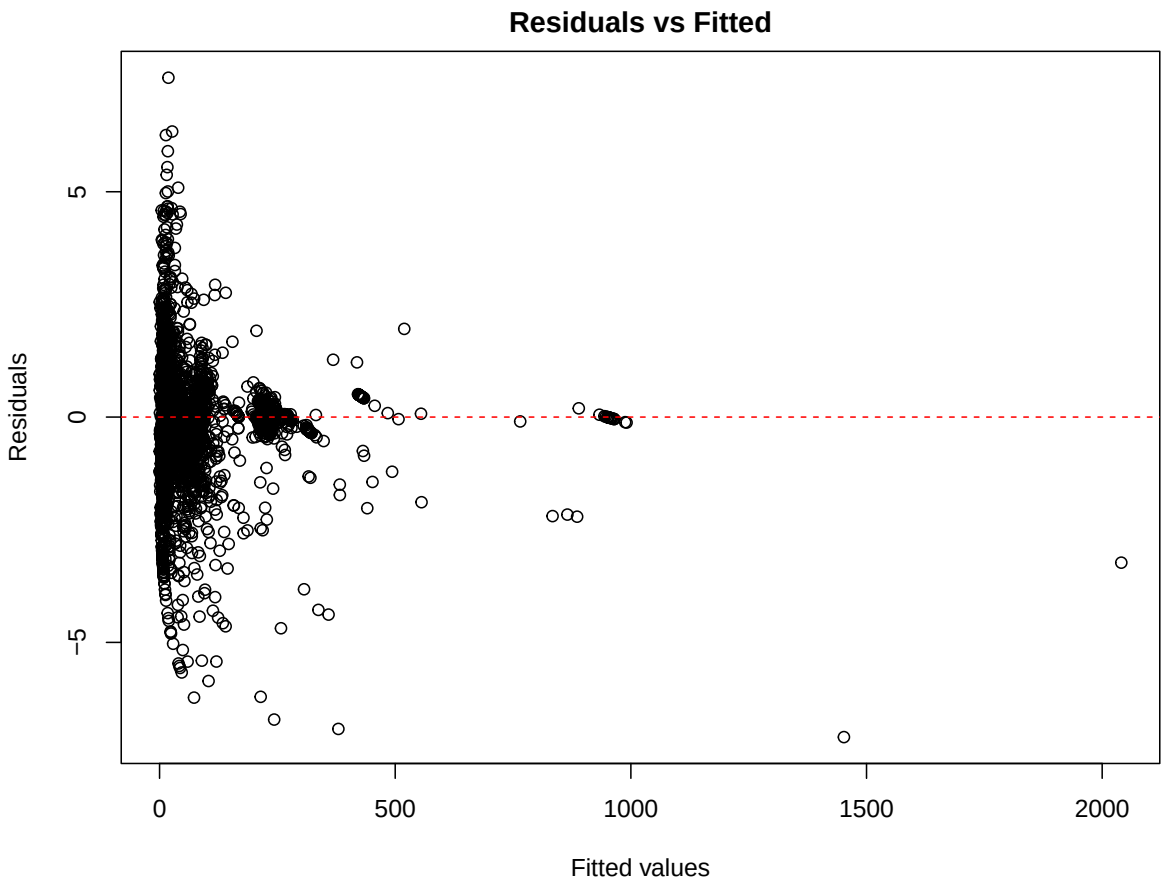

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−5
0
5
−4
−2
0
2
4

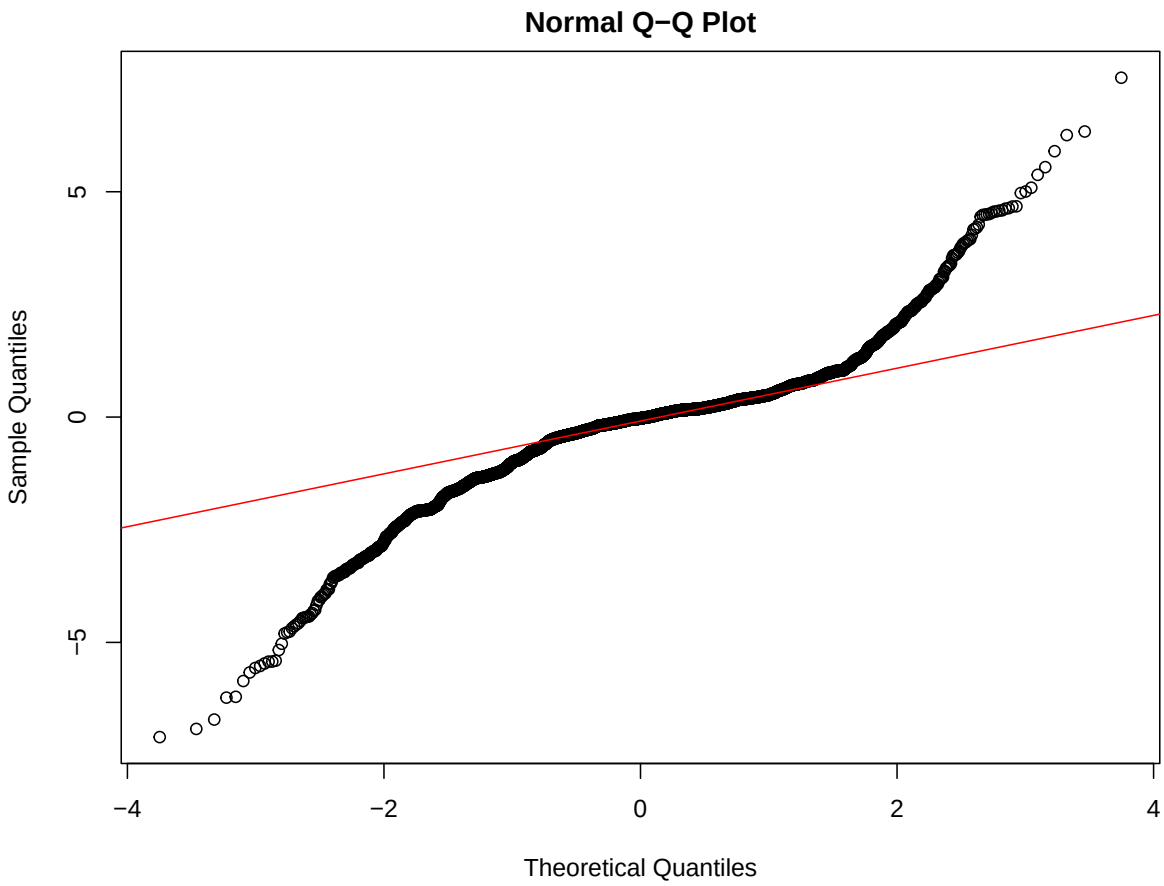

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
Fitted values

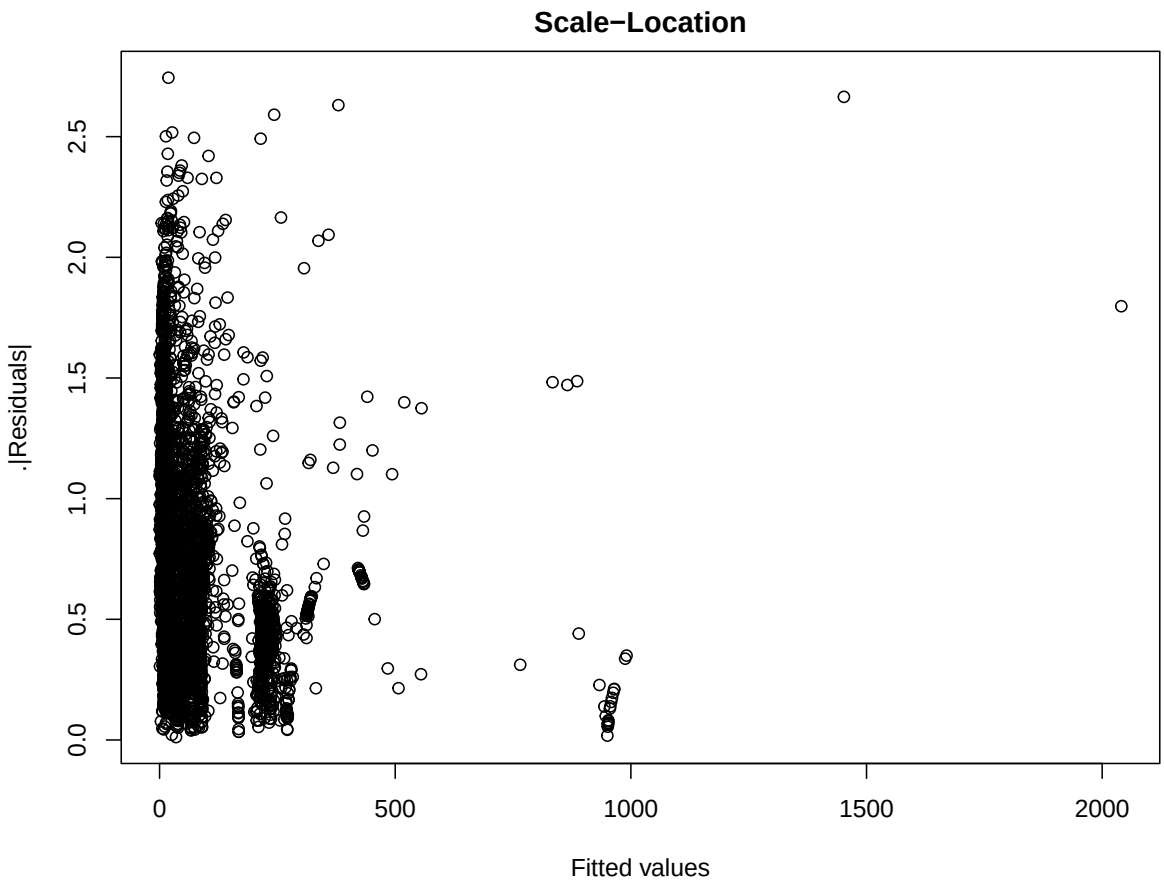


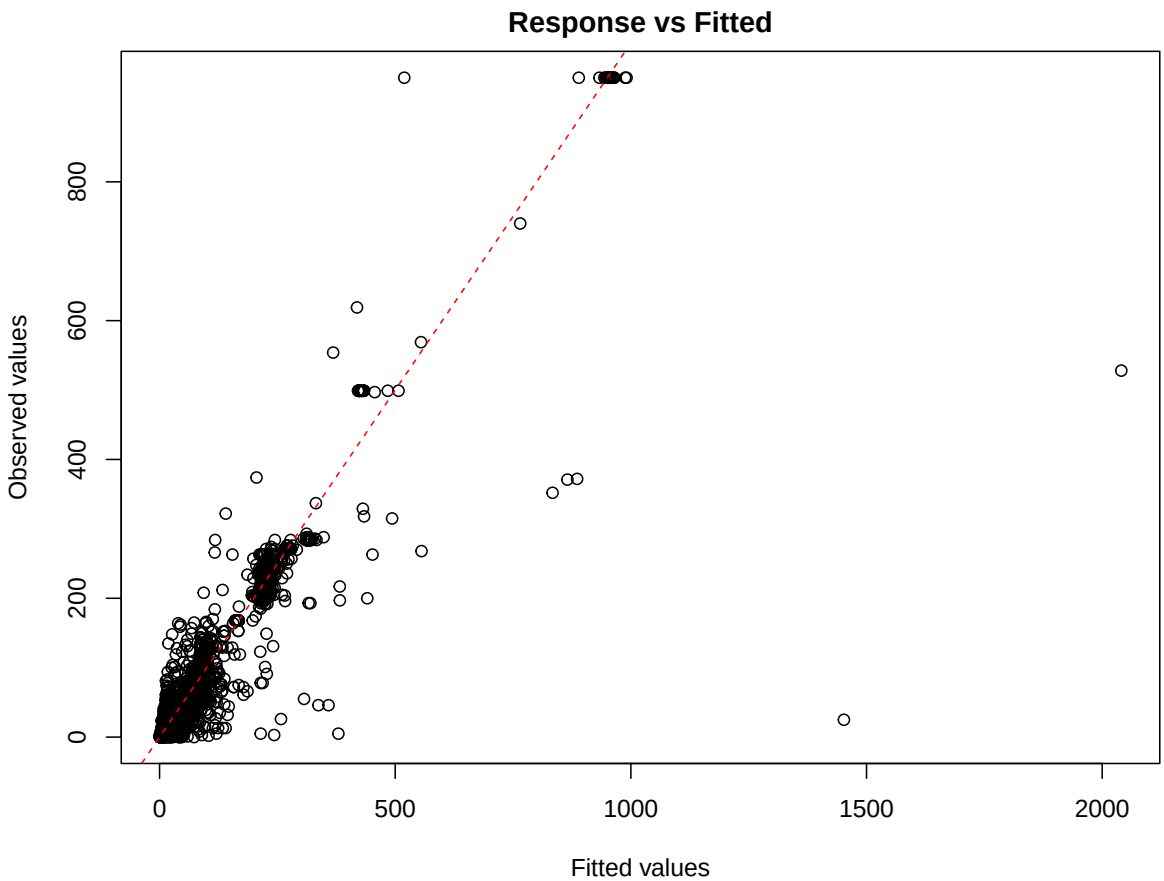

```
### Model 15 : 6. Global model

Model data dimensions after removing NAs: 5618 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(8.607)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + ti(mean_wind_speed, sunlight_proportion) + ti(mean_temp,
    sunlight_proportion) + s(view_id, bs = "re")

Parametric coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3493     0.1681   19.93   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                          edf Ref.df   Chi.sq  p-value
s(abundance_t_minus_1)                  8.924  8.997 33403.51  < 2e-16 ***
s(mean_wind_speed)                      1.002  1.004    28.48  < 2e-16 ***
s(mean_temp)                            8.356  8.807   194.60  < 2e-16 ***
s(sunlight_proportion)                  7.787  8.607    71.41  < 2e-16 ***
ti(mean_wind_speed,mean_temp)           9.102 10.479    82.46  < 2e-16 ***
ti(mean_wind_speed,sunlight_proportion) 9.915 11.977    35.62 0.000331 ***
ti(mean_temp,sunlight_proportion)      10.299 12.227   122.91  < 2e-16 ***
s(view_id)                              7.731  8.000   158.83  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.849   Deviance explained = 89.3%
-REML =  22707  Scale est. = 1         n = 5618


Diagnostic plots:
```


Residuals vs Fitted
Residuals
5
0
−5
0
500
1000
1500
2000
Fitted values


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
5
0
−5
−4
−2
0
2
4


Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
Fitted values




```r
# Model comparison table
comparison_60m <- data.frame(
  Model = names(aic_values_60m),
  AIC = round(aic_values_60m, 2)
) %>%
  arrange(AIC)

cat("\n### Model Comparison (1-hour lag)\n")
```

```
### Model Comparison (1-hour lag)
```

```r
kable(comparison_60m)
```

	Model	AIC
6. Global model	6. Global model	45226.15
5b. Wind x Temp + Temp x Sun	5b. Wind x Temp + Temp x Sun	45242.48
5c. Wind x Sun + Temp x Sun	5c. Wind x Sun + Temp x Sun	45279.34
4c. Temp x Sunlight interaction	4c. Temp x Sunlight interaction	45301.73
5a. Wind x Temp + Wind x Sun	5a. Wind x Temp + Wind x Sun	45322.71
4a. Wind x Temp interaction	4a. Wind x Temp interaction	45356.24
4b. Wind x Sunlight interaction	4b. Wind x Sunlight interaction	45364.55
3d. All additive	3d. All additive	45393.15
3c. Temp + Sunlight	3c. Temp + Sunlight	45423.55
3a. Wind + Temp	3a. Wind + Temp	45651.52
2b. Temperature only	2b. Temperature only	45670.08
3b. Wind + Sunlight	3b. Wind + Sunlight	45692.15
2c. Sunlight only	2c. Sunlight only	45722.21
2a. Wind only	2a. Wind only	45916.93
1. Baseline	1. Baseline	45936.70

2-Hour Lag Analysis

```r
data_120m <- prepare_lag_data(master_df, wind, lag_minutes = 120)
```

```
Preparing data for 120 minute lag...
Valid pairs after filtering: 6912
Calculating interval predictors efficiently...
Processing 6912 unique intervals...
Progress: 10 %
Progress: 20 %
Progress: 30 %
Progress: 40 %
Progress: 50 %
Progress: 60 %
Progress: 70 %
Progress: 80 %
Progress: 90 %
Progress: 100 %
Combining results...
Final dataset rows: 5650
Time delta range: 120 to 120 minutes
```

Exploratory Data Analysis (2 hours)

```r
cat("Data structure:\n")
```

```
Data structure:
```

```r
glimpse(data_120m)
```

```
Rows: 5,650
Columns: 30
$ deployment_id.x           <chr> "SC1", "SC1", "SC1", "SC1", "SC1", "SC1", "S~
$ image_filename            <chr> "SC1_20231117184501.JPG", "SC1_2023111718500~
$ total_butterflies         <dbl> 54, 54, 54, 54, 54, 54, 54, 54, 54, 54, 54, ~
$ butterflies_direct_sun    <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ datetime                  <dttm> 2023-11-17 18:45:01, 2023-11-17 18:50:01, 2~
$ view_id                   <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ wind_meter_name           <chr> "JazzPlay", "JazzPlay", "JazzPlay", "JazzPla~
$ deployment_id.y           <chr> "SC1", "SC1", "SC1", "SC1", "SC1", "SC1", "S~
$ timestamp                 <dbl> 2.023112e+13, 2.023112e+13, 2.023112e+13, 2.~
$ temperature               <dbl> 15, 16, 15, 16, 15, 16, 16, 15, 16, 15, 16, ~
$ confidence                <dbl> 0.4041873, 0.5208792, 0.6782328, 0.5311629, ~
$ extraction_status         <chr> "success", "success", "success", "success", ~
$ datetime_t_minus_1        <dttm> 2023-11-17 16:45:01, 2023-11-17 16:50:01, 2~
$ abundance_t_minus_1       <dbl> 39, 41, 39, 68, 73, 72, 61, 49, 57, 51, 54, ~
$ butterflies_sun_t_minus_1 <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ temperature_t_minus_1     <dbl> 15, 15, 15, 15, 15, 15, 16, 15, 15, 15, 15, ~
$ time_delta_mins           <dbl> 120, 120, 120, 120, 120, 120, 120, 120, 120,~
$ mean_wind_speed           <dbl> 0.400000, 0.900000, 1.030769, 1.177778, 1.24~
$ max_wind_speed            <dbl> 0.4, 1.9, 1.9, 1.9, 1.9, 2.3, 2.7, 2.9, 2.9,~
$ sd_wind_speed             <dbl> 0.0000000, 0.6524678, 0.5297556, 0.5241863, ~
$ gust_factor               <dbl> 2.250000, 1.722222, 1.656716, 1.622642, 1.58~
$ minutes_above_2mps        <dbl> 0, 0, 0, 0, 0, 1, 3, 6, 7, 12, 16, 20, 25, 3~
$ mean_temp                 <dbl> 15.04, 15.08, 15.08, 15.12, 15.12, 15.16, 15~
$ max_temp                  <dbl> 16, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16, ~
$ min_temp                  <dbl> 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, ~
$ sd_temp                   <dbl> 0.2000000, 0.2768875, 0.2768875, 0.3316625, ~
$ sunlight_proportion       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ max_prop_sunlight         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ min_prop_sunlight         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ sd_prop_sunlight          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
```

```r
cat("\n\nData summary:\n")
```

```
Data summary:
```

```r
summary(data_120m)
```

```
 deployment_id.x    image_filename     total_butterflies butterflies_direct_sun
 Length:5650        Length:5650        Min.   :  0.00    Min.   :  0.000
 Class :character   Class :character   1st Qu.: 13.00    1st Qu.:  0.000
 Mode  :character   Mode  :character   Median : 34.00    Median :  0.000
                                       Mean   : 65.33    Mean   :  1.647
                                       3rd Qu.: 72.00    3rd Qu.:  0.000
                                       Max.   :950.00    Max.   :150.000

    datetime                          view_id     wind_meter_name
 Min.   :2023-11-17 18:45:01.00   4      :2921   Length:5650
 1st Qu.:2023-12-11 03:07:31.00   2      :1080   Class :character
 Median :2023-12-29 07:50:01.00   1      : 662   Mode  :character
 Mean   :2023-12-29 03:20:27.64   5      : 460
 3rd Qu.:2024-01-23 08:12:31.00   11     : 287
 Max.   :2024-02-04 10:50:01.00   9      : 171
                                  (Other):  69
 deployment_id.y      timestamp          temperature      confidence
 Length:5650        Min.   :2.023e+13   Min.   : 4.00   Min.   :0.1222
 Class :character   1st Qu.:2.023e+13   1st Qu.:11.00   1st Qu.:0.4967
 Mode  :character   Median :2.023e+13   Median :13.00   Median :0.5934
                    Mean   :2.024e+13   Mean   :13.26   Mean   :0.5960
                    3rd Qu.:2.024e+13   3rd Qu.:15.00   3rd Qu.:0.6951
                    Max.   :2.024e+13   Max.   :33.00   Max.   :1.0000

 extraction_status  datetime_t_minus_1               abundance_t_minus_1
 Length:5650        Min.   :2023-11-17 16:45:01.00   Min.   :  0.00
 Class :character   1st Qu.:2023-12-11 01:07:31.00   1st Qu.: 14.00
 Mode  :character   Median :2023-12-29 05:50:01.00   Median : 35.00
                    Mean   :2023-12-29 01:20:27.64   Mean   : 65.93
                    3rd Qu.:2024-01-23 06:12:31.00   3rd Qu.: 73.75
                    Max.   :2024-02-04 08:50:01.00   Max.   :950.00

 butterflies_sun_t_minus_1 temperature_t_minus_1 time_delta_mins
```

```
 Min.   :  0.000           Min.   : 4.00         Min.   :120
 1st Qu.:  0.000           1st Qu.:11.00         1st Qu.:120
 Median :  0.000           Median :13.00         Median :120
 Mean   :  1.707           Mean   :13.25         Mean   :120
 3rd Qu.:  0.000           3rd Qu.:15.00         3rd Qu.:120
 Max.   :150.000           Max.   :30.00         Max.   :120

 mean_wind_speed  max_wind_speed sd_wind_speed     gust_factor
 Min.   :0.0000   Min.   :0.00   Min.   :0.0000   Min.   :1.181
 1st Qu.:0.2429   1st Qu.:1.00   1st Qu.:0.2049   1st Qu.:1.357
 Median :0.7521   Median :1.50   Median :0.3170   Median :1.447
 Mean   :0.8837   Mean   :1.71   Mean   :0.3288   Mean   :  Inf
 3rd Qu.:1.3033   3rd Qu.:2.30   3rd Qu.:0.4395   3rd Qu.:1.571
 Max.   :5.0092   Max.   :6.70   Max.   :1.2932   Max.   :  Inf
                                 NA's   :1        NA's   :498
 minutes_above_2mps   mean_temp        max_temp        min_temp
 Min.   :  0.00     Min.   : 4.20   Min.   : 5.00   Min.   : 4.00
 1st Qu.:  0.00     1st Qu.:11.00   1st Qu.:12.00   1st Qu.:10.00
 Median :  0.00     Median :13.00   Median :14.00   Median :12.00
 Mean   : 11.58     Mean   :13.26   Mean   :14.22   Mean   :12.35
 3rd Qu.:  4.00     3rd Qu.:15.48   3rd Qu.:16.00   3rd Qu.:15.00
 Max.   :120.00     Max.   :29.85   Max.   :33.00   Max.   :28.00

    sd_temp       sunlight_proportion max_prop_sunlight min_prop_sunlight
 Min.   :0.0000   Min.   :0.00000     Min.   :0.00000   Min.   :0.00000
 1st Qu.:0.4472   1st Qu.:0.00000     1st Qu.:0.00000   1st Qu.:0.00000
 Median :0.5477   Median :0.00000     Median :0.00000   Median :0.00000
 Mean   :0.7375   Mean   :0.07285     Mean   :0.15567   Mean   :0.03173
 3rd Qu.:0.8944   3rd Qu.:0.01923     3rd Qu.:0.07868   3rd Qu.:0.00000
 Max.   :6.8775   Max.   :1.00000     Max.   :1.00000   Max.   :1.00000

 sd_prop_sunlight
 Min.   :0.00000
 1st Qu.:0.00000
 Median :0.00000
 Mean   :0.04904
 3rd Qu.:0.00221
 Max.   :0.70711
 NA's   :37
```

```r
# Check for data quality issues
cat("\n\nData quality checks:\n")
```

```
Data quality checks:
```

```r
cat("Total rows:", nrow(data_120m), "\n")
```

```
Total rows: 5650
```

```r
cat("Complete cases:", sum(complete.cases(data_120m)), "\n")
```

```
Complete cases: 5122
```

```r
cat("Rows with abundance_t_minus_1 > 0:", sum(data_120m$abundance_t_minus_1 > 0, na.rm = TRU
```

```
Rows with abundance_t_minus_1 > 0: 5552
```

```r
cat("Rows with total_butterflies > 0:", sum(data_120m$total_butterflies > 0, na.rm = TRUE), "
```

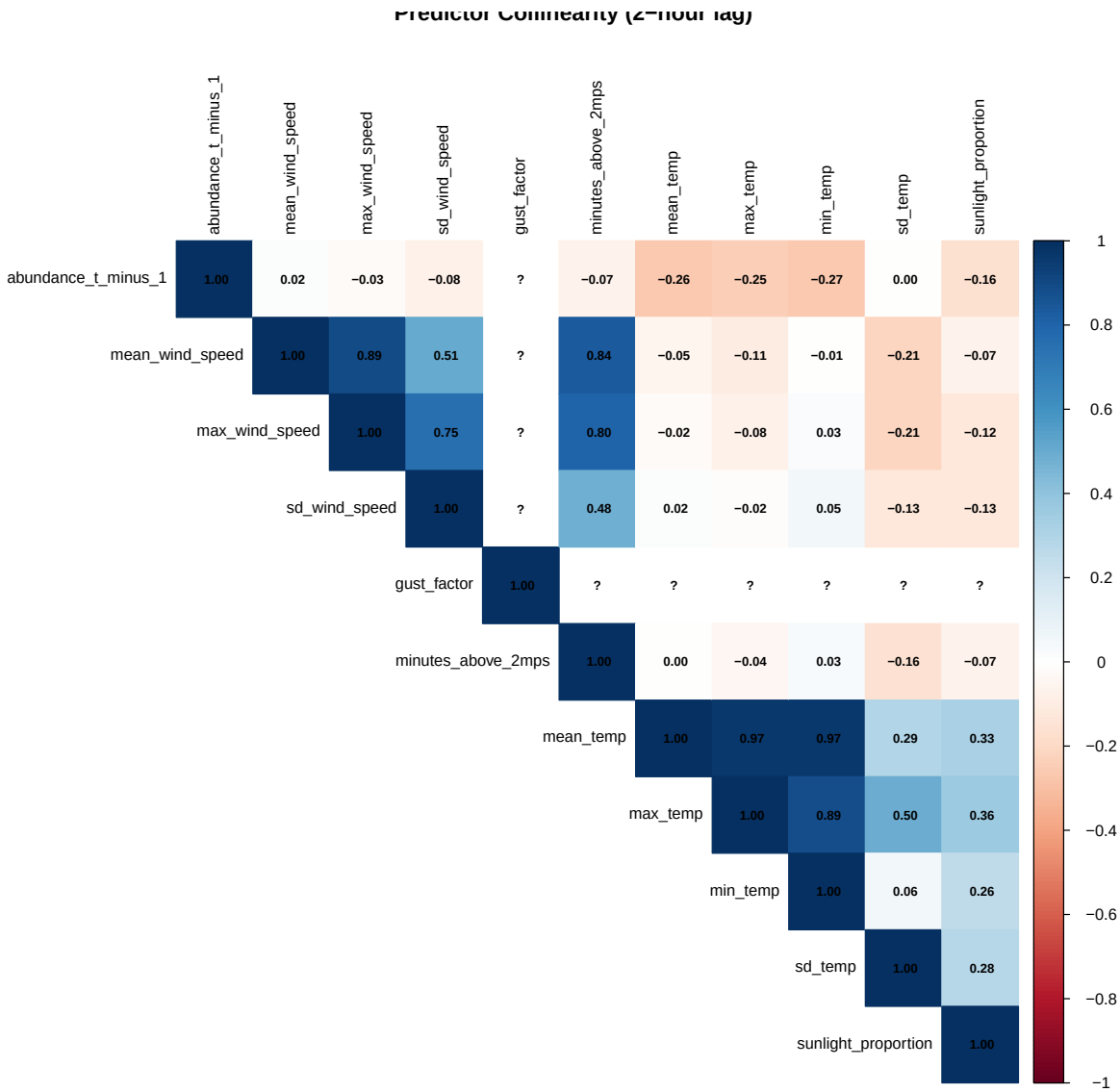
```
Rows with total_butterflies > 0: 5499
```

```r
cat("Unique view_ids:", length(unique(data_120m$view_id)), "\n")
```

```
Unique view_ids: 9
```

```r
# Select numeric predictors for correlation analysis
predictors_120m <- data_120m %>%
  select(abundance_t_minus_1, mean_wind_speed, max_wind_speed, sd_wind_speed,
         gust_factor, minutes_above_2mps, mean_temp, max_temp, min_temp, sd_temp,
         sunlight_proportion) %>%
  na.omit()

if (nrow(predictors_120m) > 1) {
  cor_matrix_120m <- cor(predictors_120m, use = "complete.obs")
  if (!any(is.na(cor_matrix_120m) | is.infinite(cor_matrix_120m))) {
    corrplot(cor_matrix_120m, method = "color", type = "upper",
             order = "hclust", tl.cex = 0.8, tl.col = "black",
             addCoef.col = "black", number.cex = 0.7,
             title = "Predictor Collinearity (2-hour lag)")
```

```r
  } else {
    corrplot(cor_matrix_120m, method = "color", type = "upper",
             order = "original", tl.cex = 0.8, tl.col = "black",
             addCoef.col = "black", number.cex = 0.7,
             title = "Predictor Collinearity (2-hour lag)")
  }
} else {
  cat("Not enough data for correlation analysis\n")
}
```



Model Fitting (2 hours)

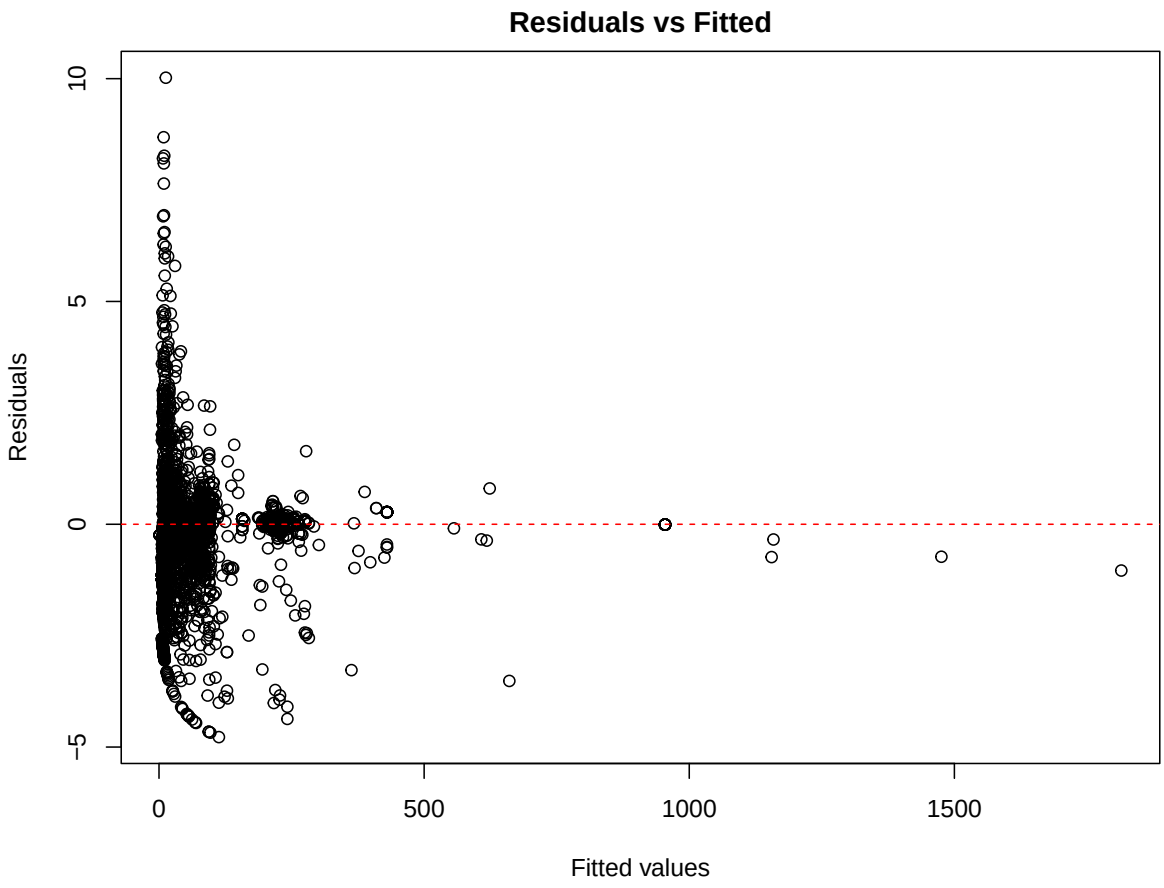
```
# Define comprehensive model formulas
model_formulas_120m <- list(
  # 1. Baseline Model
  "1. Baseline" = total_butterflies ~ s(abundance_t_minus_1) + s(view_id, bs = "re"),
```

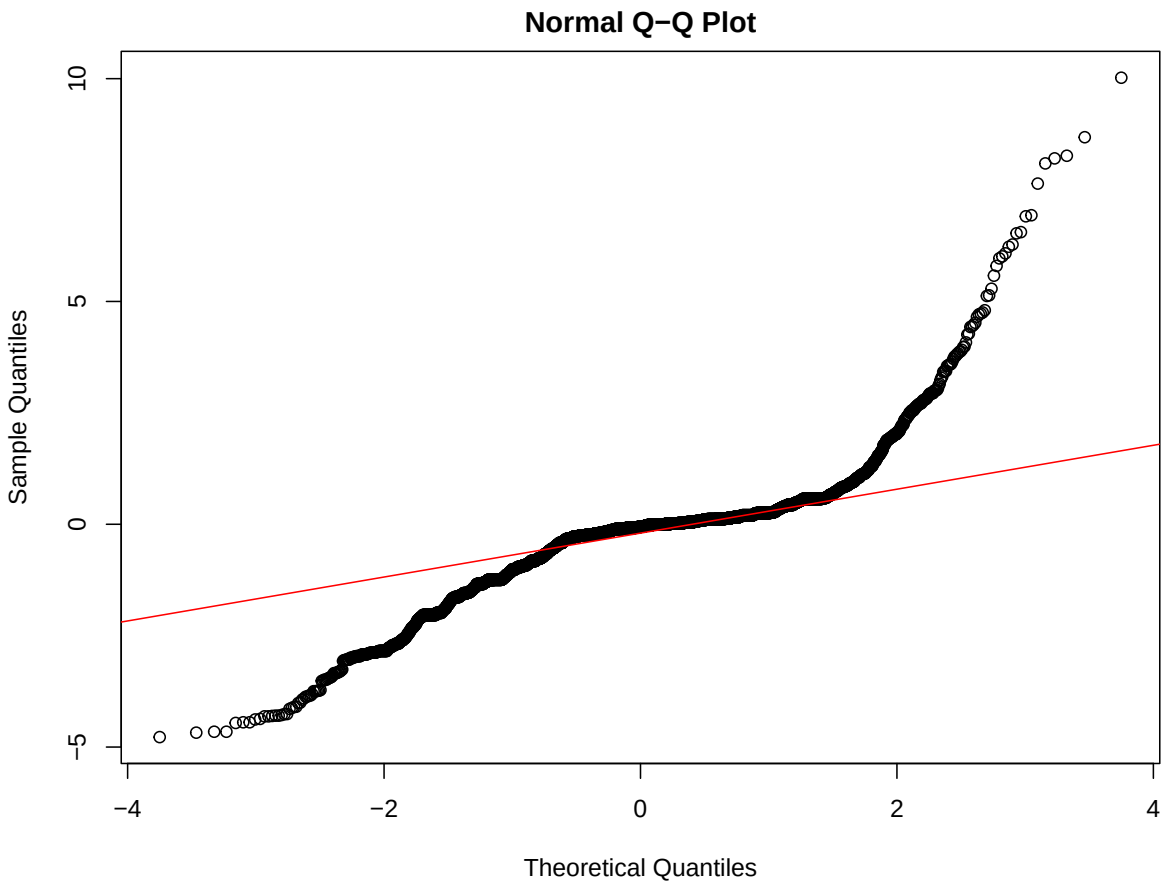
```
  # 2. Single Environmental Factor Models
  "2a. Wind only" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(view_
  "2b. Temperature only" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(vie
  "2c. Sunlight only" = total_butterflies ~ s(abundance_t_minus_1) + s(sunlight_proportion) +

  # 3. Additive Models (No Interactions)
  "3a. Wind + Temp" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(mea
  "3b. Wind + Sunlight" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s
  "3c. Temp + Sunlight" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(sunli
  "3d. All additive" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(me

  # 4. Single Interaction Models
  "4a. Wind x Temp interaction" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_sp
  "4b. Wind x Sunlight interaction" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wi
  "4c. Temp x Sunlight interaction" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wi

  # 5. Multiple Interaction Models
  "5a. Wind x Temp + Wind x Sun" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_s
  "5b. Wind x Temp + Temp x Sun" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_s
  "5c. Wind x Sun + Temp x Sun" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_sp

  # 6. Global Model with All Interactions
  "6. Global model" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(mea
)

# Storage for results
models_120m <- list()
aic_values_120m <- numeric()

# Fit each model
for (i in seq_along(model_formulas_120m)) {
  model_name <- names(model_formulas_120m)[i]
  formula <- model_formulas_120m[[i]]

  cat("\n### Model", i, ":", model_name, "\n\n")

  tryCatch({
    # Check data before fitting
    model_vars <- all.vars(formula)
    model_data <- data_120m %>%
      select(all_of(model_vars), view_id) %>%
      na.omit()
```

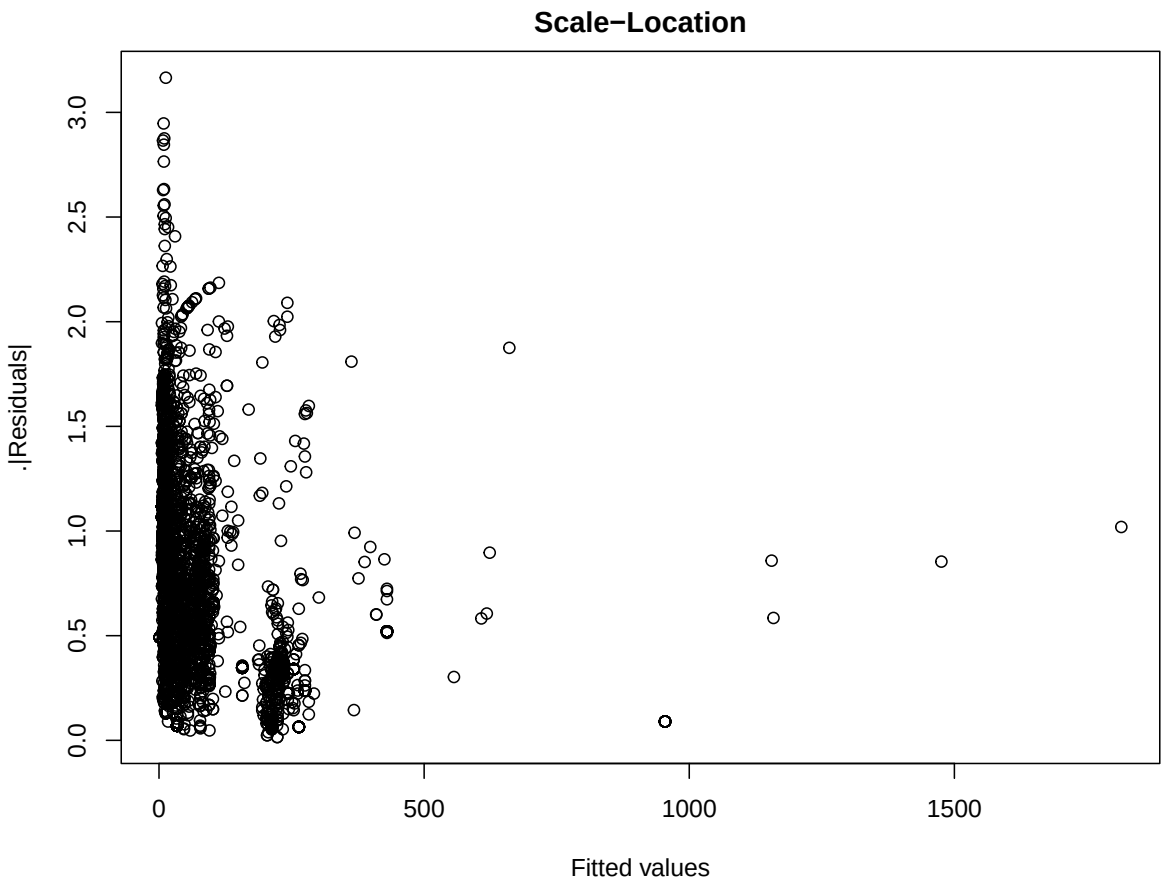
```r
cat("Model data dimensions after removing NAs:", nrow(model_data), "x", ncol(model_data)

# Fit GAM
gam_model <- gam(
  formula,
  family = nb(),
  data = model_data
)

cat("GAM fitted successfully\n")

# Store model
models_120m[[model_name]] <- gam_model
aic_values_120m[model_name] <- AIC(gam_model)

# Print summary
cat("Model Summary:\n")
print(summary(gam_model))

# Diagnostic plots - individual plots for better visibility in PDF
cat("\n\nDiagnostic plots:\n")

# Set up for individual plots
par(mfrow = c(1, 1), mar = c(4, 4, 2, 1))

# Plot 1: Residuals vs Fitted
plot(fitted(gam_model), residuals(gam_model),
     xlab = "Fitted values", ylab = "Residuals",
     main = "Residuals vs Fitted")
abline(h = 0, col = "red", lty = 2)

# Plot 2: Q-Q plot
qqnorm(residuals(gam_model))
qqline(residuals(gam_model), col = "red")

# Plot 3: Scale-Location (if appropriate)
plot(fitted(gam_model), sqrt(abs(residuals(gam_model))),
     xlab = "Fitted values", ylab = "√|Residuals|",
     main = "Scale-Location")

# Plot 4: Response vs Fitted
plot(fitted(gam_model), model_data$total_butterflies,
```

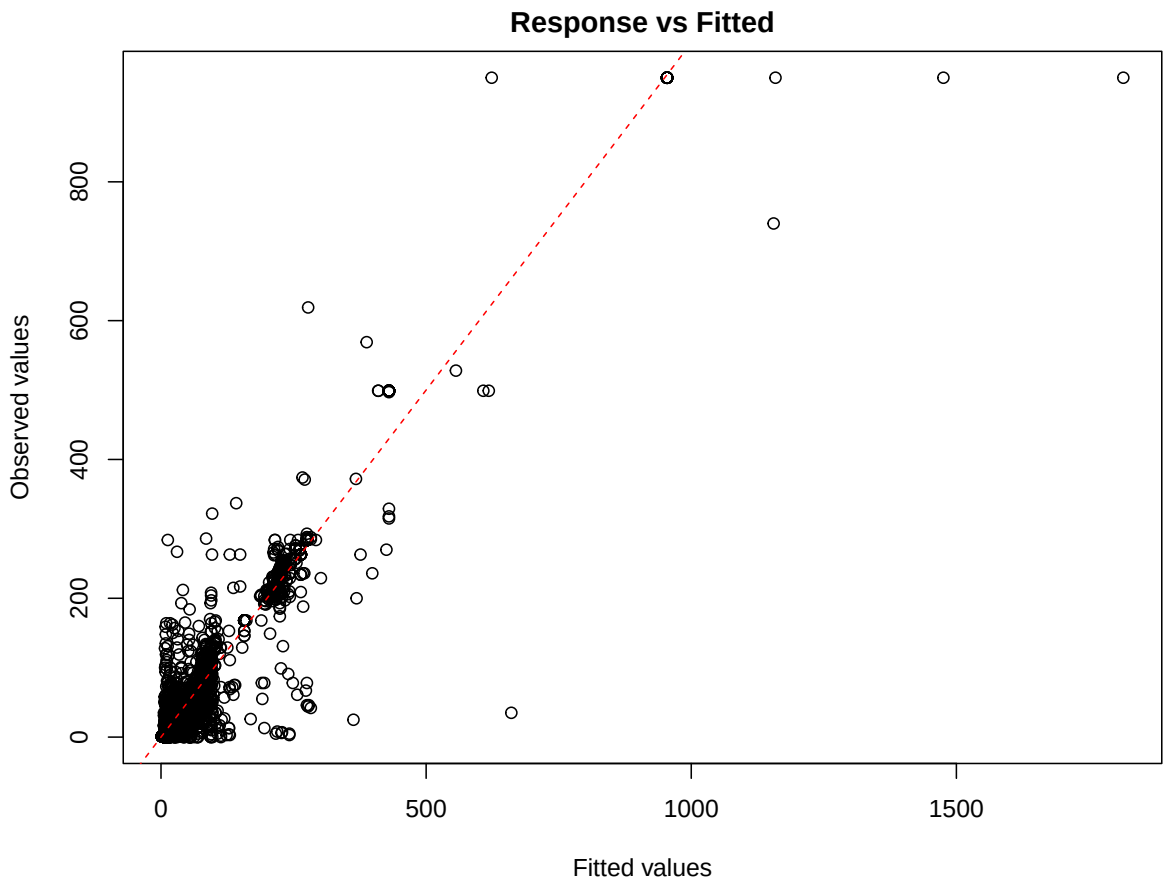
```r
        xlab = "Fitted values", ylab = "Observed values",
        main = "Response vs Fitted")
    abline(a = 0, b = 1, col = "red", lty = 2)

  }, error = function(e) {
    cat("Error fitting model:", e$message, "\n")
    aic_values_120m[model_name] <- NA
  })
}
```

```
### Model 1 : 1. Baseline

Model data dimensions after removing NAs: 5650 x 3
GAM fitted successfully
Model Summary:

Family: Negative Binomial(3.167)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.4360     0.2534   13.56   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                         edf Ref.df  Chi.sq p-value
s(abundance_t_minus_1) 8.740  8.973 16839.9  <2e-16 ***
s(view_id)             7.797  8.000   229.6  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.898   Deviance explained = 76.2%
-REML =  24694  Scale est. = 1         n = 5650


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
10
5
0
−5
0
500
1000
1500

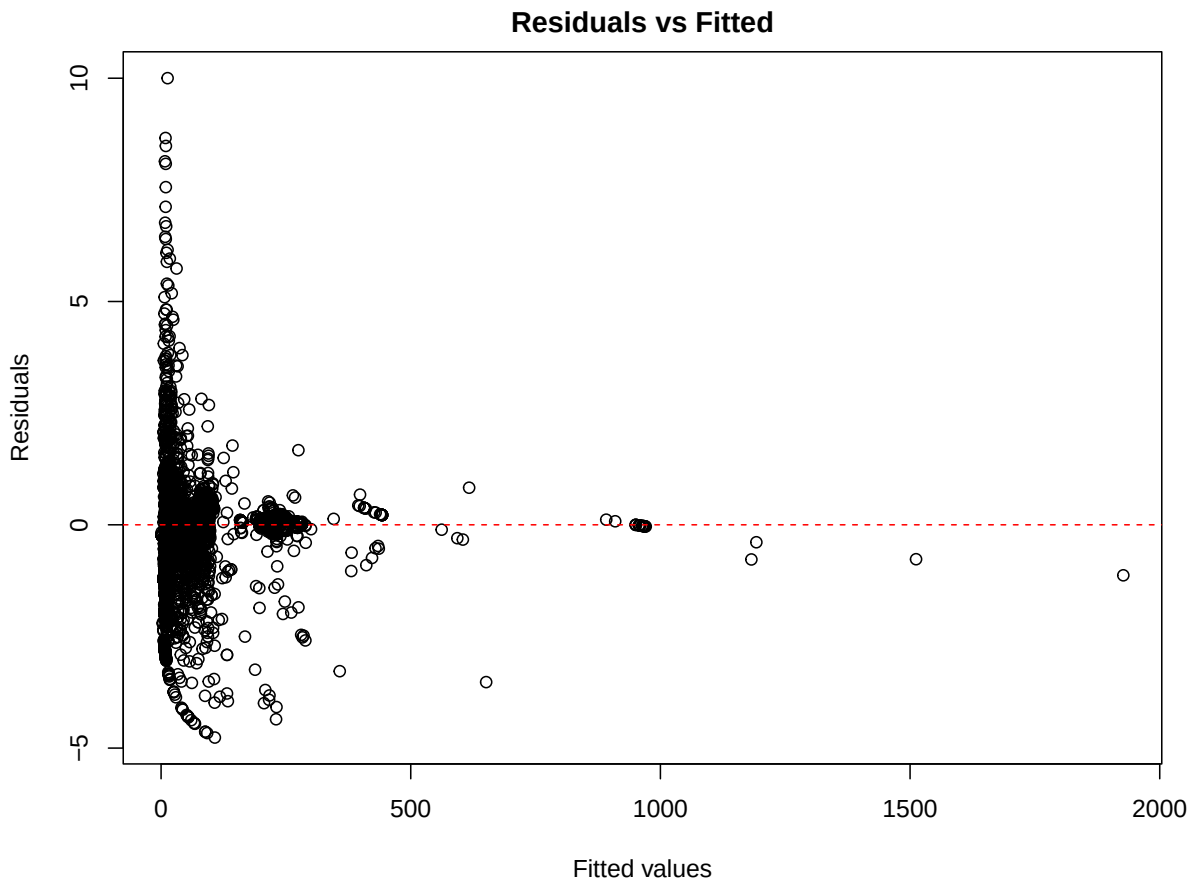

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4

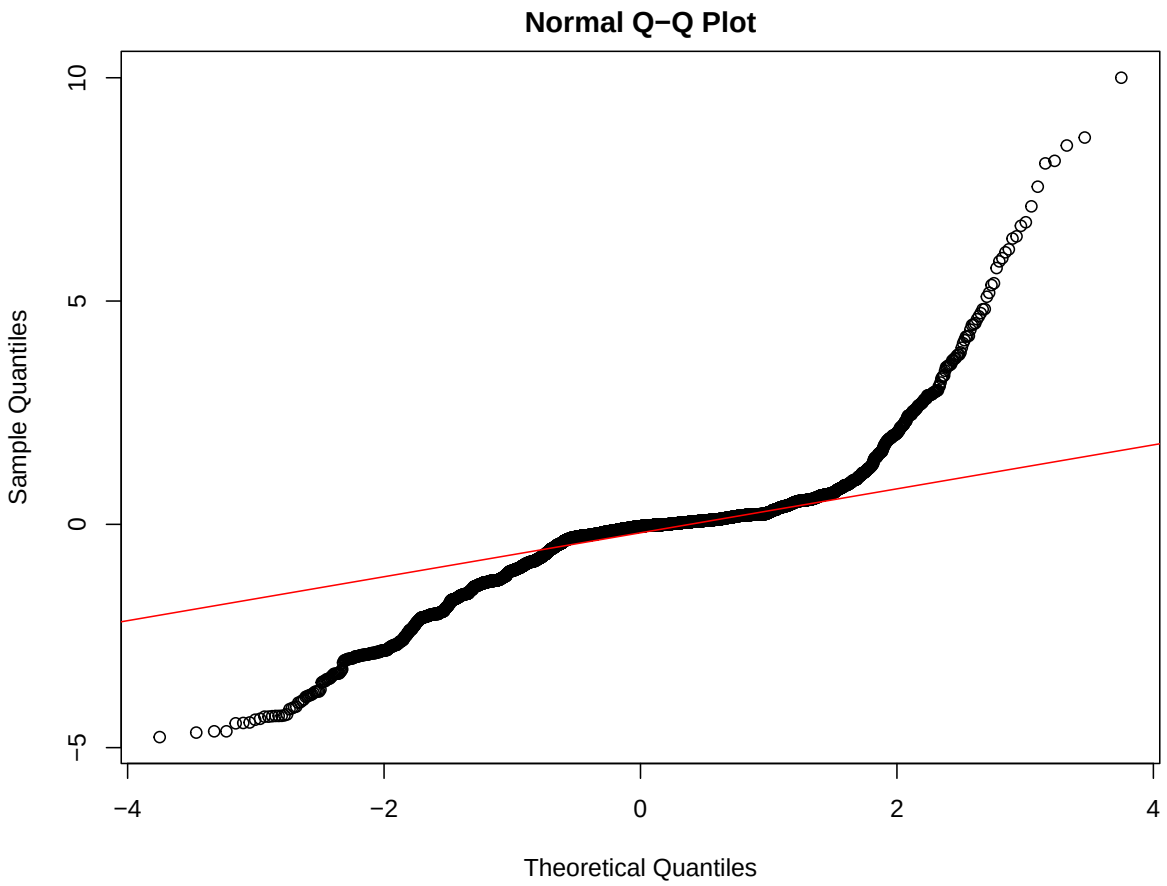

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
500
1000
1500
Fitted values

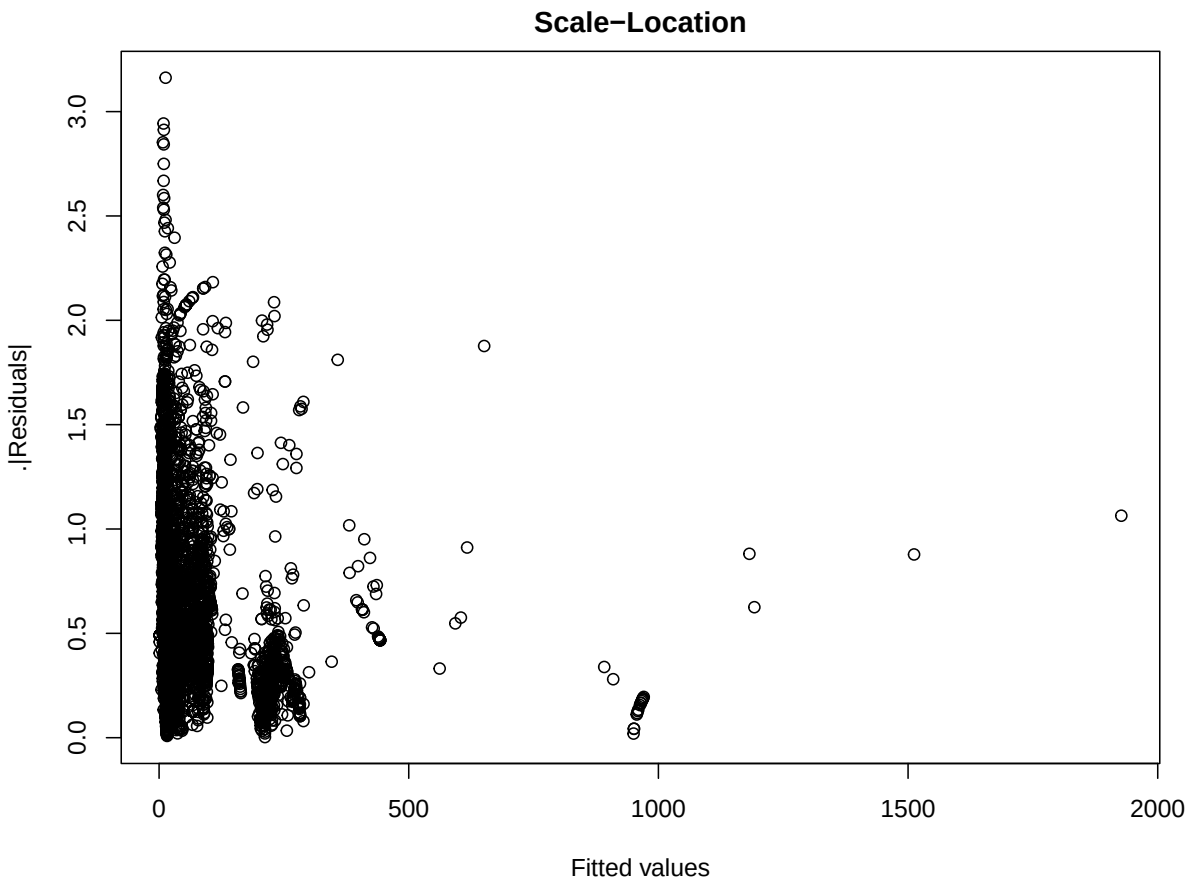


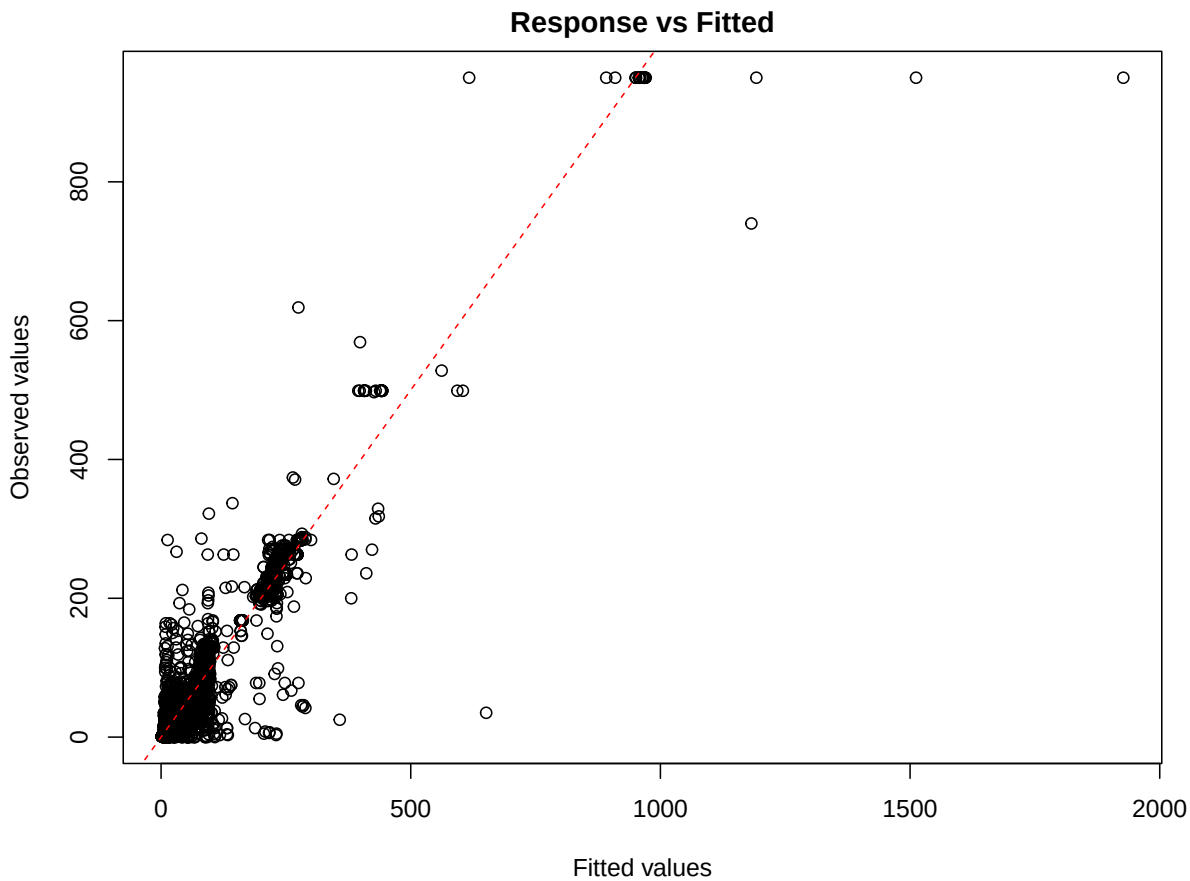

```
### Model 2 : 2a. Wind only

Model data dimensions after removing NAs: 5650 x 4
GAM fitted successfully
Model Summary:

Family: Negative Binomial(3.202)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    3.447      0.262   13.15   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df   Chi.sq p-value
s(abundance_t_minus_1) 8.749  8.975 16437.51  <2e-16 ***
s(mean_wind_speed)     5.286  6.428    53.94  <2e-16 ***
s(view_id)             7.808  8.000   230.52  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.893   Deviance explained = 76.4%
-REML =  24679  Scale est. = 1         n = 5650


Diagnostic plots:
```



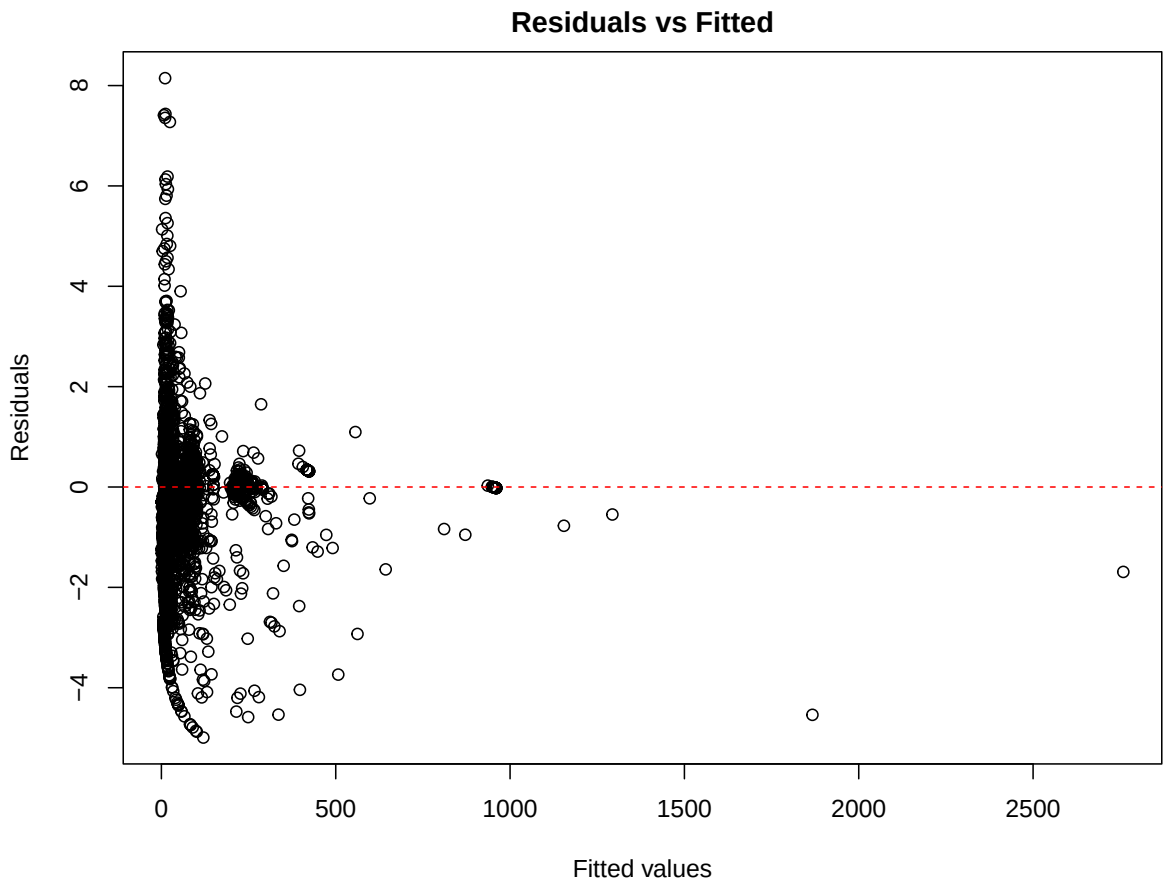

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4

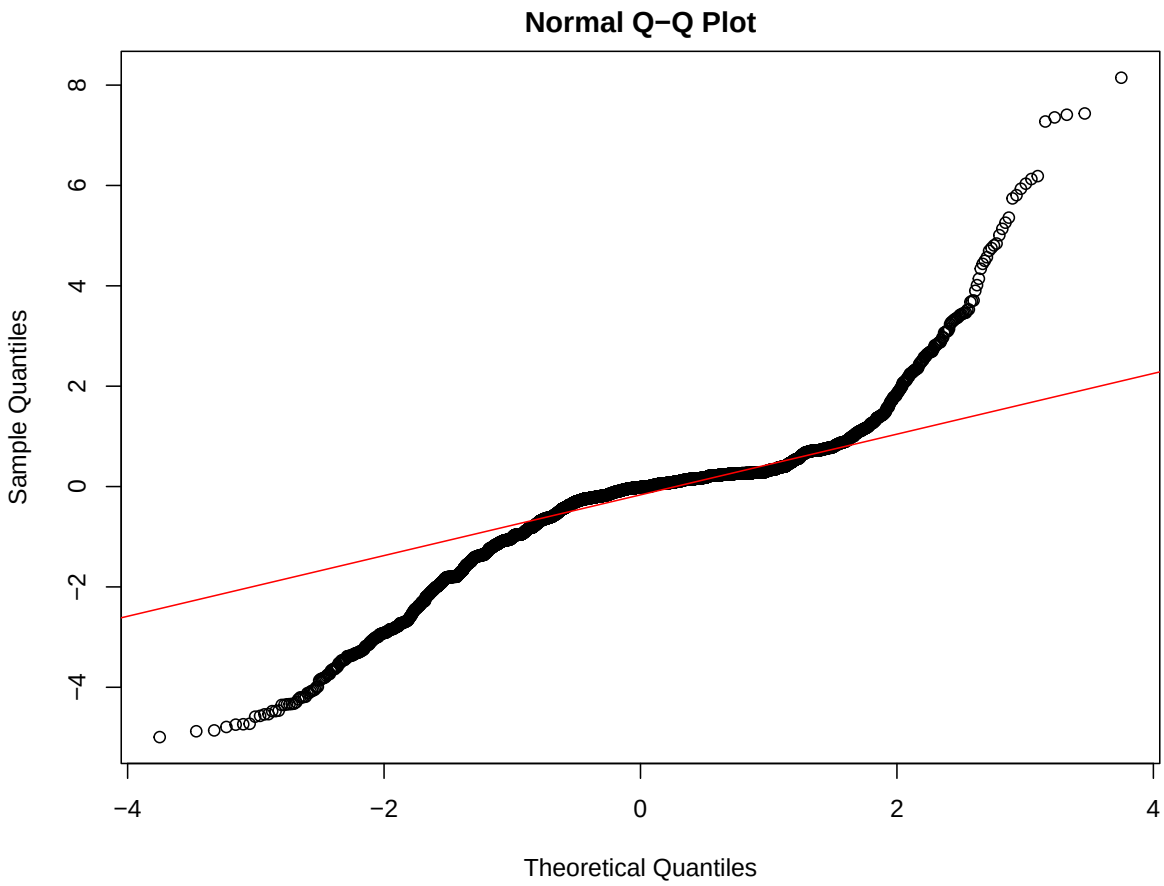

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
500
1000
1500
2000
Fitted values

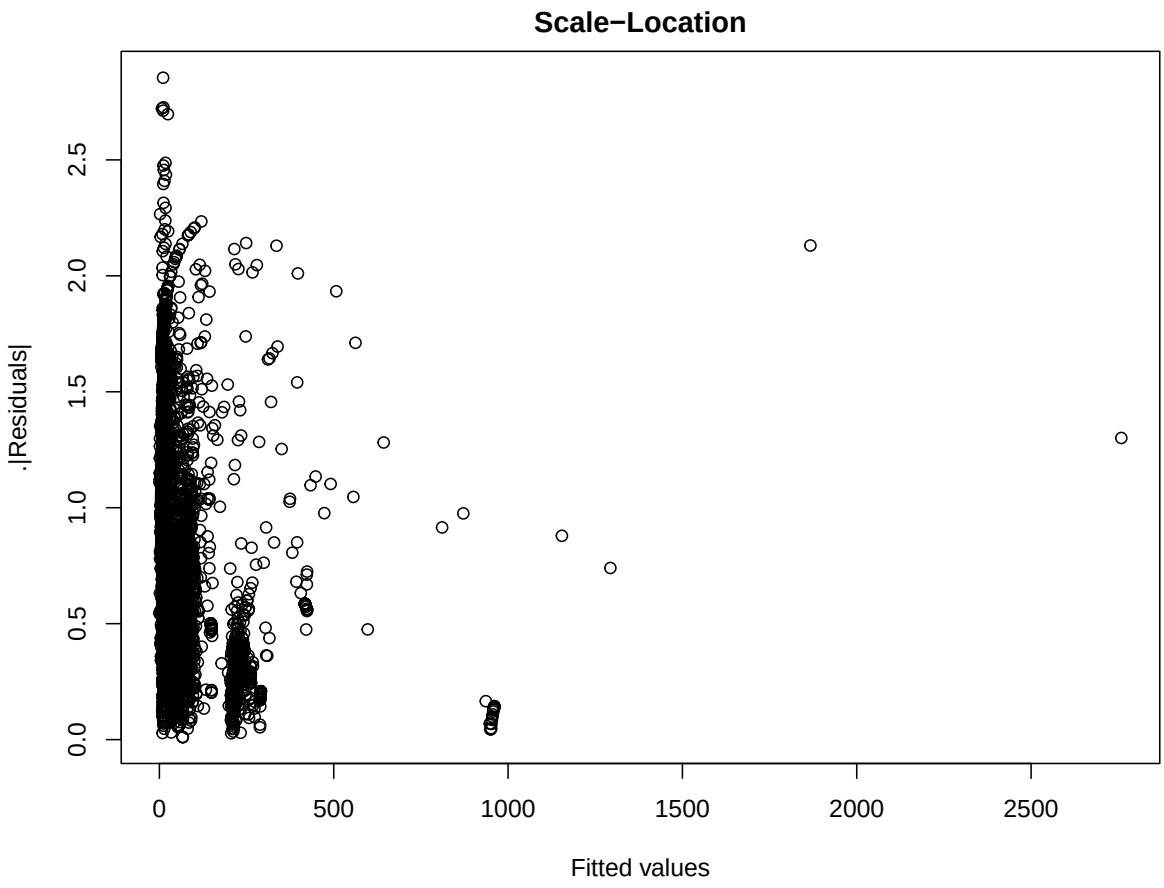


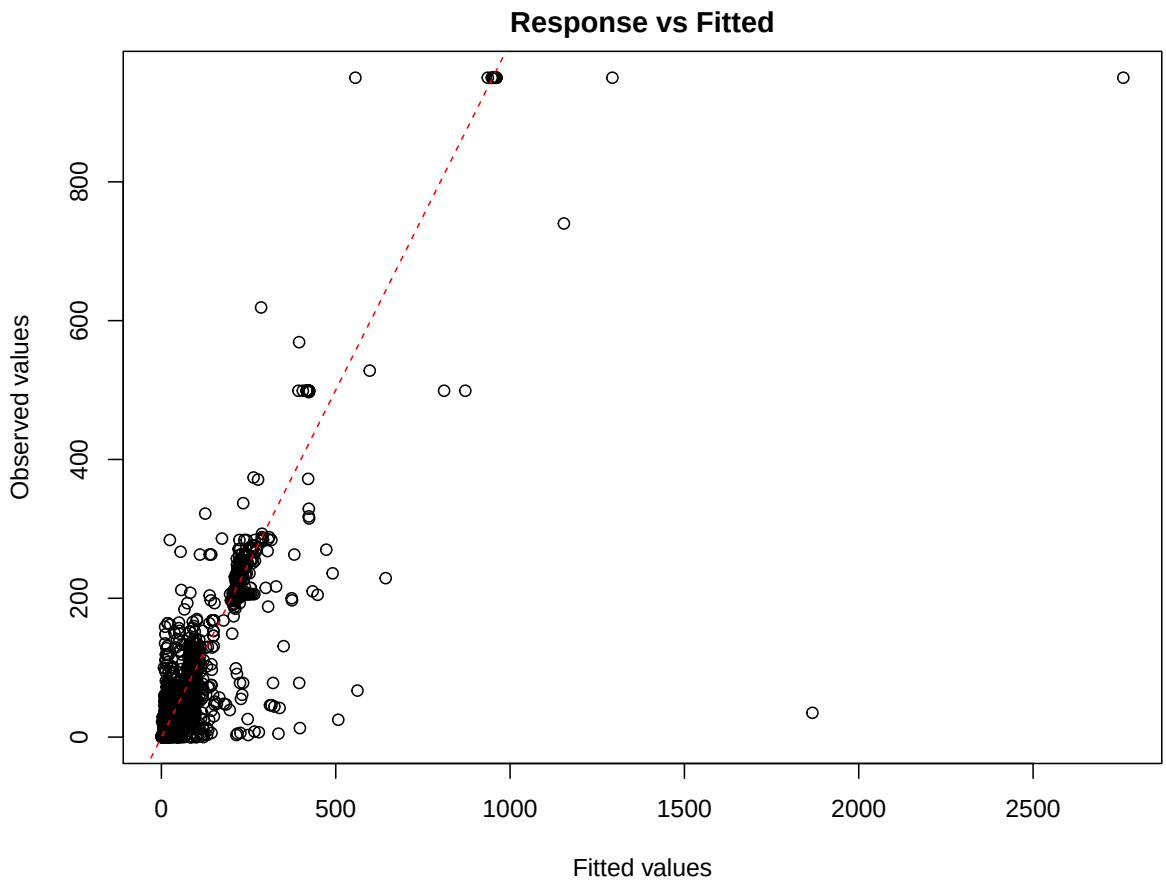

```
### Model 3 : 2b. Temperature only

Model data dimensions after removing NAs: 5650 x 4
GAM fitted successfully
Model Summary:

Family: Negative Binomial(3.495)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(view_id,
    bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3938     0.2034   16.69   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df  Chi.sq p-value
s(abundance_t_minus_1) 8.821  8.987 18535.7  <2e-16 ***
s(mean_temp)           8.053  8.715   531.0  <2e-16 ***
s(view_id)             7.713  8.000   293.3  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.766   Deviance explained = 78.3%
-REML =  24445  Scale est. = 1         n = 5650


Diagnostic plots:
```



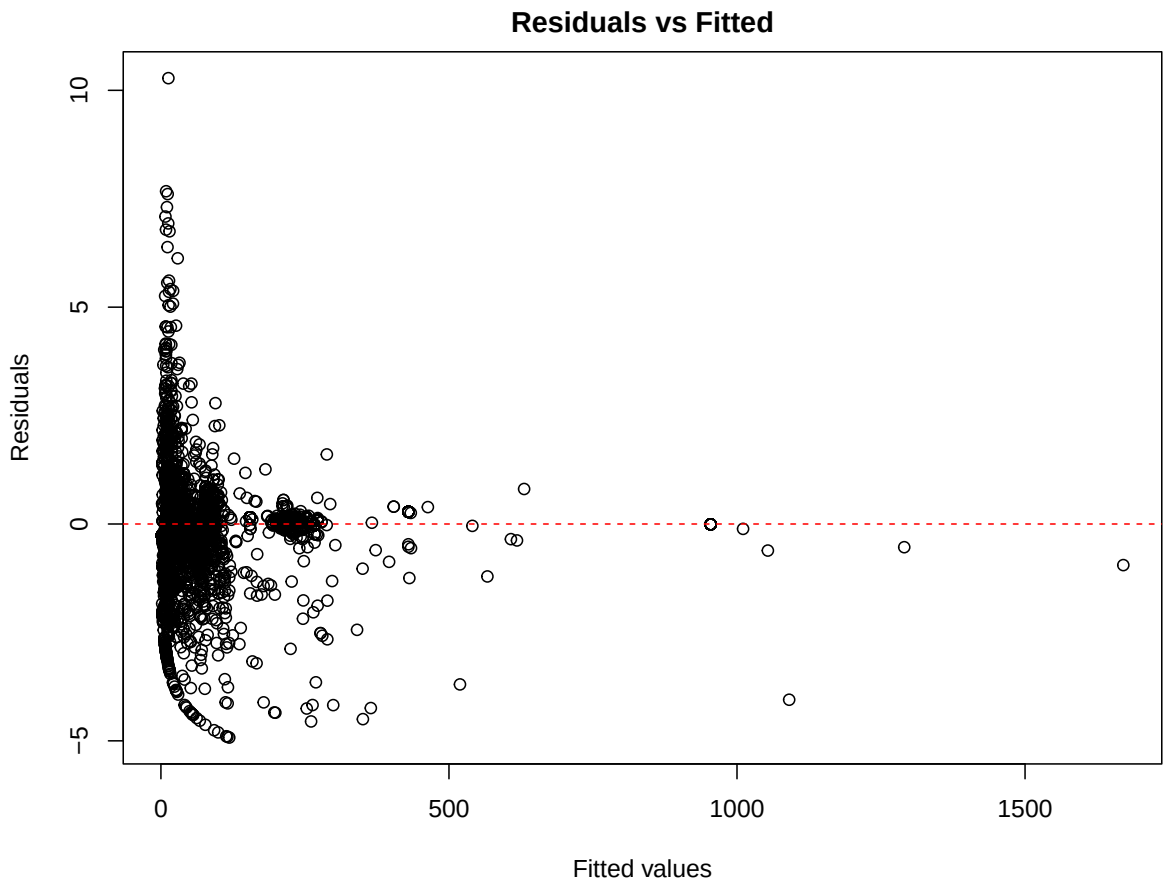

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
8
6
4
2
0
−2
−4
−4
−2
0
2
4

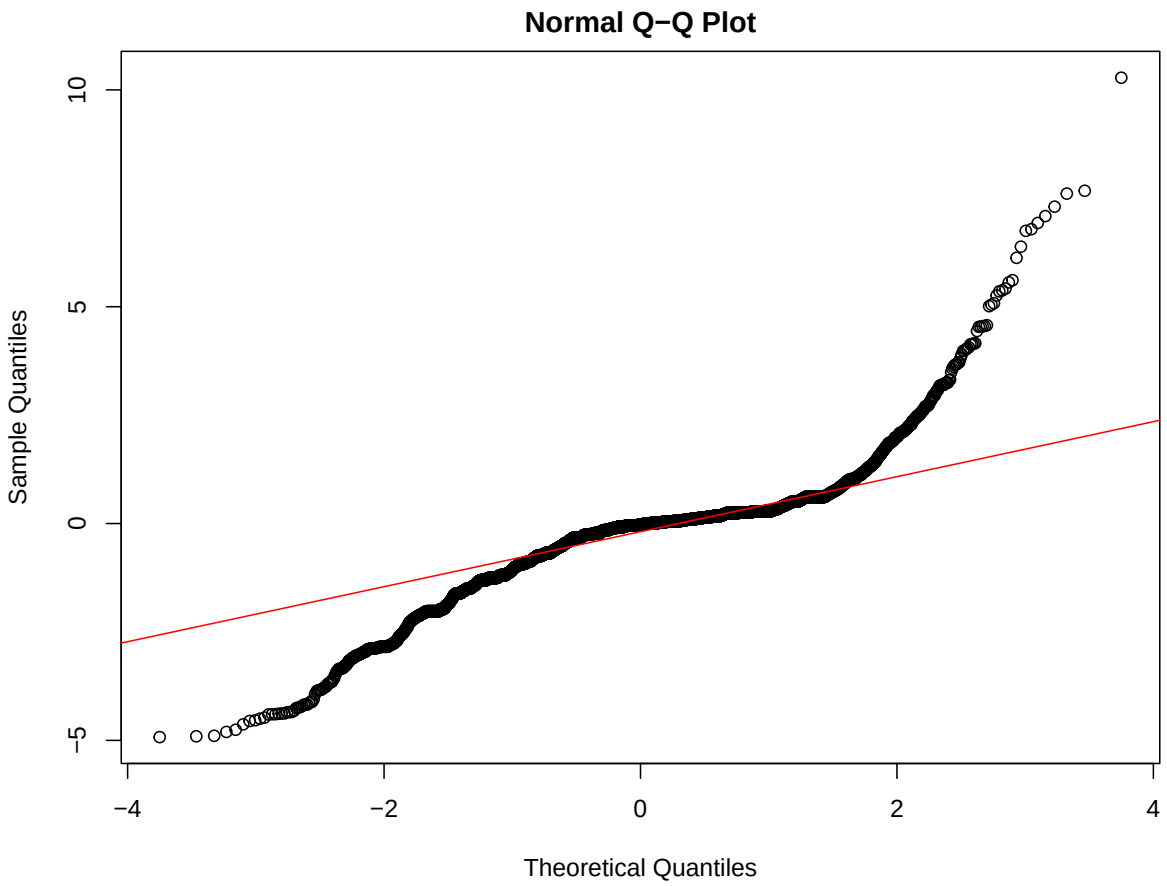

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
2500
Fitted values

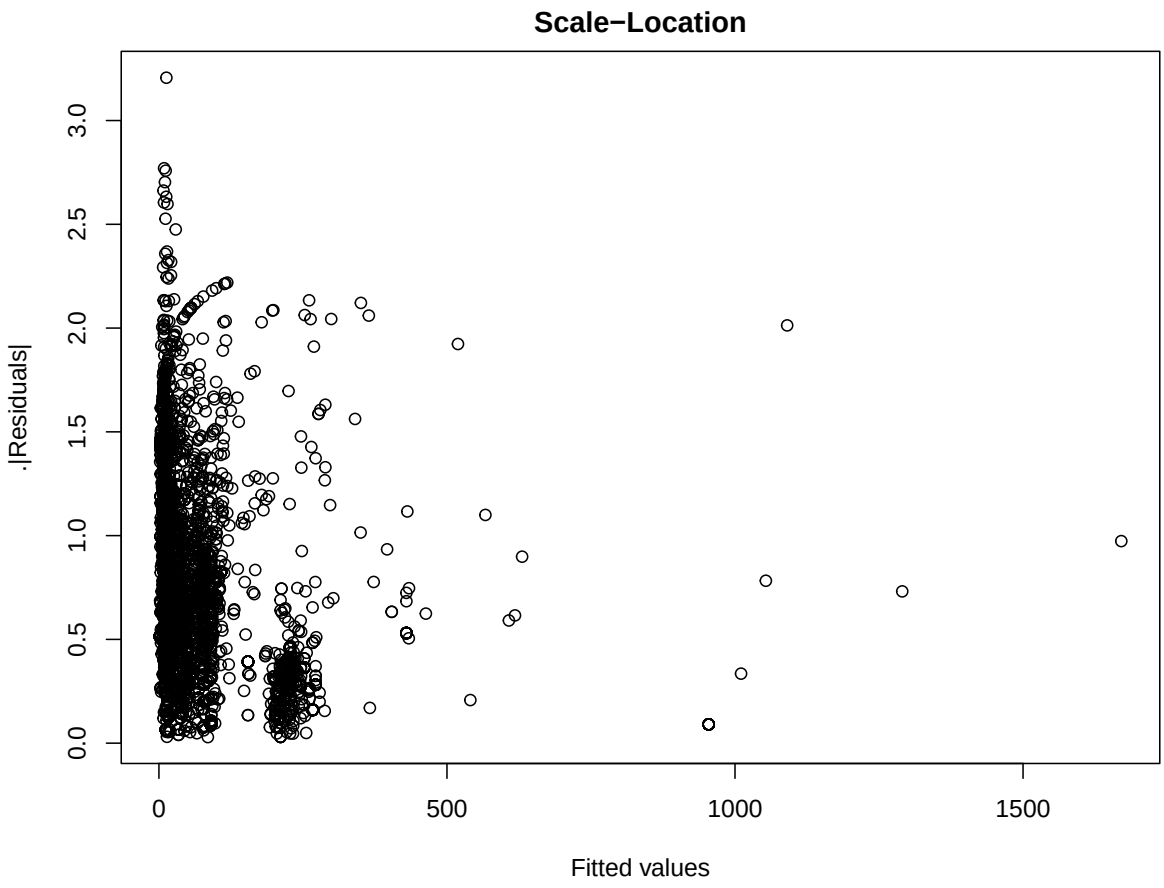


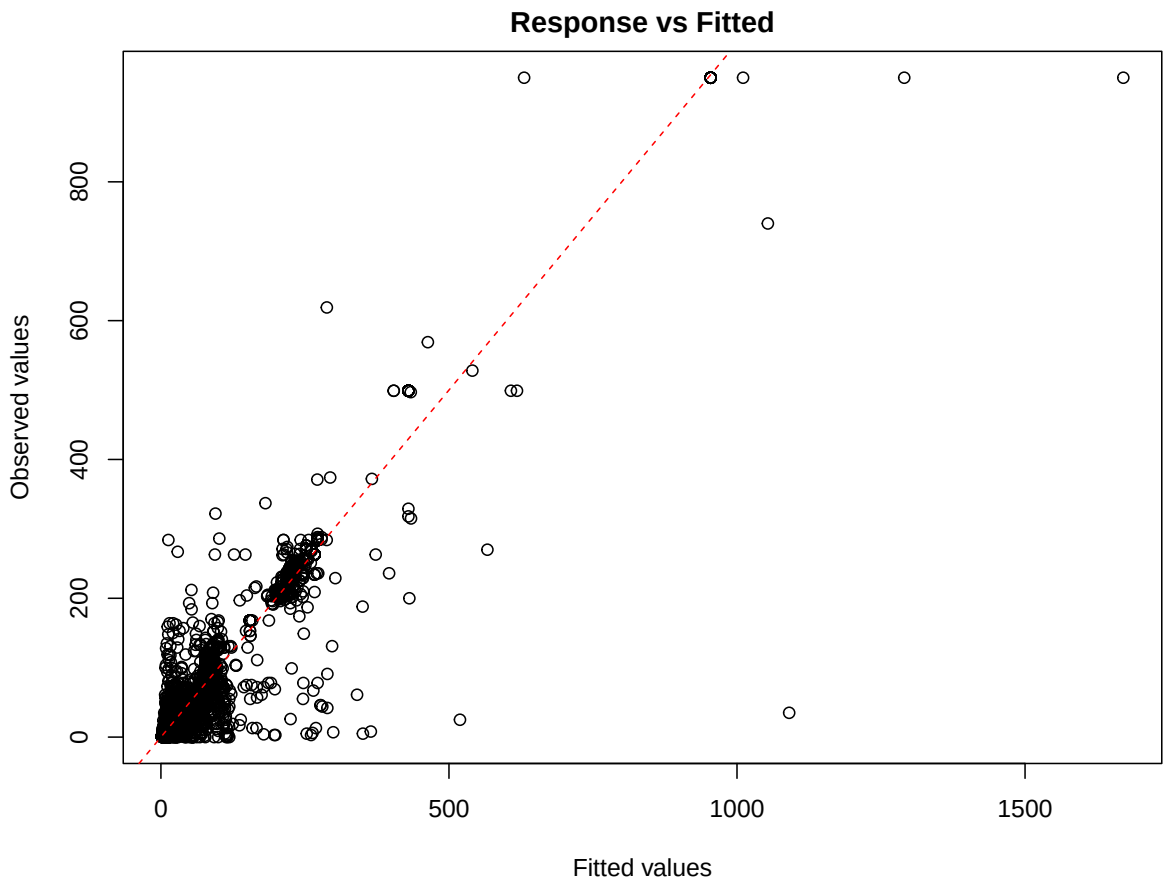

```
### Model 4 : 2c. Sunlight only

Model data dimensions after removing NAs: 5650 x 4
GAM fitted successfully
Model Summary:

Family: Negative Binomial(3.382)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(sunlight_proportion) +
    s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.4289     0.2358   14.54   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df  Chi.sq p-value
s(abundance_t_minus_1) 8.743  8.974 17069.4  <2e-16 ***
s(sunlight_proportion) 7.780  8.605   367.7  <2e-16 ***
s(view_id)             7.777  8.000   183.7  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.877   Deviance explained = 77.7%
-REML =  24530  Scale est. = 1         n = 5650


Diagnostic plots:
```



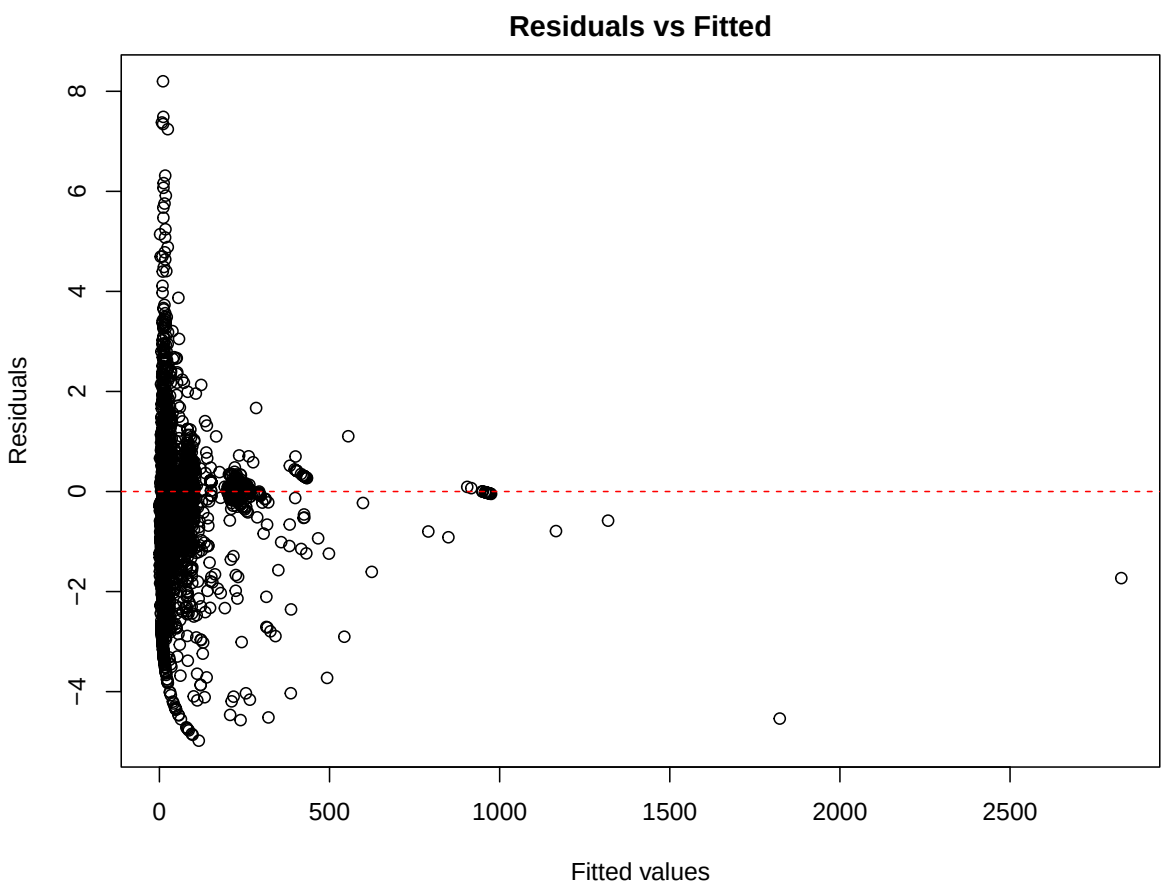

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4

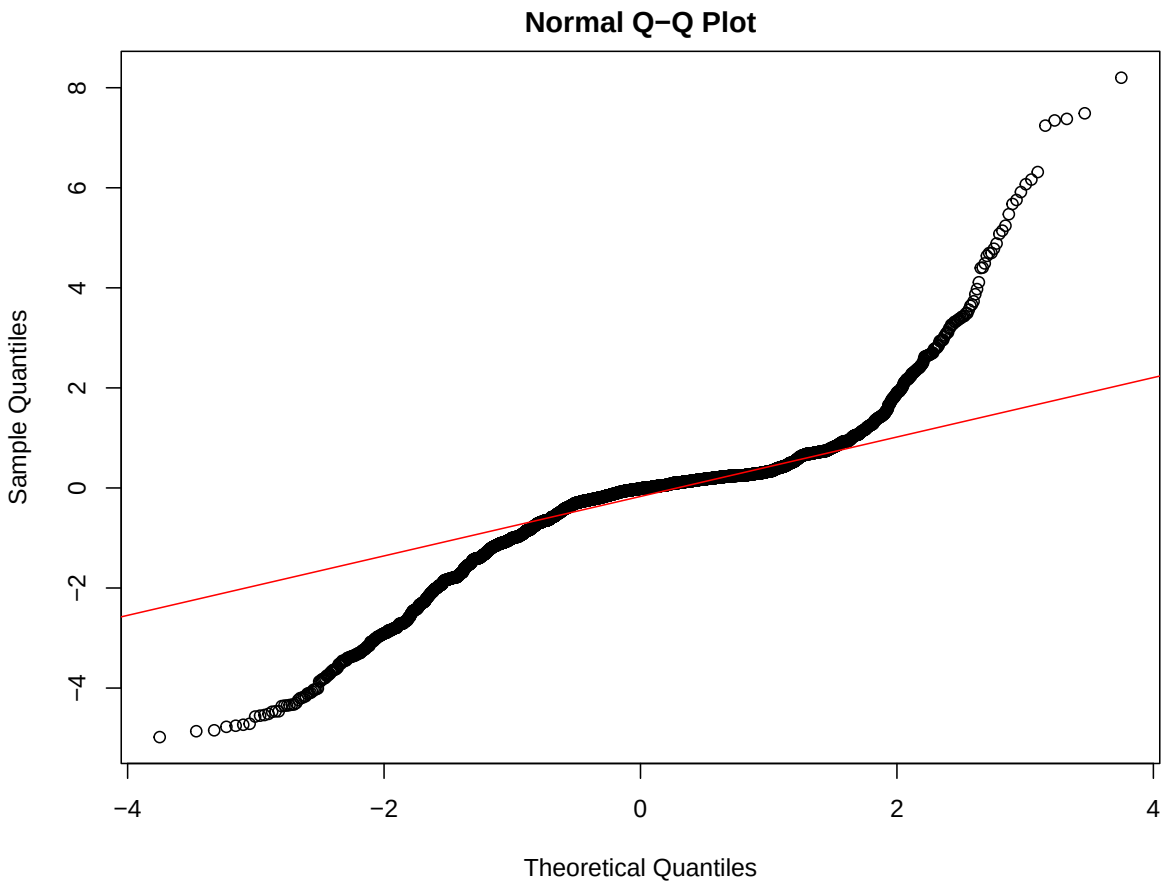

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
500
1000
1500
Fitted values

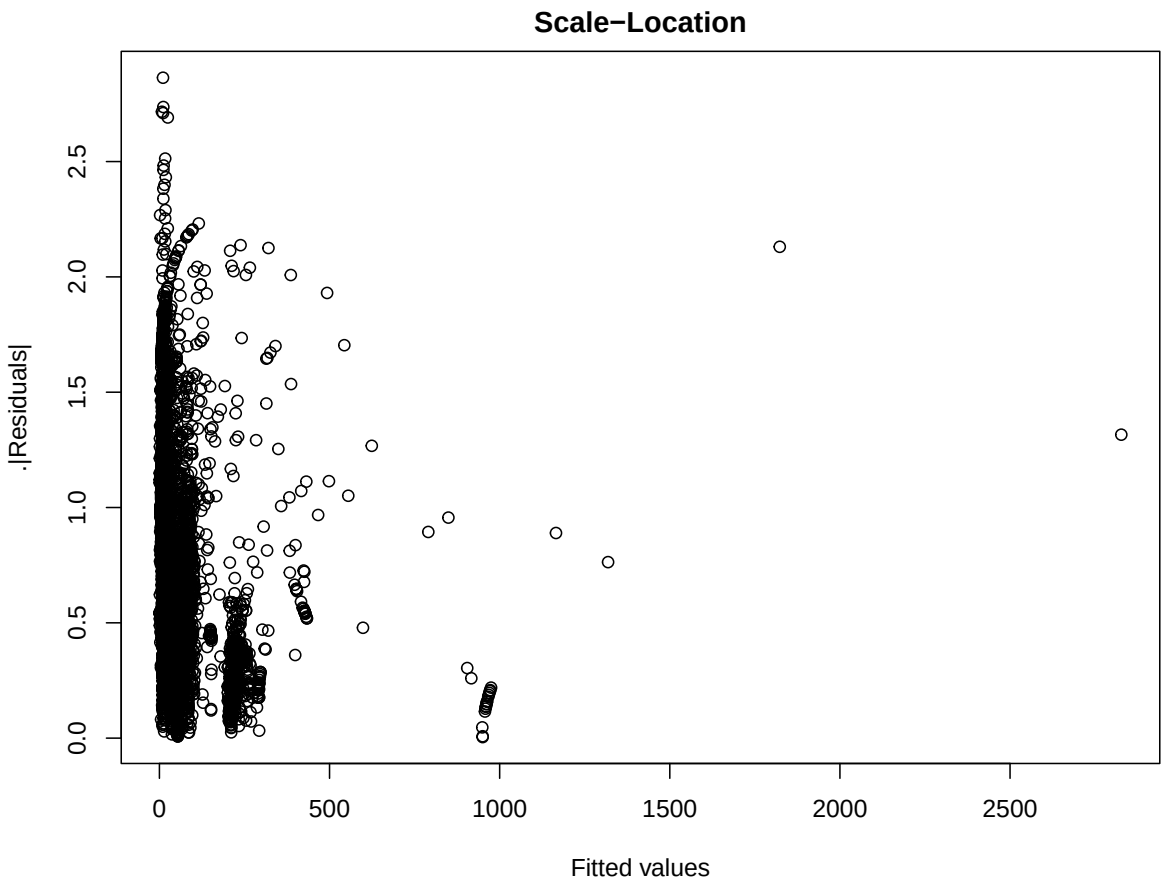


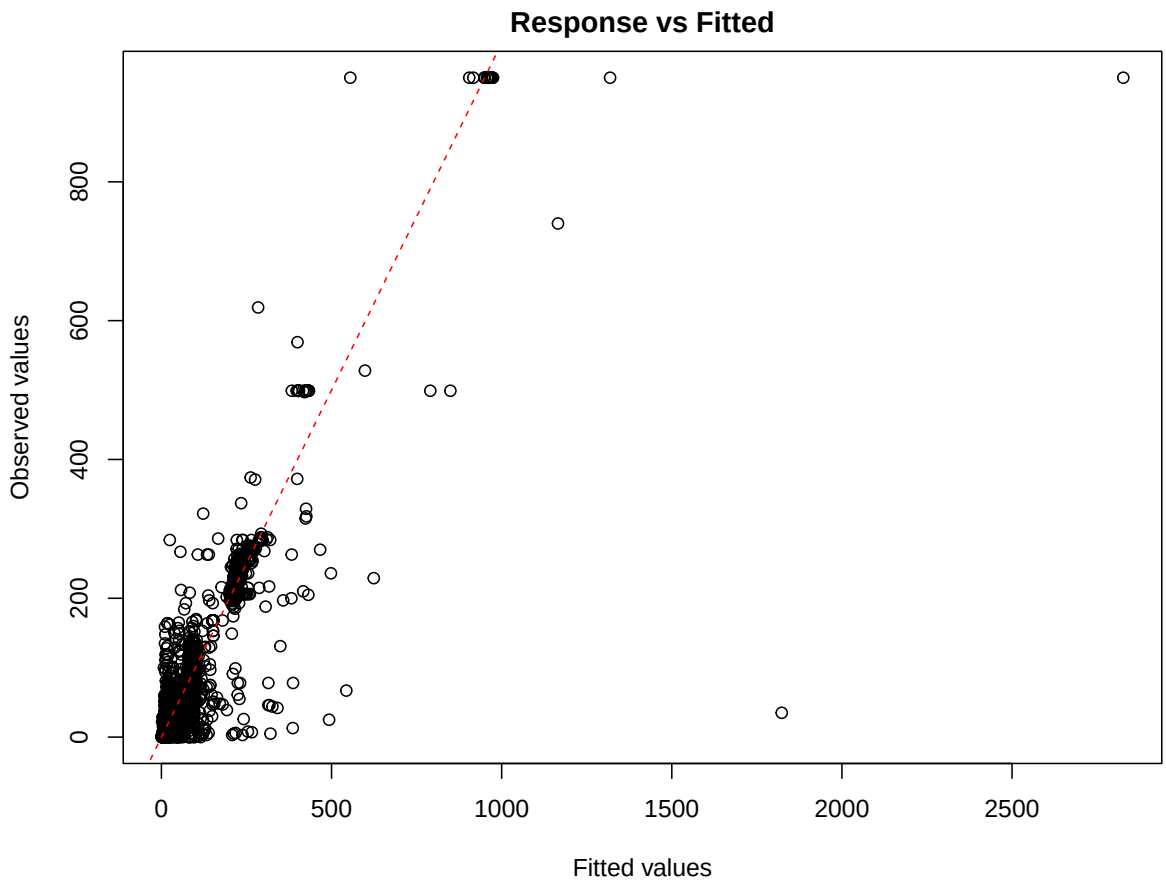

```
### Model 5 : 3a. Wind + Temp

Model data dimensions after removing NAs: 5650 x 5
GAM fitted successfully
Model Summary:

Family: Negative Binomial(3.521)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.4015     0.2101   16.19   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df   Chi.sq p-value
s(abundance_t_minus_1) 8.824  8.987 18061.44 < 2e-16 ***
s(mean_wind_speed)     4.794  5.886    33.54 8.9e-06 ***
s(mean_temp)           8.074  8.726   512.71 < 2e-16 ***
s(view_id)             7.728  8.000   272.17 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.766   Deviance explained = 78.5%
-REML =  24438  Scale est. = 1         n = 5650


Diagnostic plots:
```



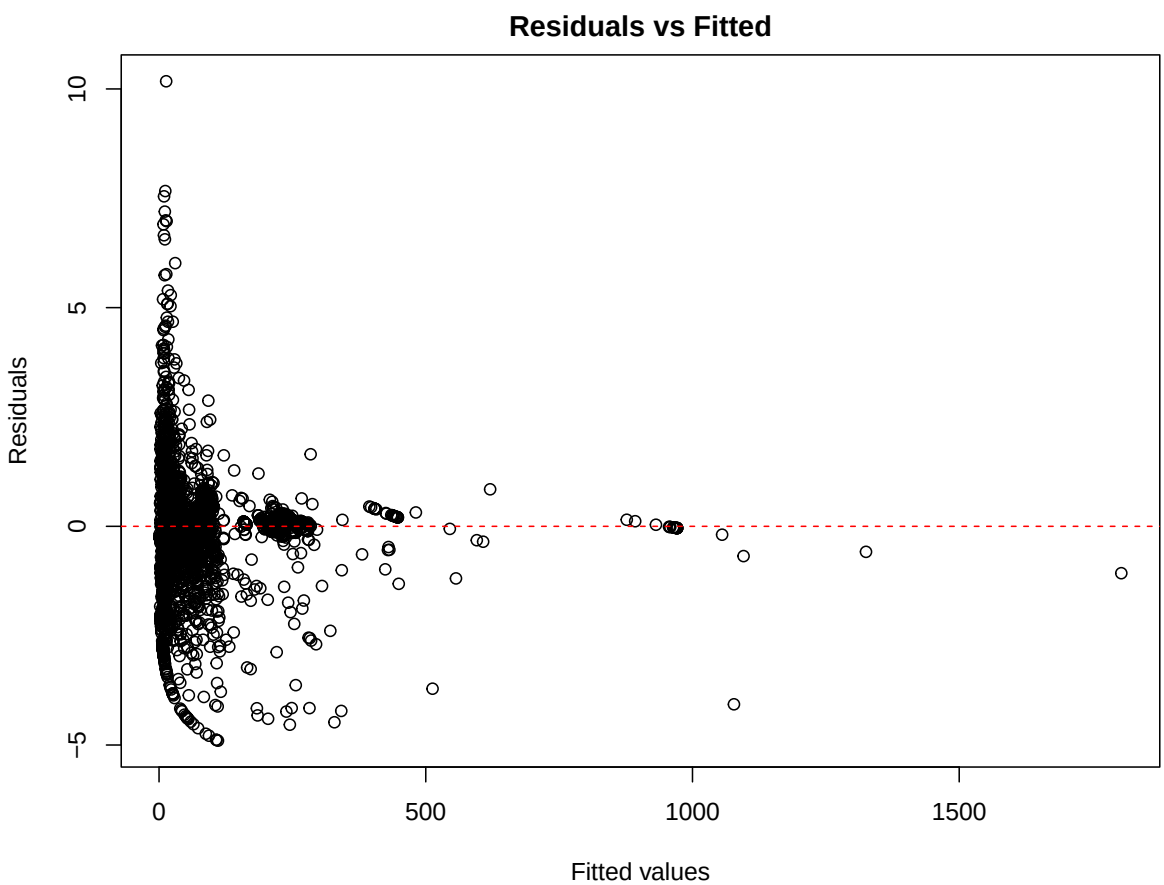

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
8
6
4
2
0
−2
−4
−4
−2
0
2
4

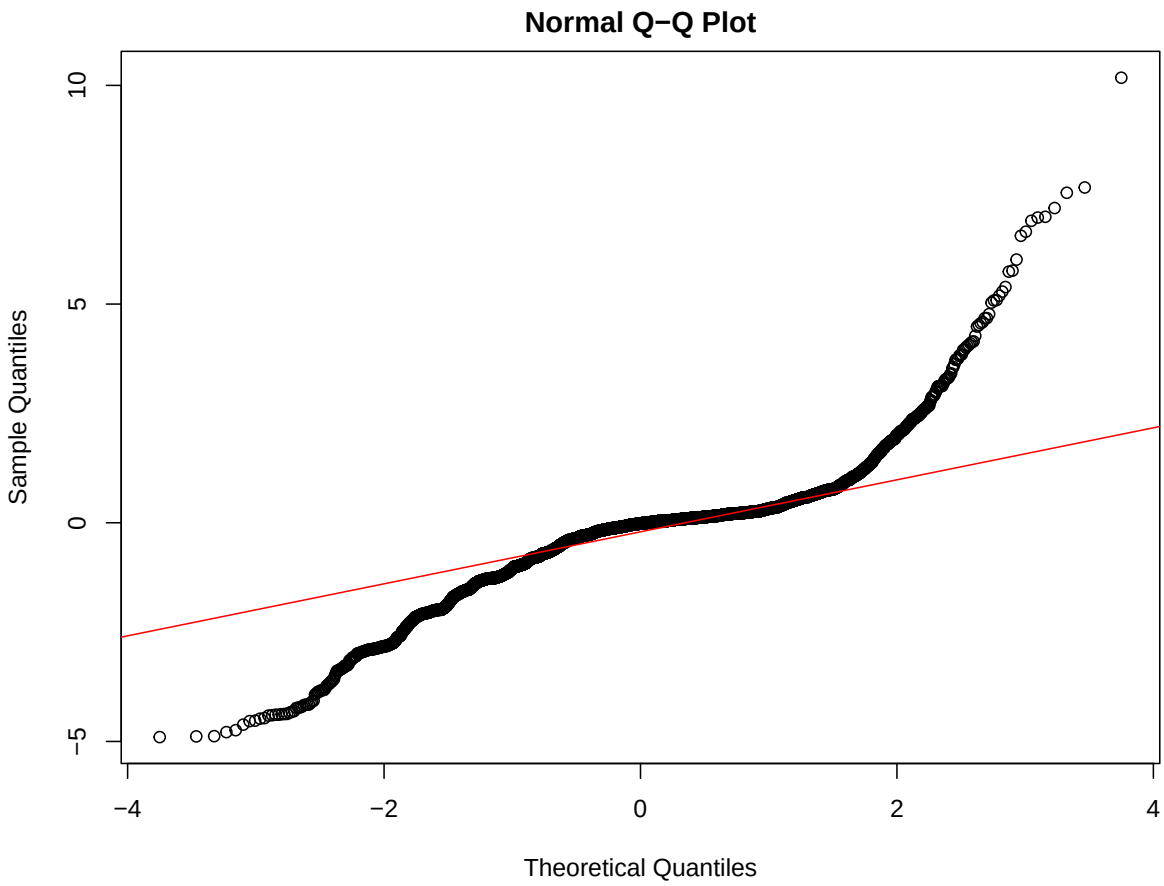

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
2500
Fitted values

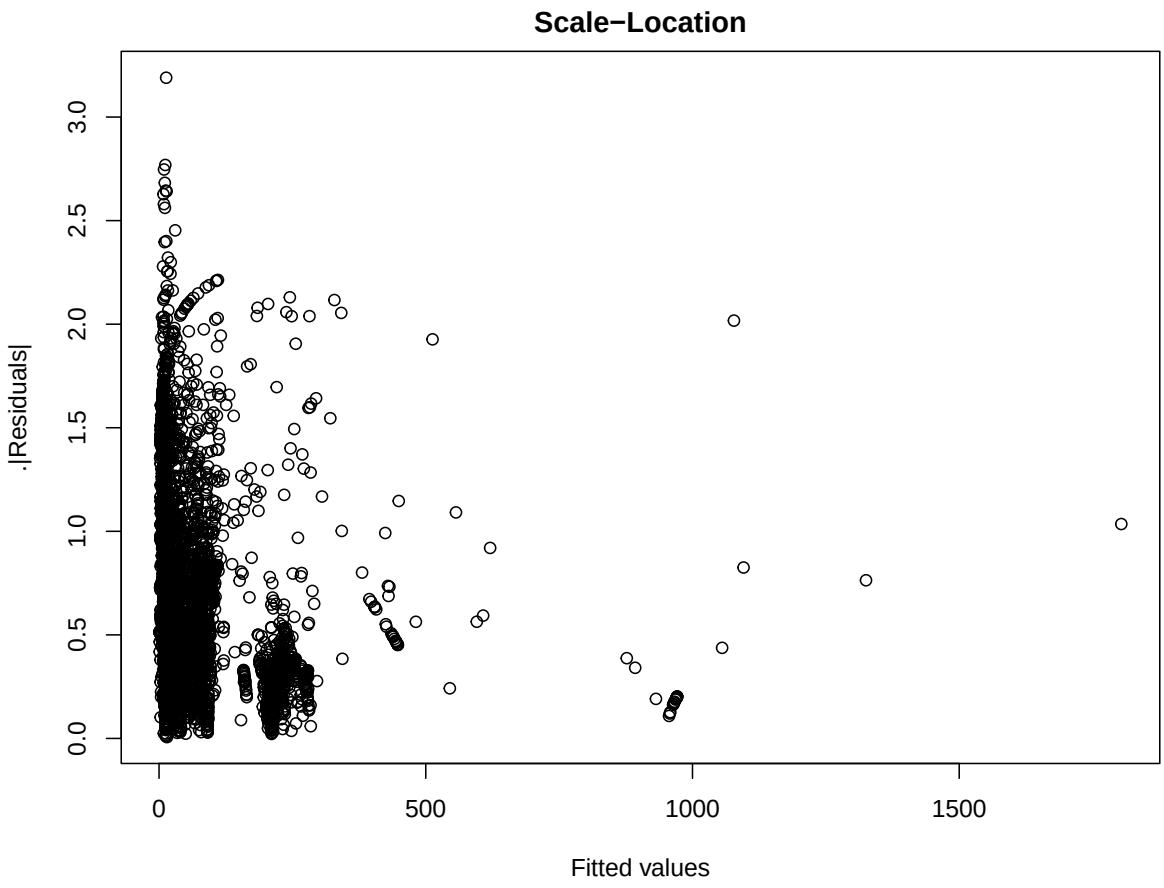


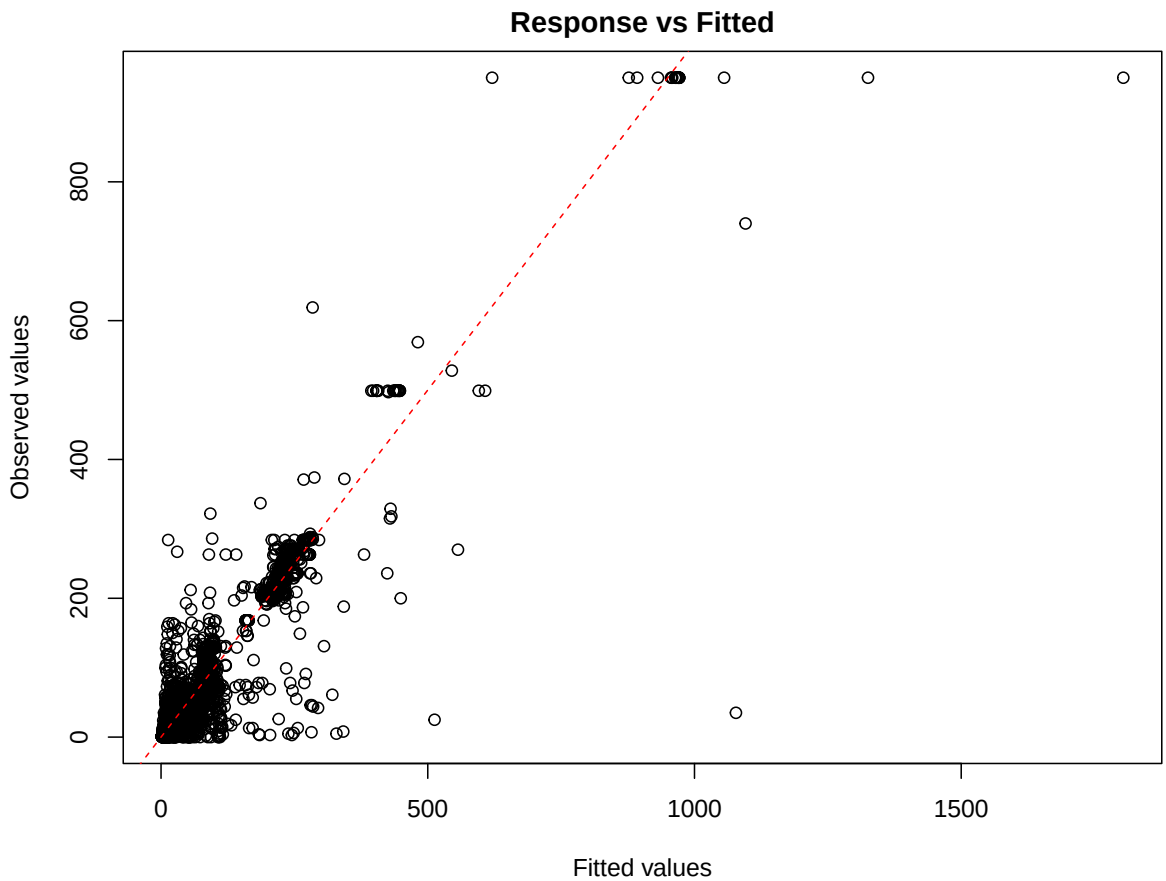

```
### Model 6 : 3b. Wind + Sunlight

Model data dimensions after removing NAs: 5650 x 5
GAM fitted successfully
Model Summary:

Family: Negative Binomial(3.427)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(sunlight_proportion) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.4433     0.2456   14.02   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df   Chi.sq p-value
s(abundance_t_minus_1) 8.757  8.976 16564.25  <2e-16 ***
s(mean_wind_speed)     5.472  6.628    63.34  <2e-16 ***
s(sunlight_proportion) 7.780  8.604   376.98  <2e-16 ***
s(view_id)             7.793  8.000   198.03  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.873   Deviance explained = 77.9%
-REML =  24511  Scale est. = 1         n = 5650


Diagnostic plots:
```



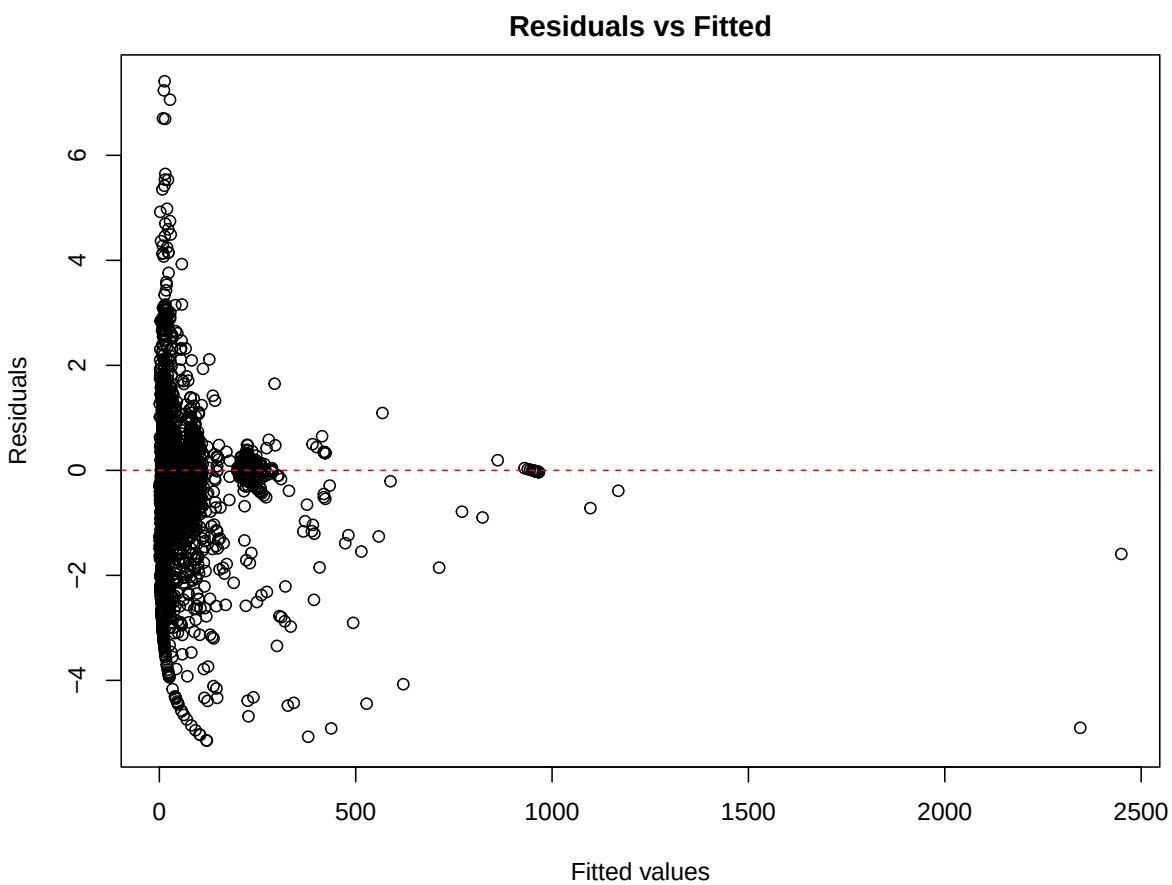

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−4
−2
0
2
4

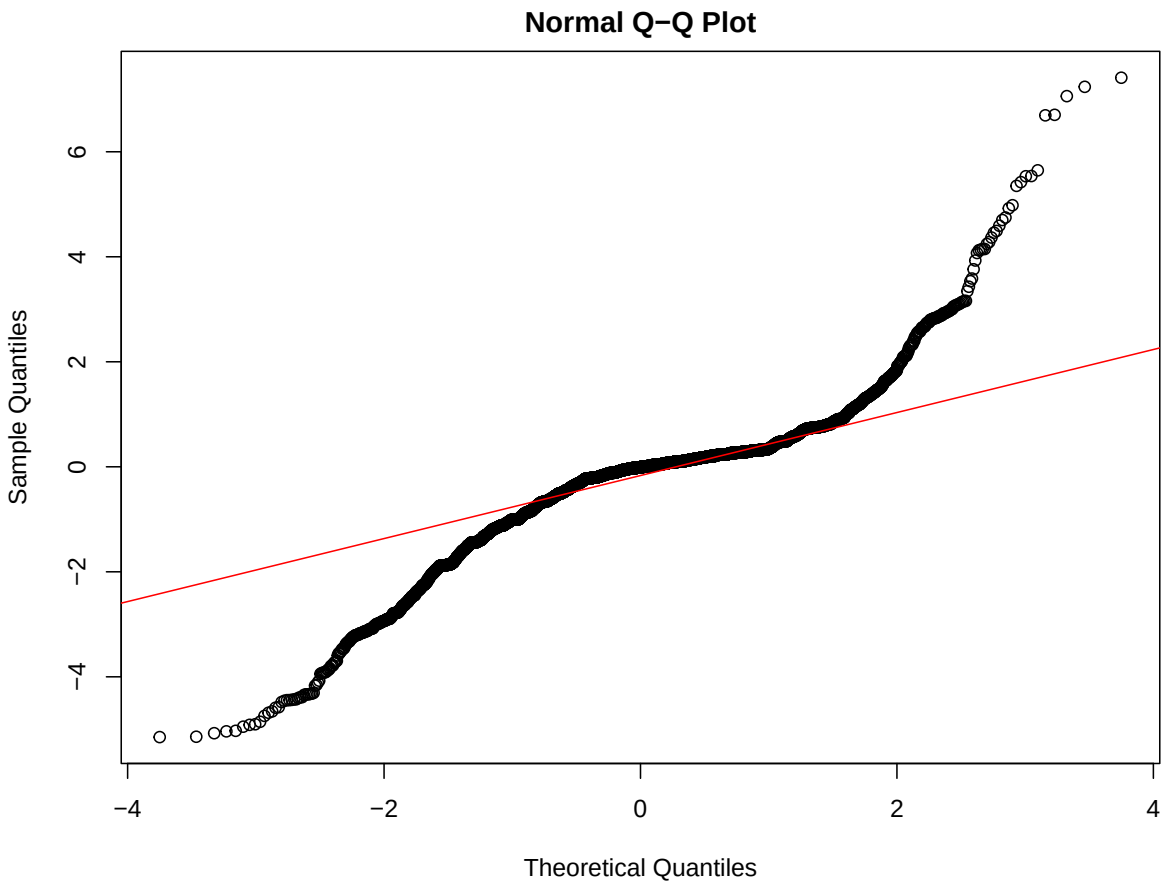

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
500
1000
1500
Fitted values

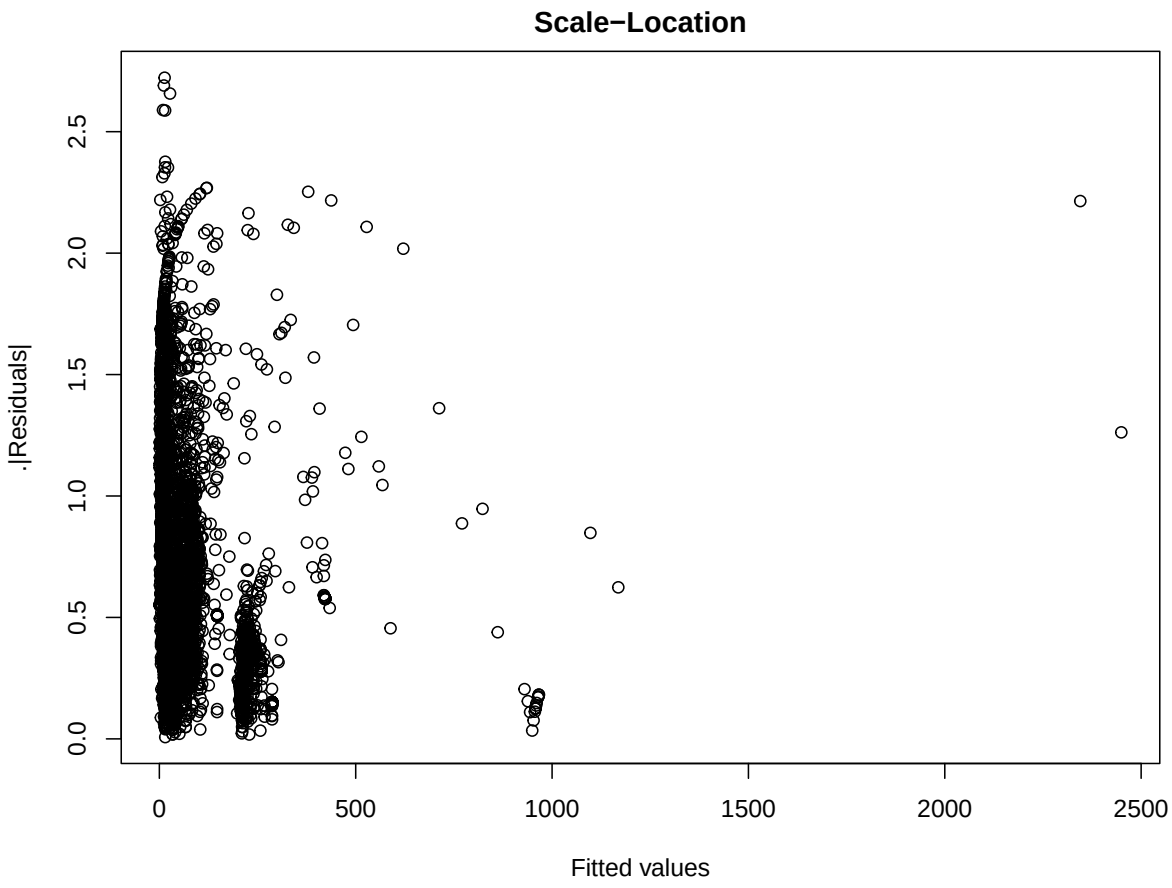


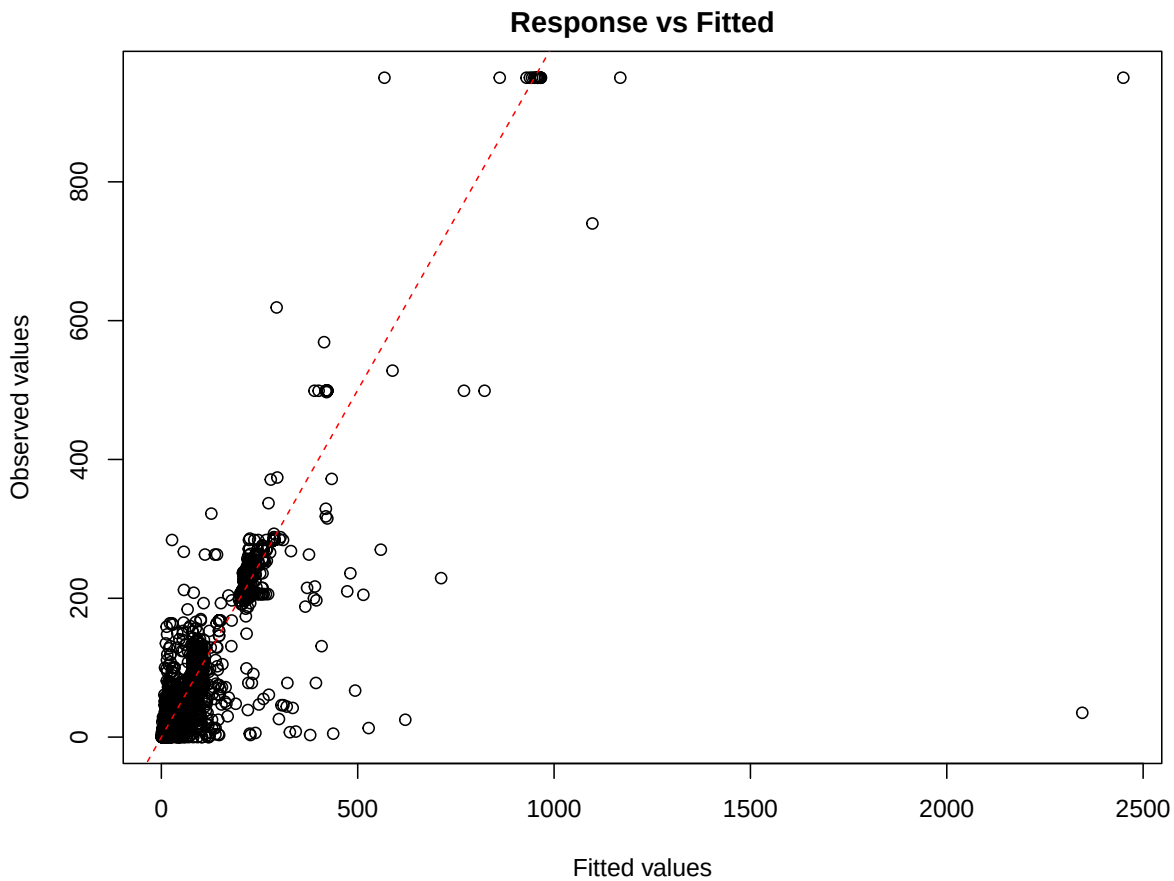

```
### Model 7 : 3c. Temp + Sunlight

Model data dimensions after removing NAs: 5650 x 5
GAM fitted successfully
Model Summary:

Family: Negative Binomial(3.794)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(sunlight_proportion) +
    s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3801     0.1999   16.91   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df  Chi.sq p-value
s(abundance_t_minus_1) 8.819  8.987 19088.9  <2e-16 ***
s(mean_temp)           8.258  8.813   553.2  <2e-16 ***
s(sunlight_proportion) 7.753  8.589   415.0  <2e-16 ***
s(view_id)             7.716  8.000   282.3  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.735   Deviance explained = 79.9%
-REML =  24265  Scale est. = 1         n = 5650


Diagnostic plots:
```



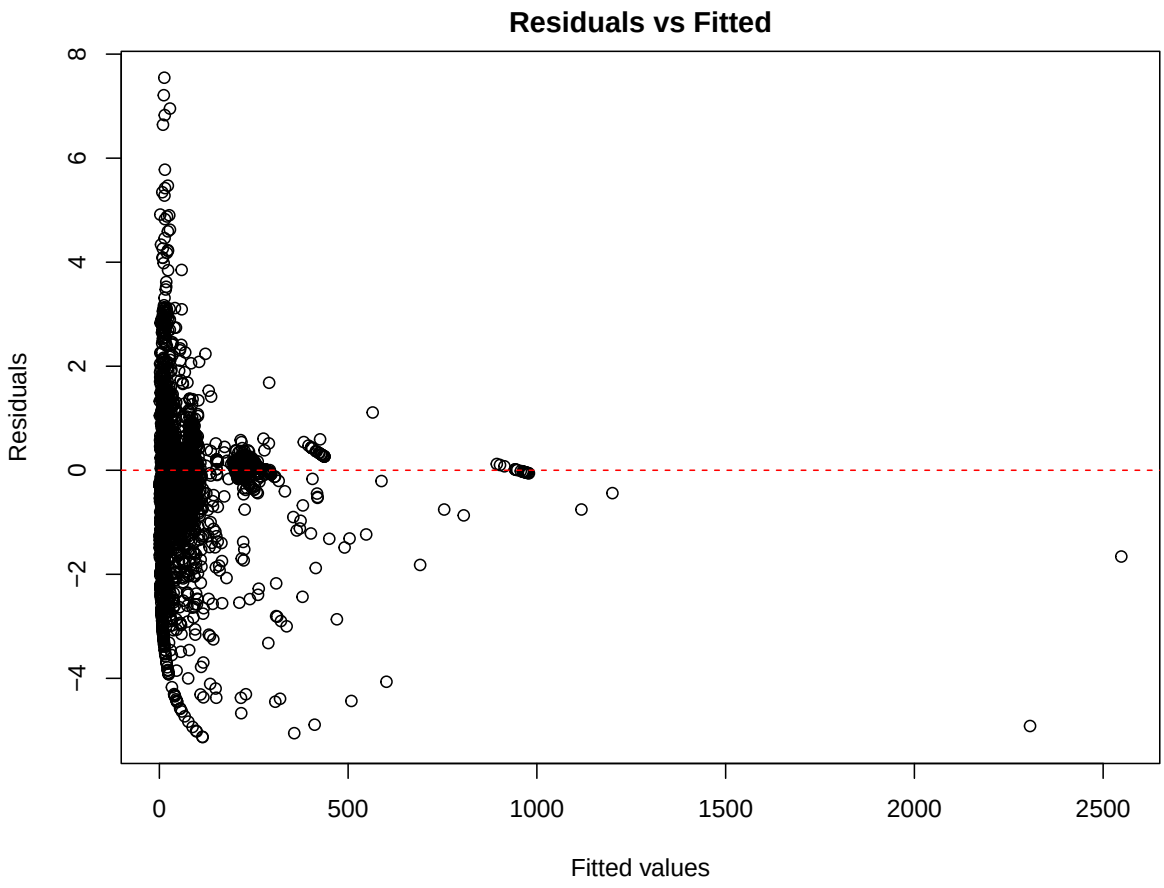

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6
−4
−2
0
2
4

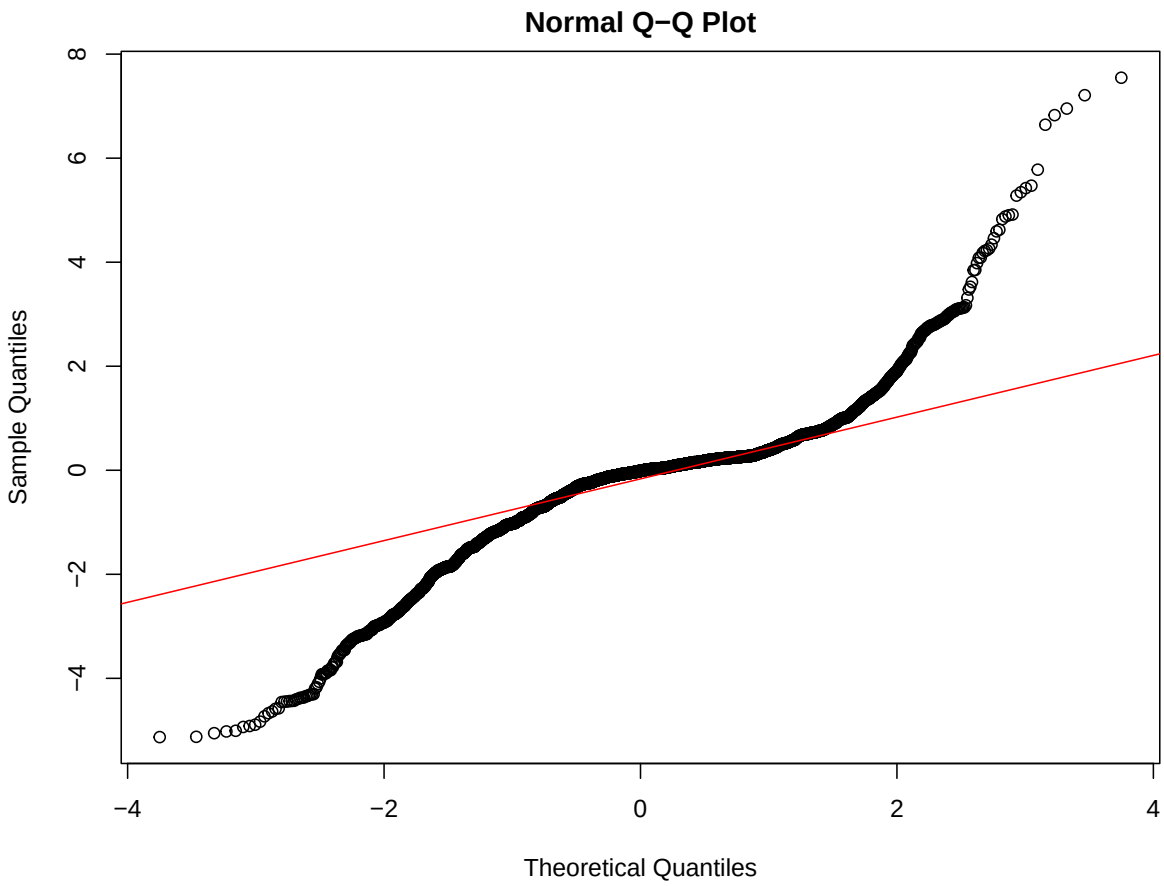

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
2500
Fitted values

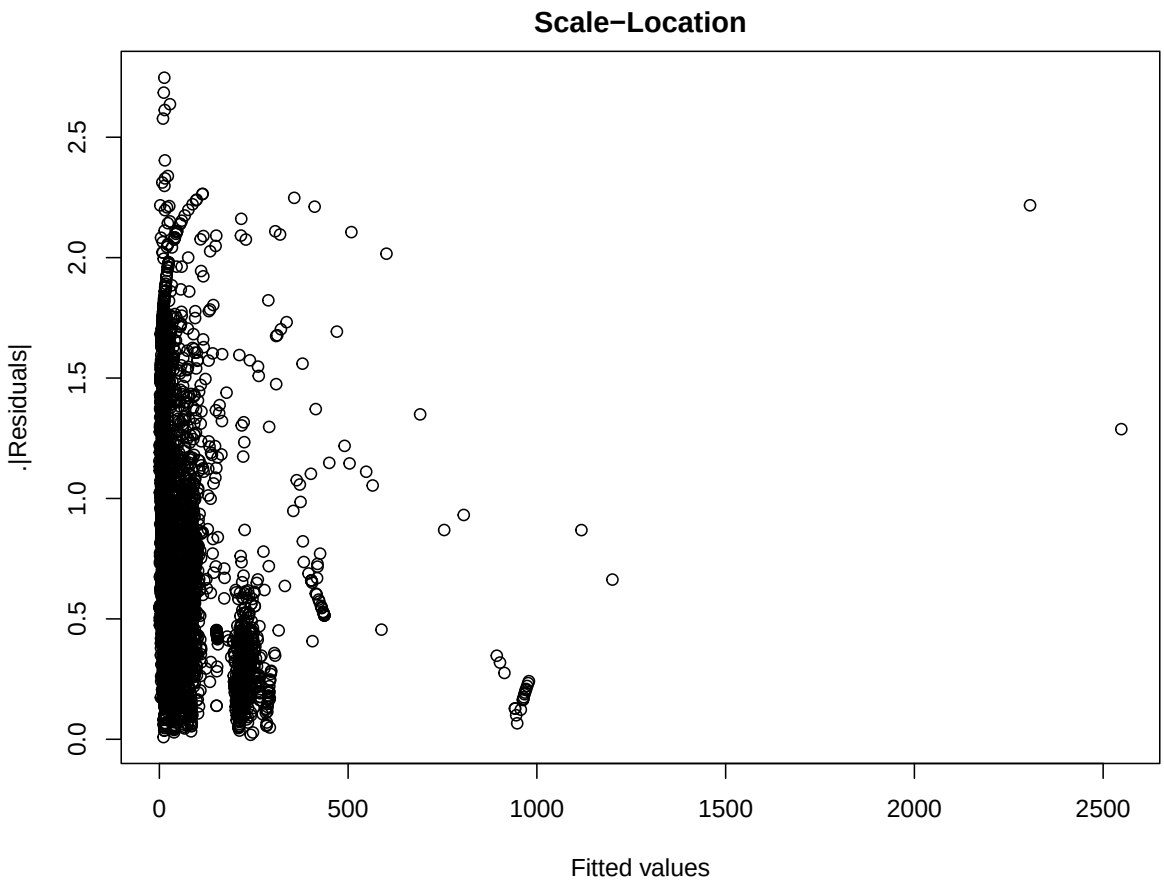


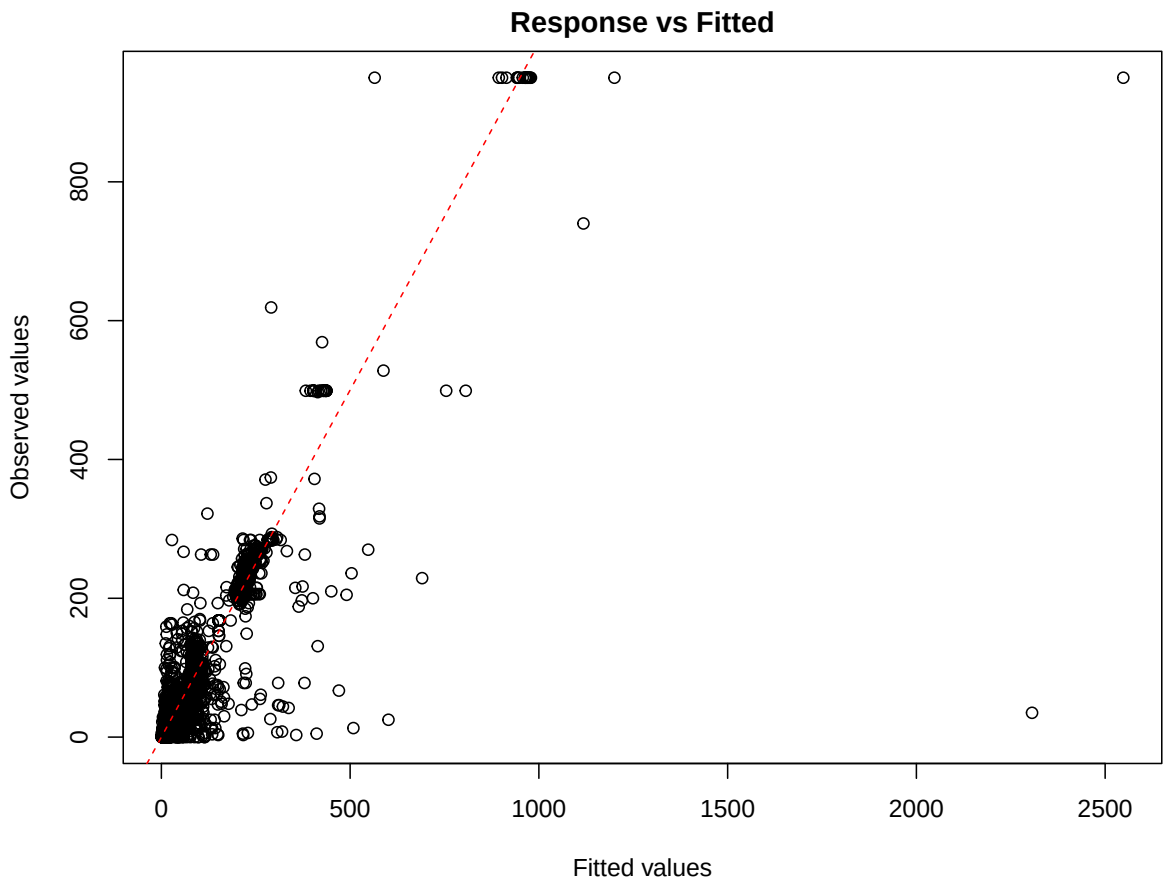

```
### Model 8 : 3d. All additive

Model data dimensions after removing NAs: 5650 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(3.835)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3922     0.2083   16.28   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df   Chi.sq p-value
s(abundance_t_minus_1) 8.825  8.987 18554.57  <2e-16 ***
s(mean_wind_speed)     5.088  6.214    45.62  <2e-16 ***
s(mean_temp)           8.264  8.816   537.10  <2e-16 ***
s(sunlight_proportion) 7.720  8.569   425.48  <2e-16 ***
s(view_id)             7.735  8.000   270.19  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.737   Deviance explained =   80%
-REML =  24253  Scale est. = 1         n = 5650


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
8
6
4
2
0
−2
−4
0
500
1000
1500
2000
2500

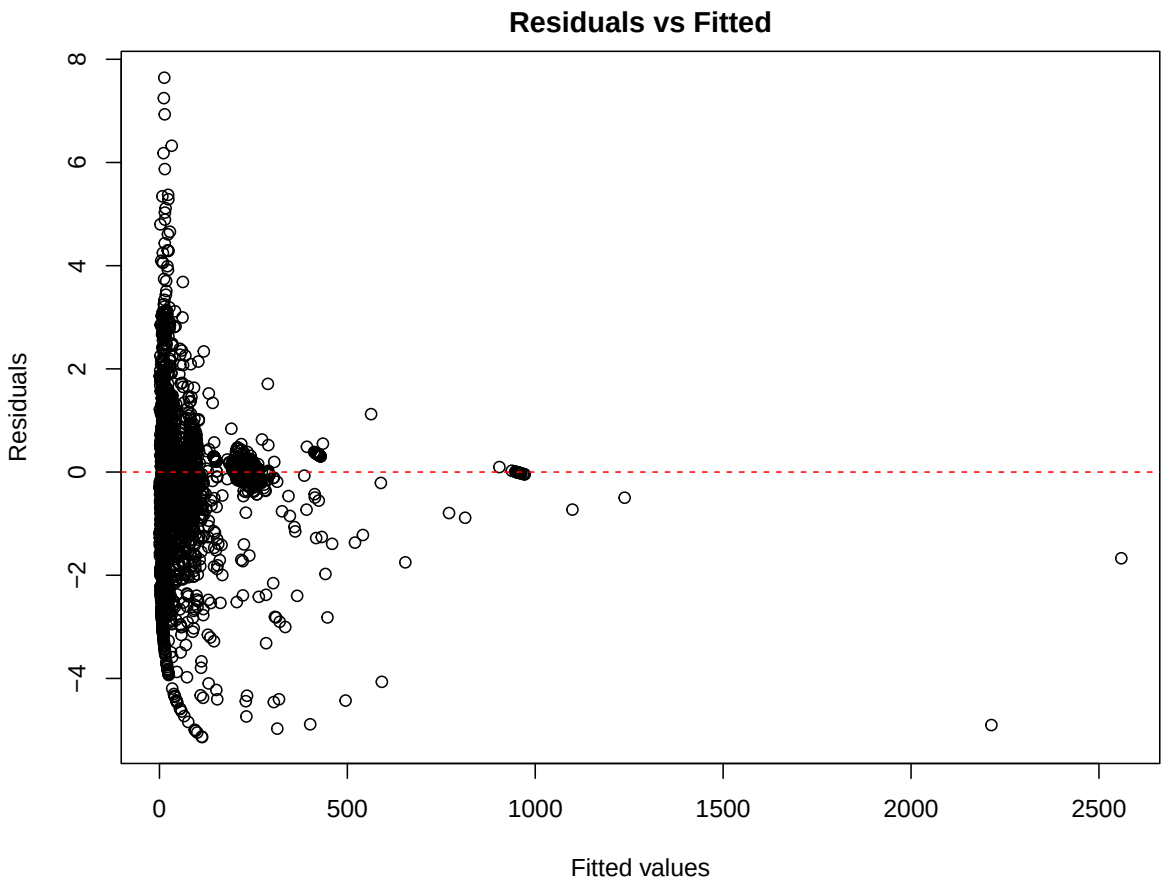

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
8
6
4
2
0
−2
−4
−4
−2
0
2
4

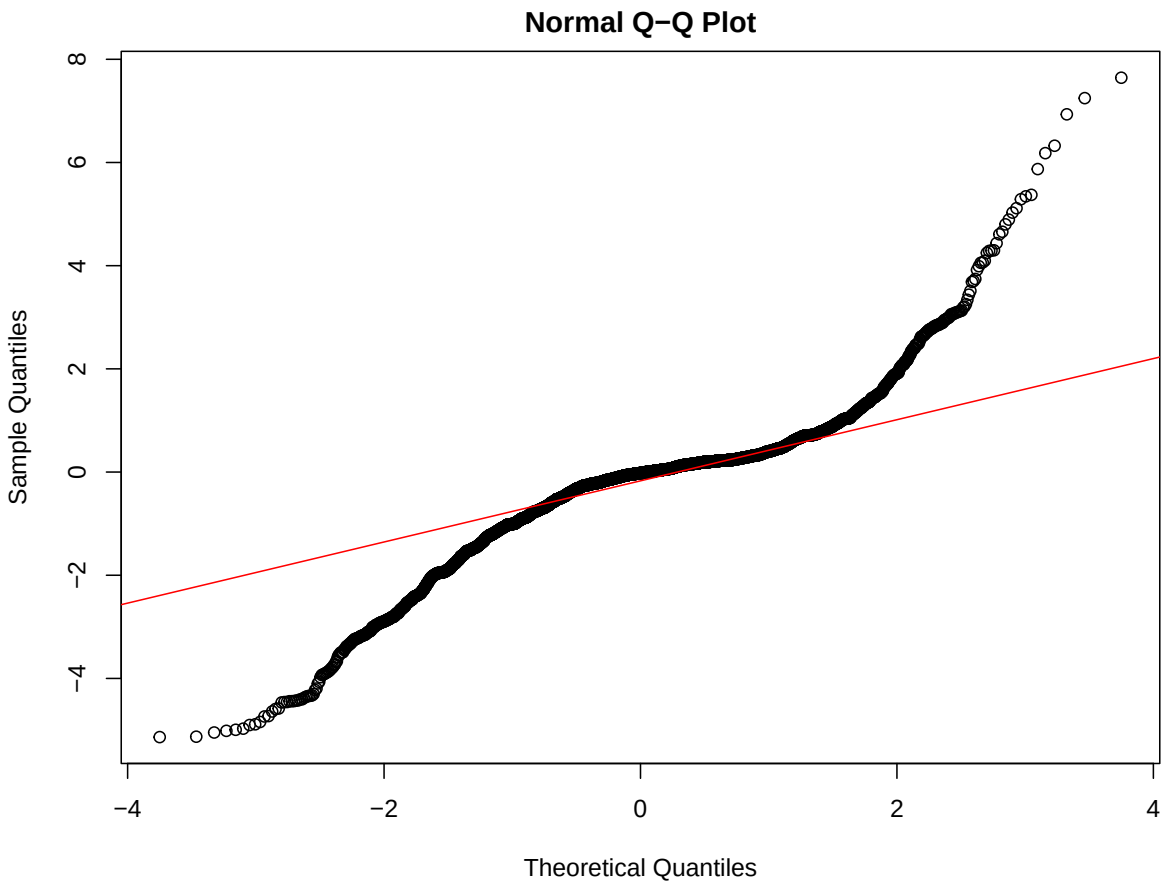

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
2500
Fitted values

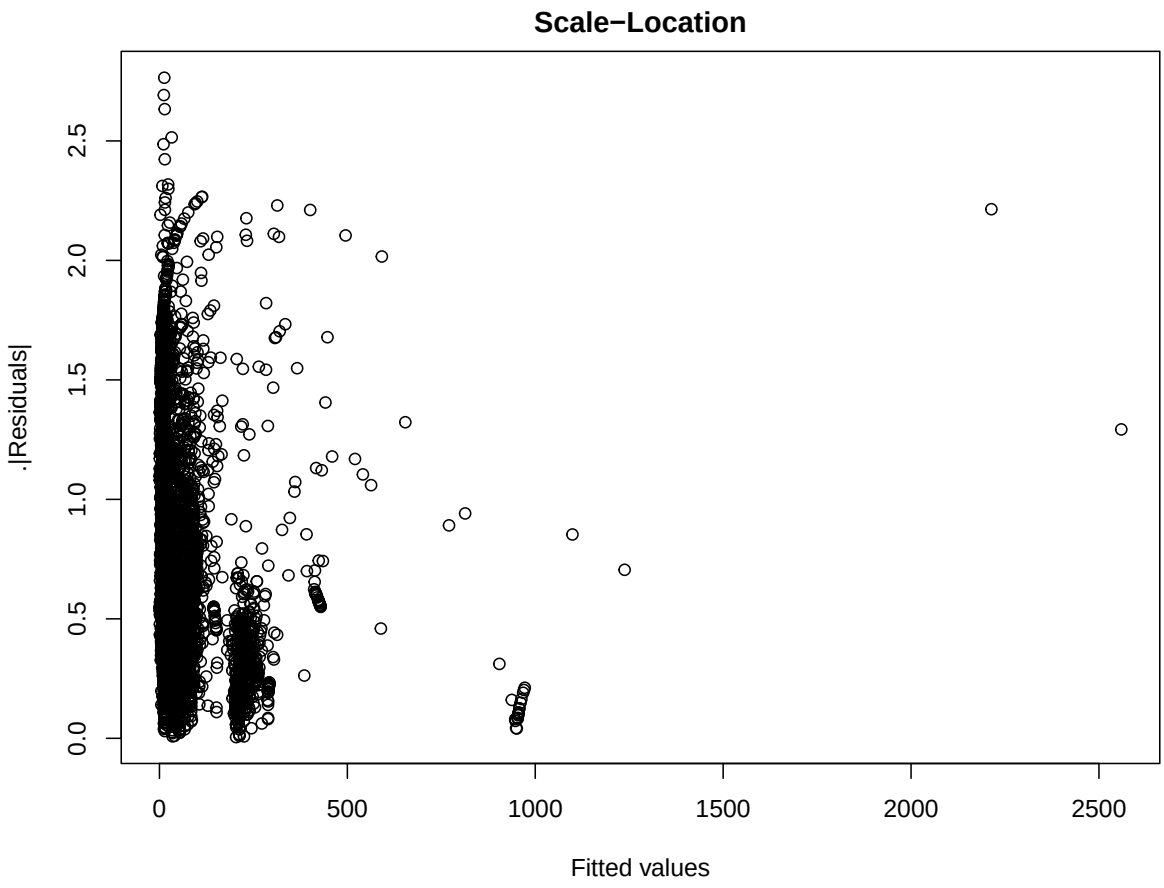


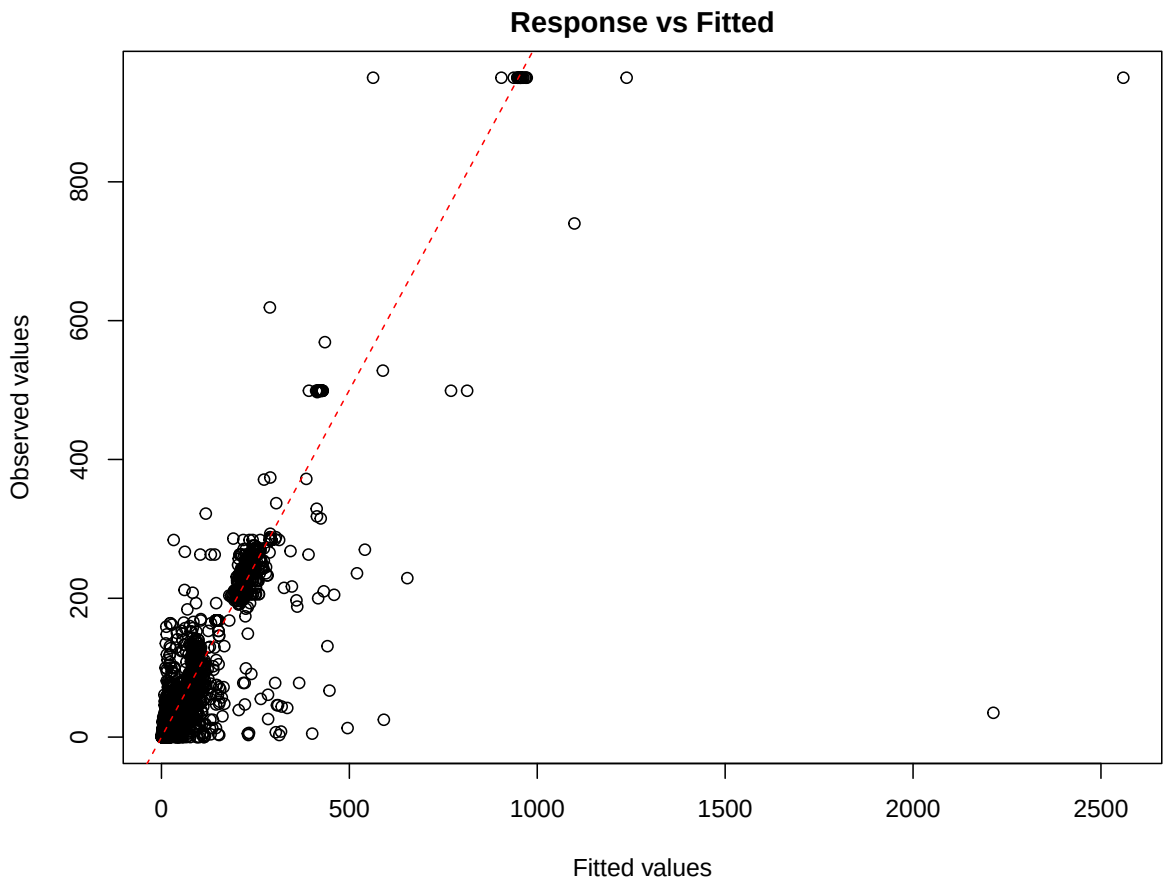

```
### Model 9 : 4a. Wind x Temp interaction

Model data dimensions after removing NAs: 5650 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(3.863)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)   3.3984     0.2112    16.1   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                edf Ref.df   Chi.sq  p-value
s(abundance_t_minus_1)        8.829  8.988 17504.69  < 2e-16 ***
s(mean_wind_speed)            4.672  5.747    23.73  0.00062 ***
s(mean_temp)                  8.329  8.842   485.51  < 2e-16 ***
s(sunlight_proportion)        7.759  8.592   425.47  < 2e-16 ***
ti(mean_wind_speed,mean_temp) 4.512  5.498    32.05 1.22e-05 ***
s(view_id)                    7.740  8.000   272.78  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.746   Deviance explained = 80.2%
-REML =  24242  Scale est. = 1         n = 5650


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
8
6
4
2
0
−2
−4
0
500
1000
1500
2000
2500

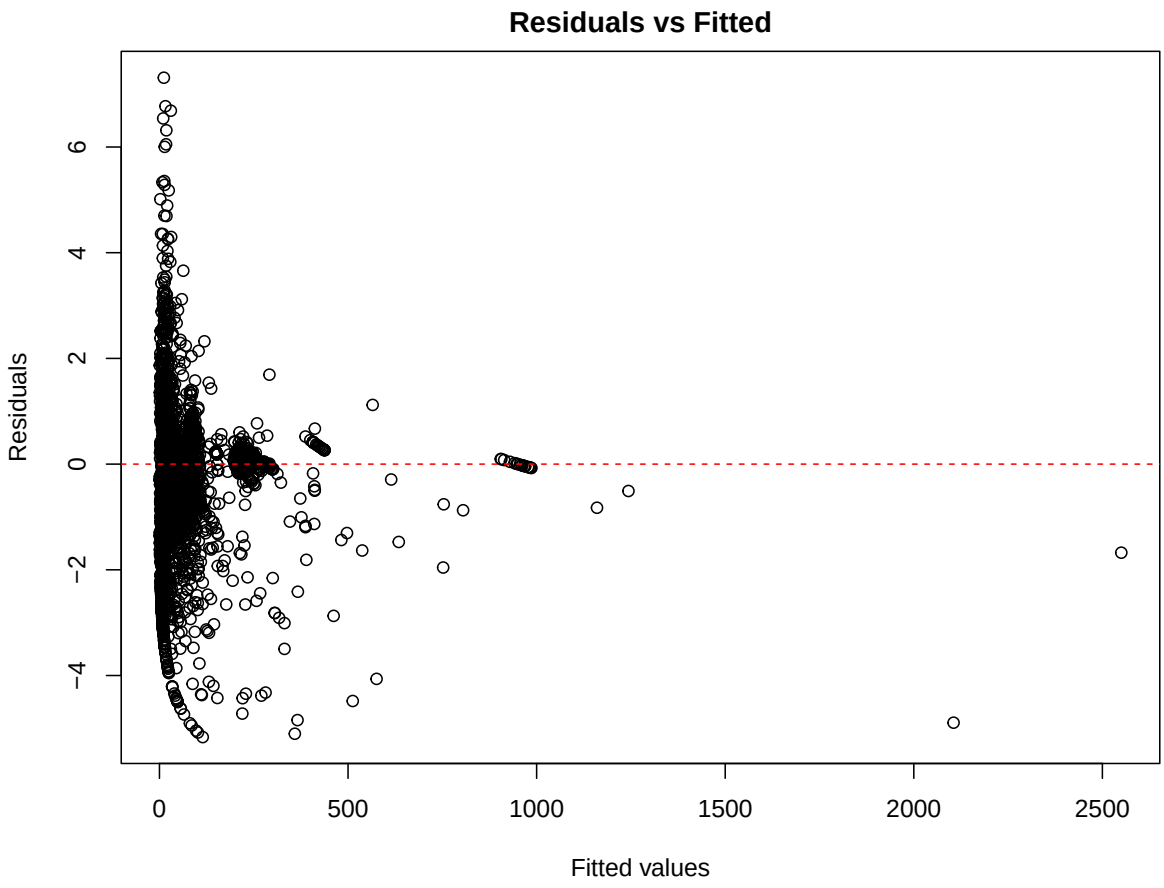

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
8
6
4
2
0
−2
−4
−4
−2
0
2
4

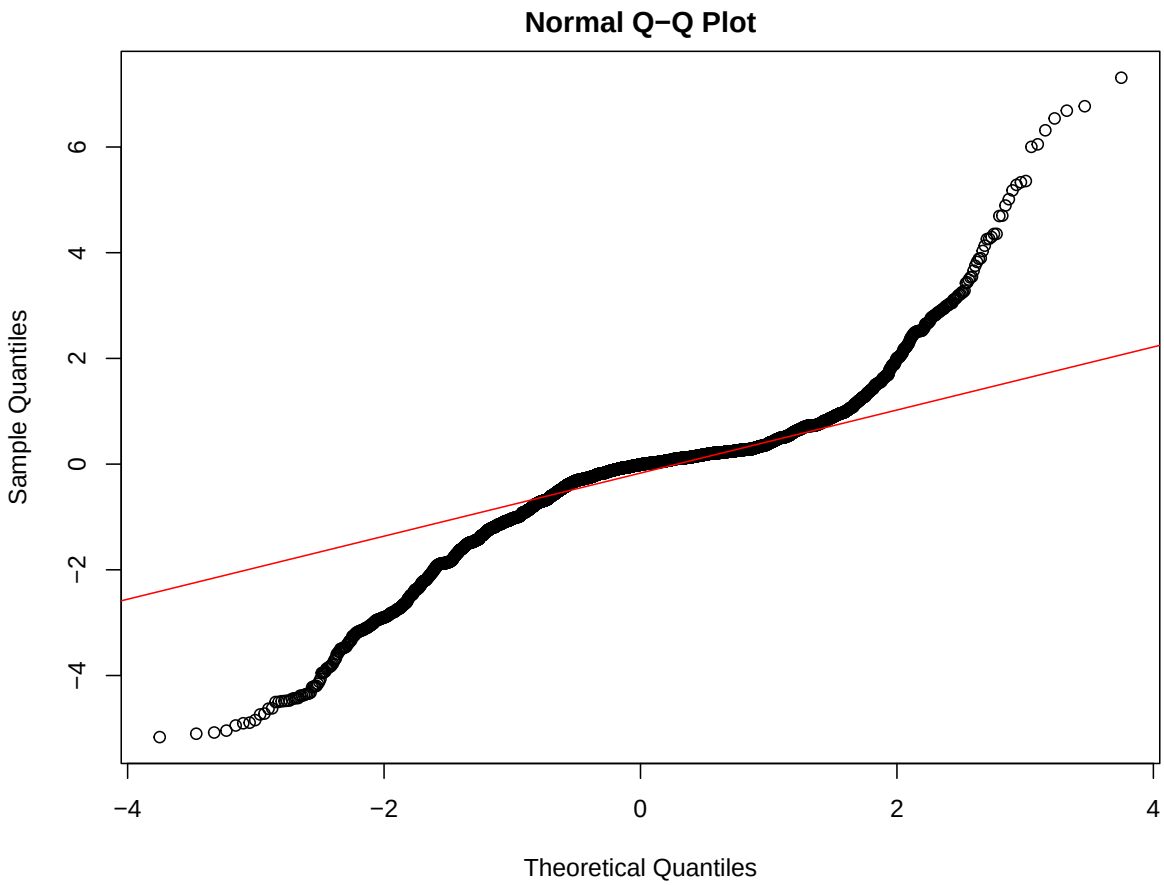

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
2500
Fitted values

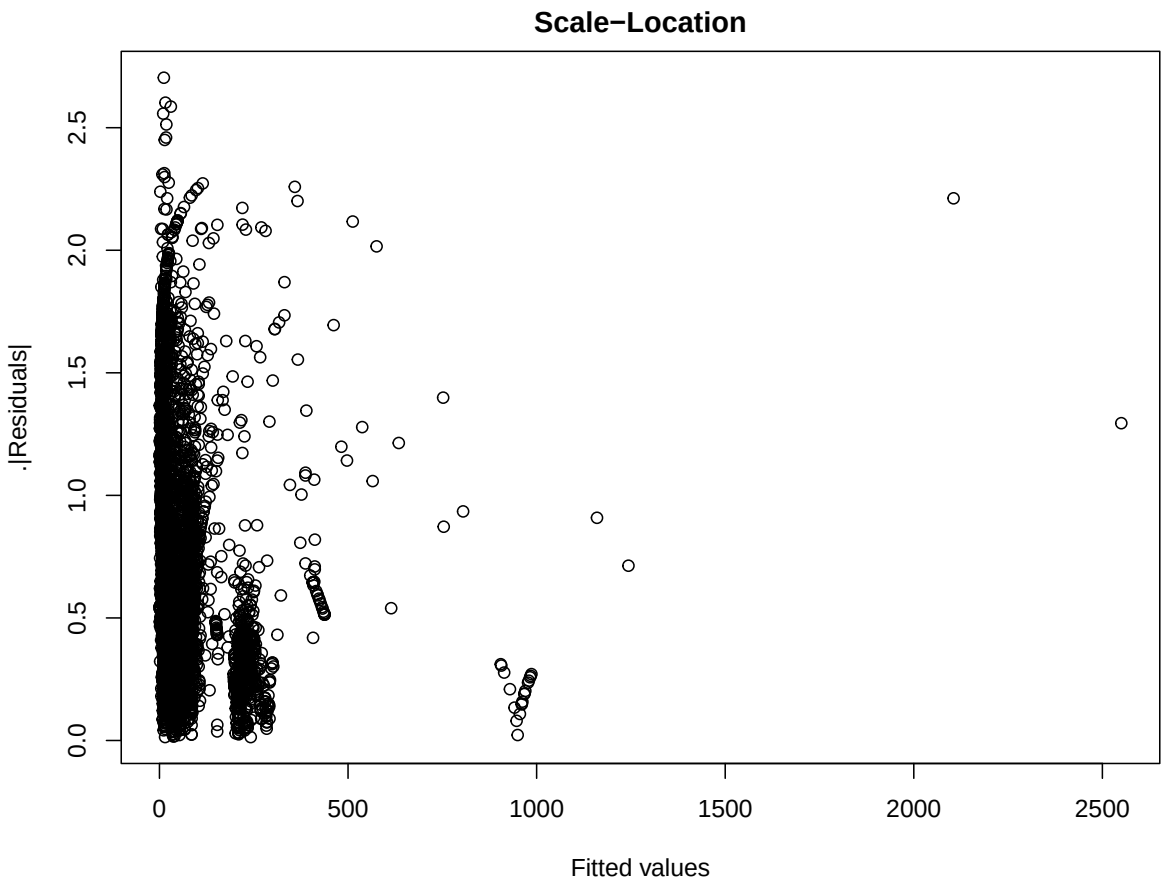


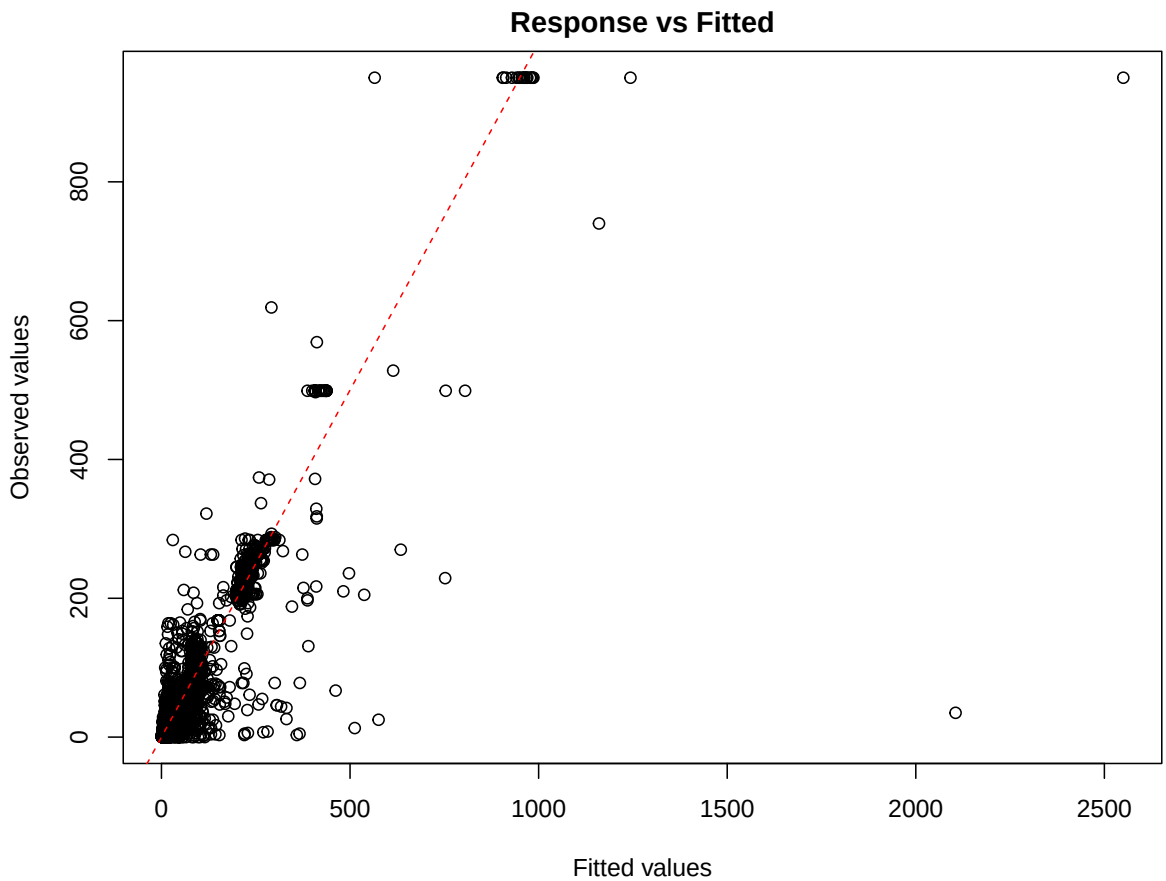

```
### Model 10 : 4b. Wind x Sunlight interaction

Model data dimensions after removing NAs: 5650 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(3.909)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    sunlight_proportion) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)   3.3889     0.2106    16.1   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                          edf Ref.df   Chi.sq  p-value
s(abundance_t_minus_1)                  8.827  8.988 18667.24  < 2e-16 ***
s(mean_wind_speed)                      5.476  6.638    32.82 1.95e-05 ***
s(mean_temp)                            8.345  8.849   535.55  < 2e-16 ***
s(sunlight_proportion)                  7.672  8.538   344.52  < 2e-16 ***
ti(mean_wind_speed,sunlight_proportion) 10.905 12.752    83.76  < 2e-16 ***
s(view_id)                              7.741  8.000   264.83  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.751   Deviance explained = 80.4%
-REML =  24226  Scale est. = 1         n = 5650


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
6
4
2
0
−2
−4
0
500
1000
1500
2000
2500

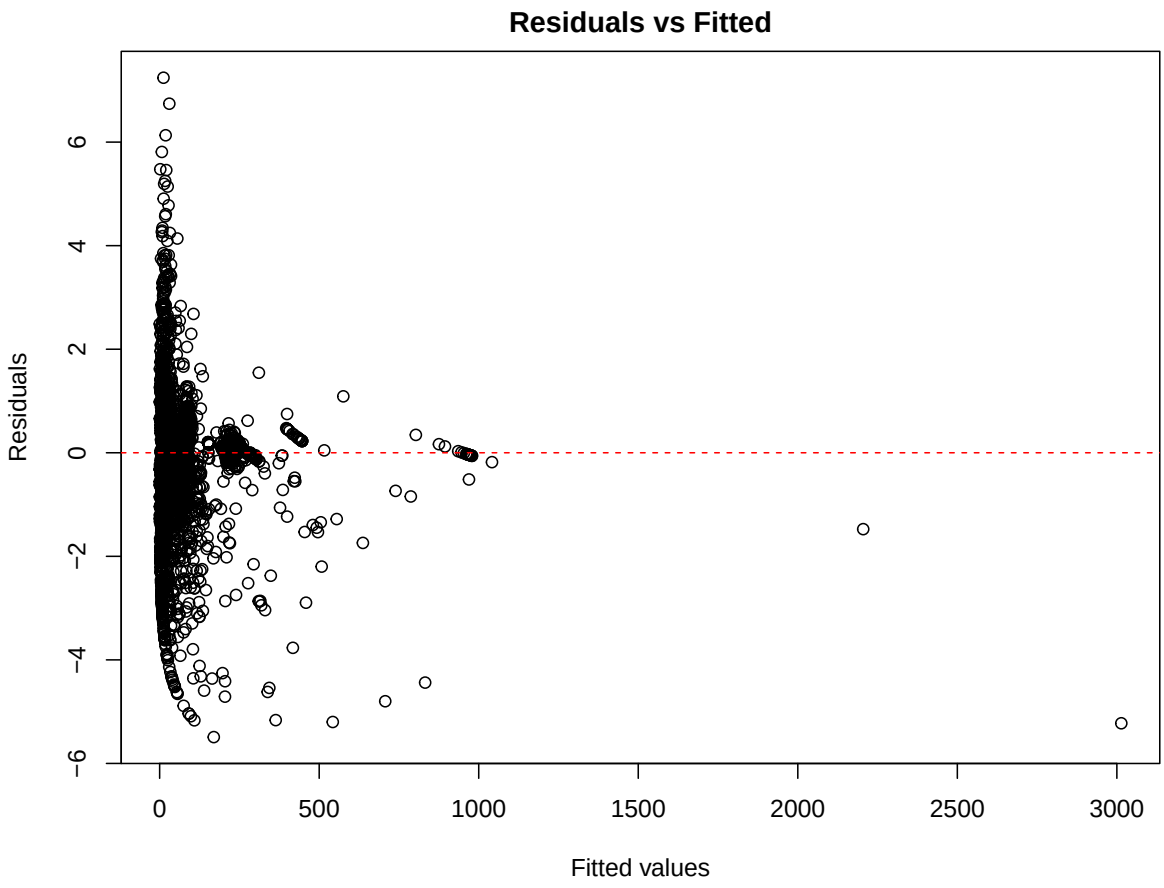

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6
−4
−2
0
2
4

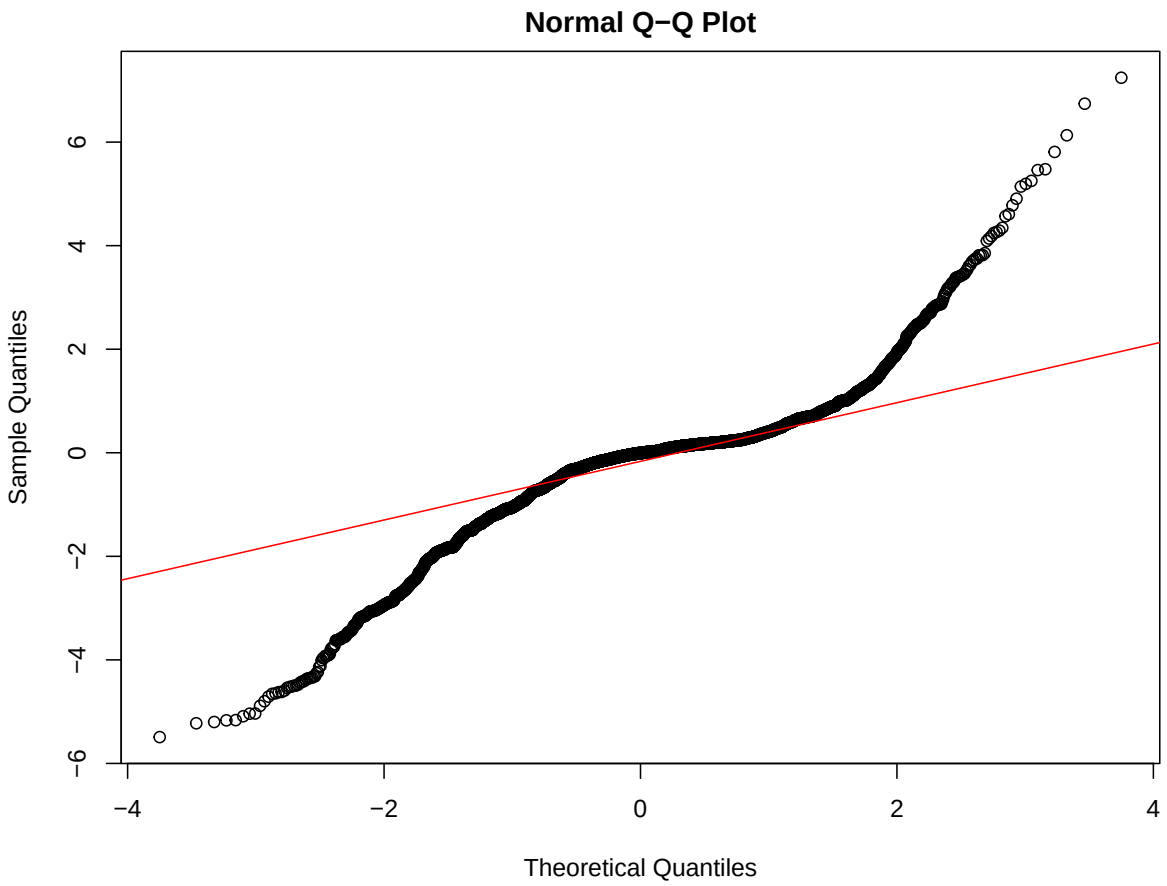

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
2500
Fitted values

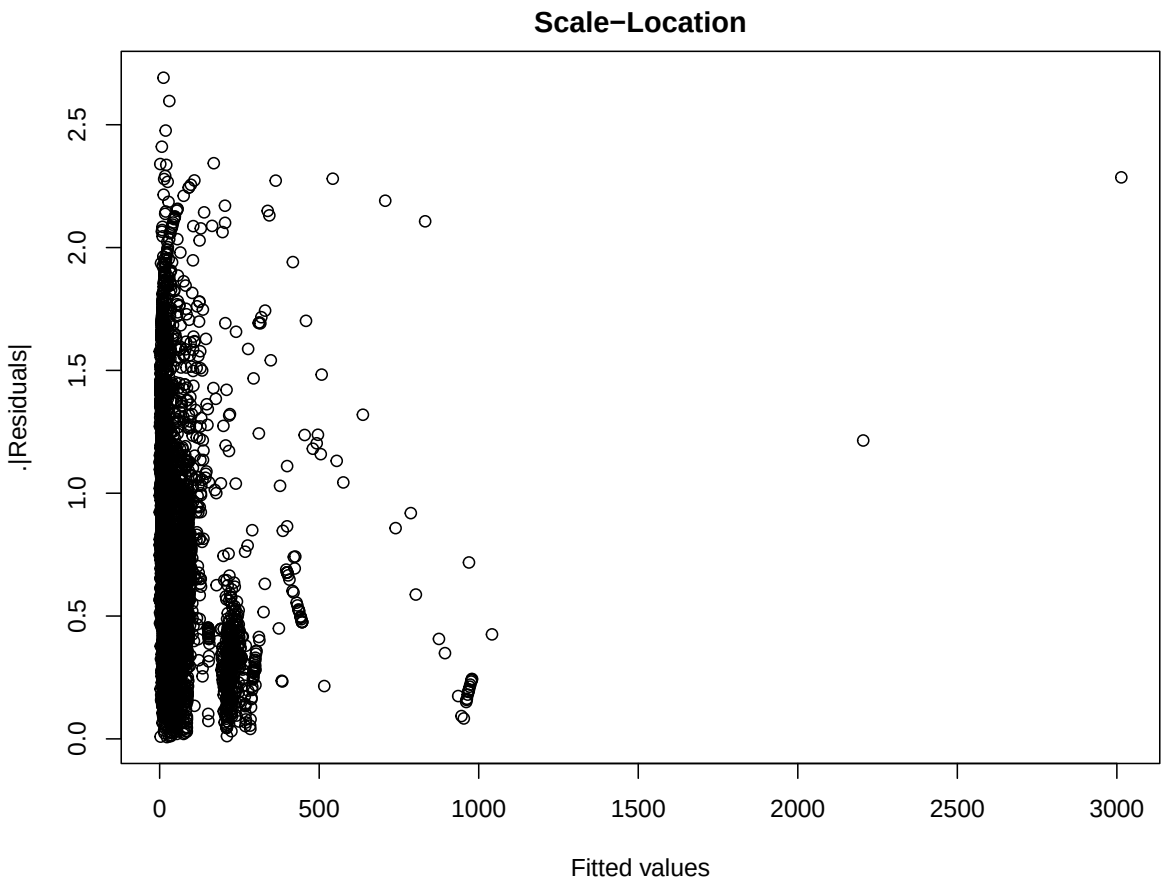


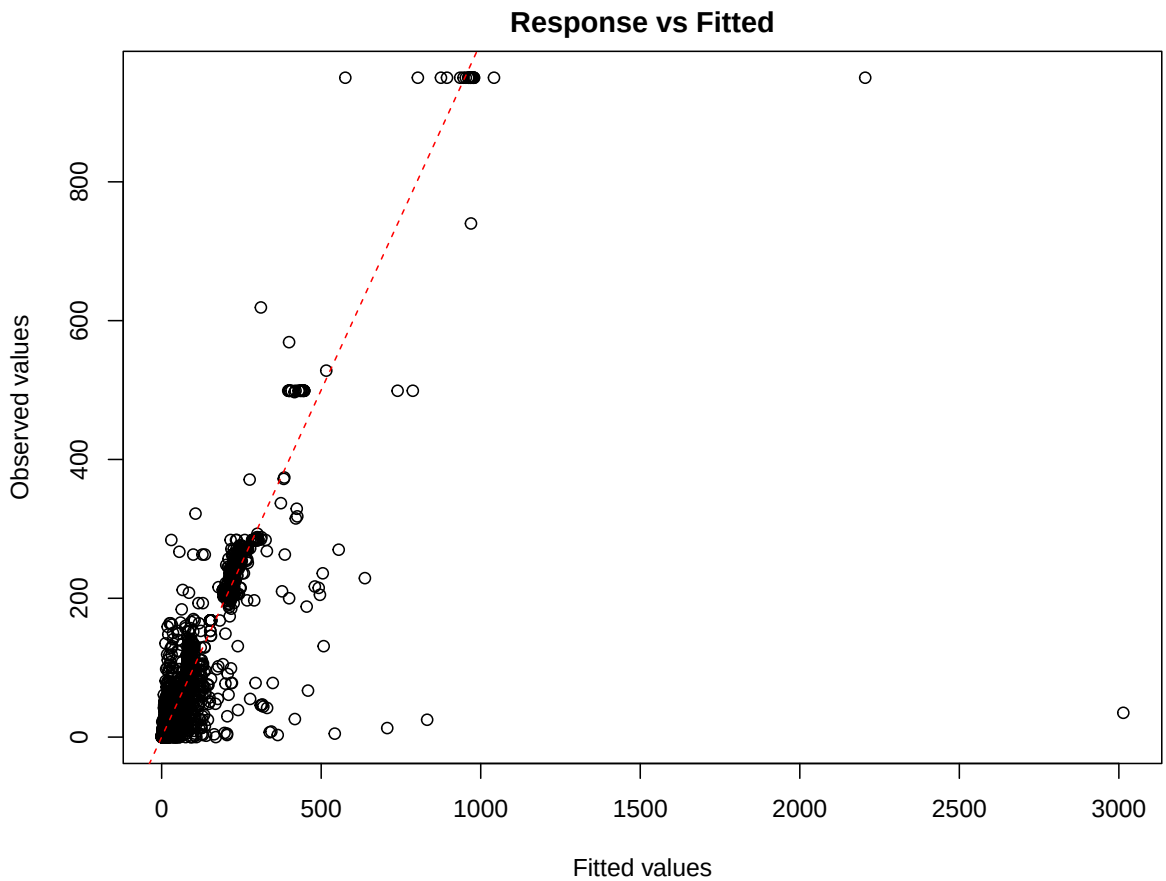

```
### Model 11 : 4c. Temp x Sunlight interaction

Model data dimensions after removing NAs: 5650 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(3.998)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_temp, sunlight_proportion) +
    s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)   3.3844     0.2009   16.84   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                    edf Ref.df   Chi.sq p-value
s(abundance_t_minus_1)            8.831  8.988 18872.43  <2e-16 ***
s(mean_wind_speed)                5.302  6.447    53.80  <2e-16 ***
s(mean_temp)                      8.323  8.832   313.72  <2e-16 ***
s(sunlight_proportion)            6.074  7.208    61.32  <2e-16 ***
ti(mean_temp,sunlight_proportion) 13.010 14.358   202.07  <2e-16 ***
s(view_id)                        7.724  8.000   251.18  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =   0.67   Deviance explained = 80.8%
-REML =  24170  Scale est. = 1         n = 5650


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
6
4
2
0
−2
−4
−6
0
500
1000
1500
2000
2500
3000

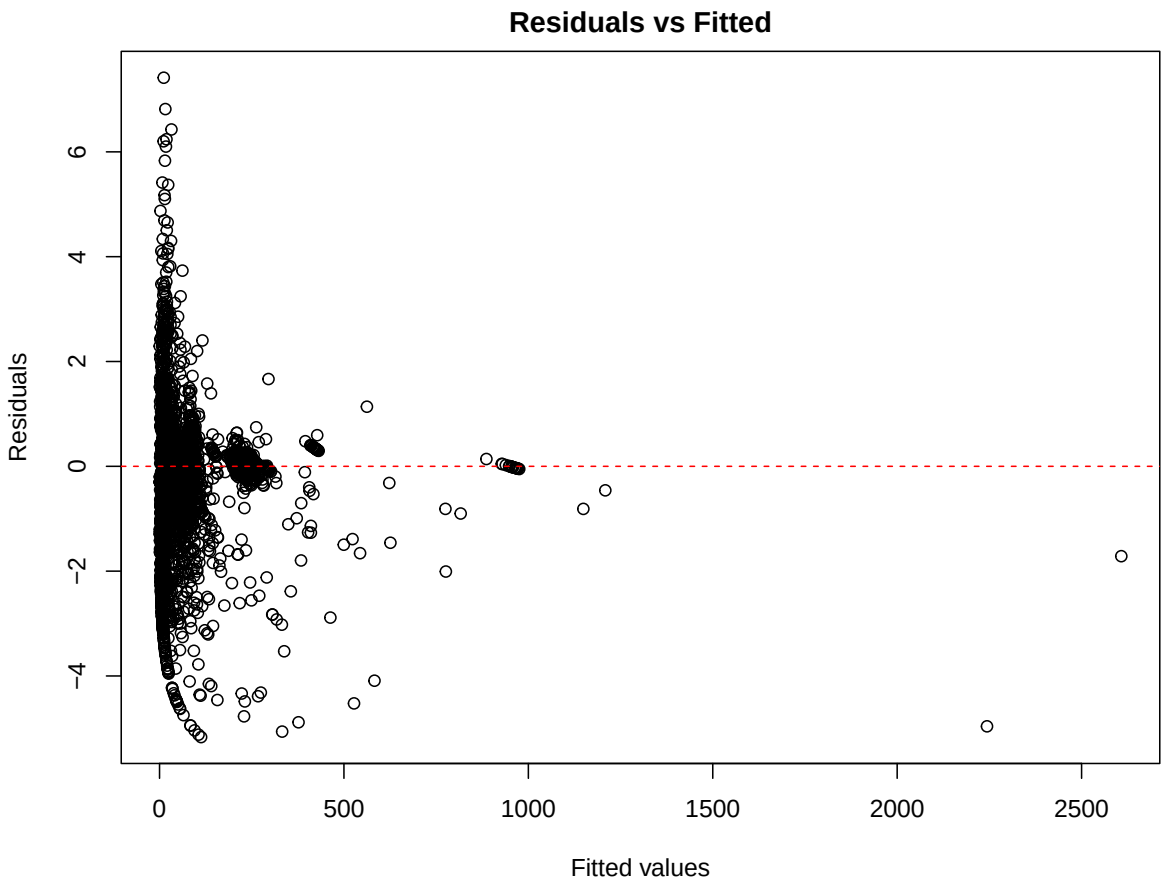

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−6
−4
−2
0
2
4
6
−4
−2
0
2
4

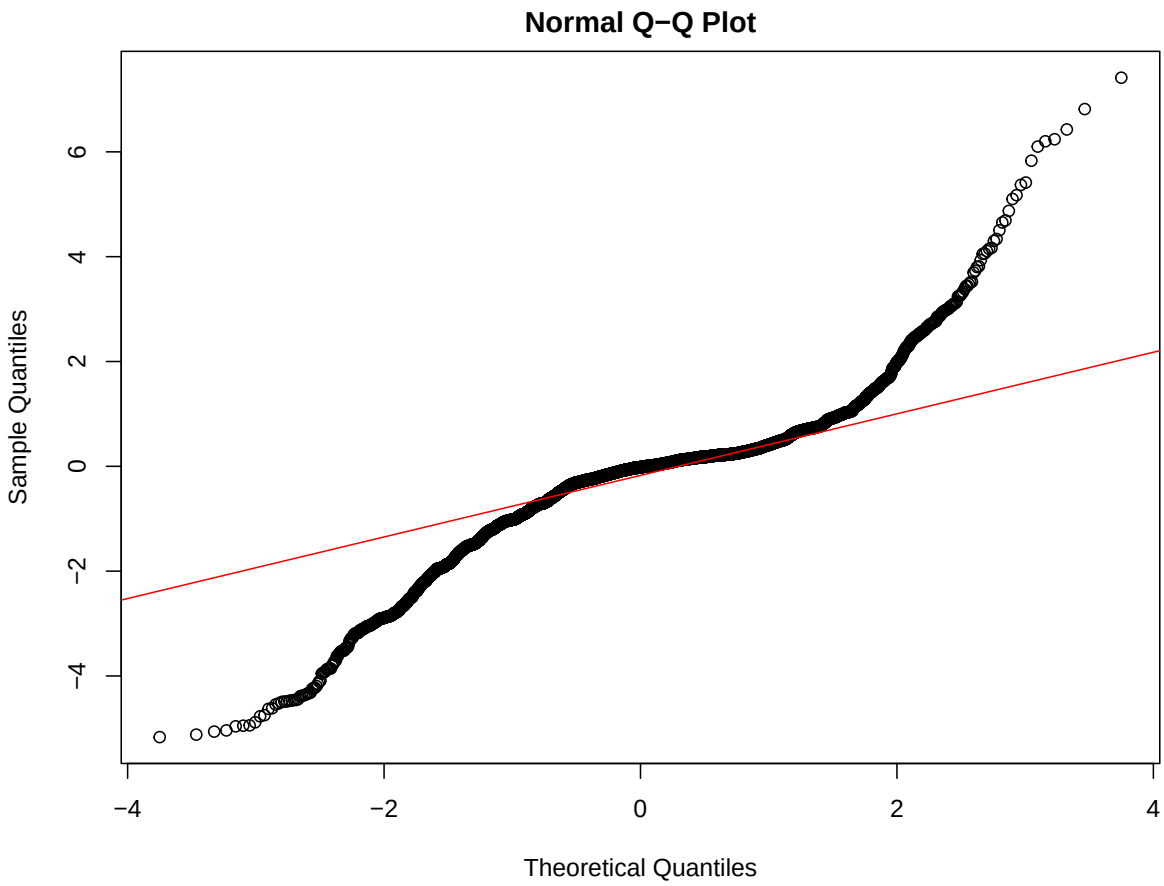

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
2500
3000
Fitted values

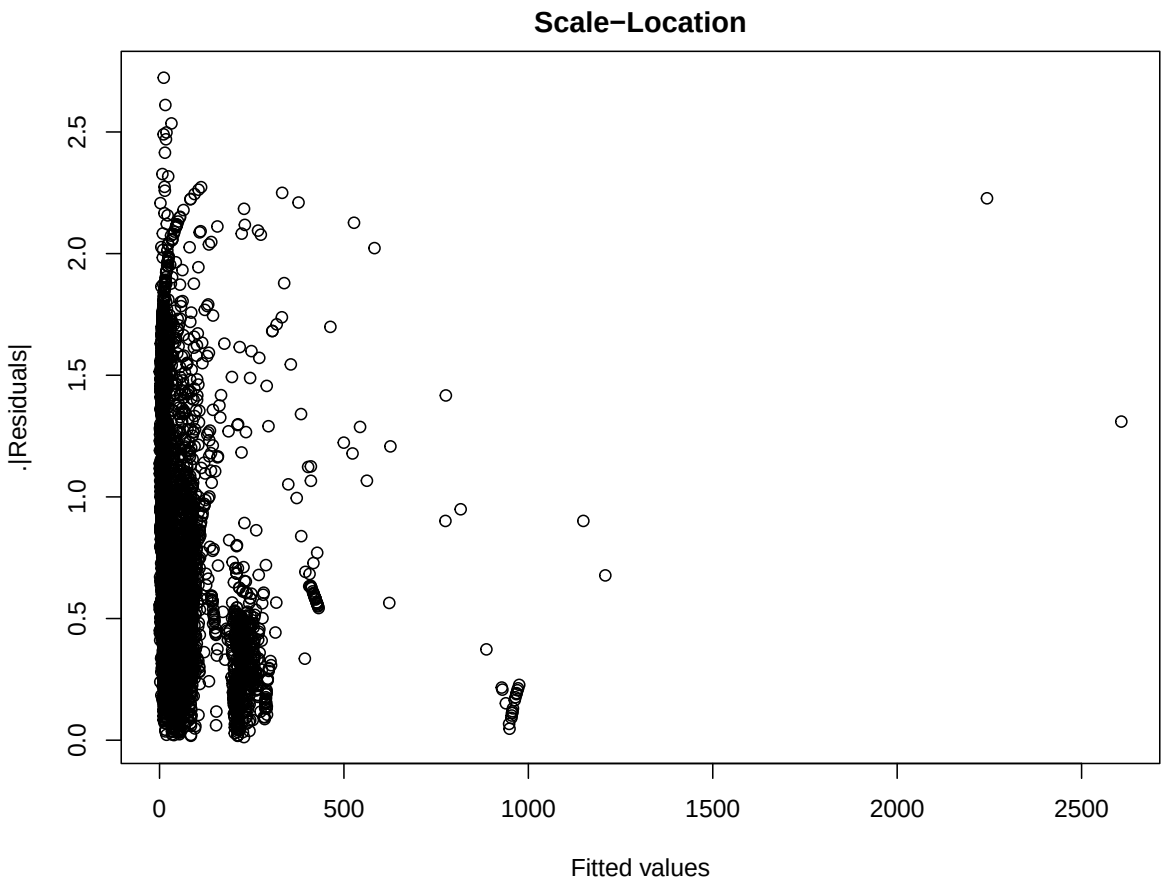


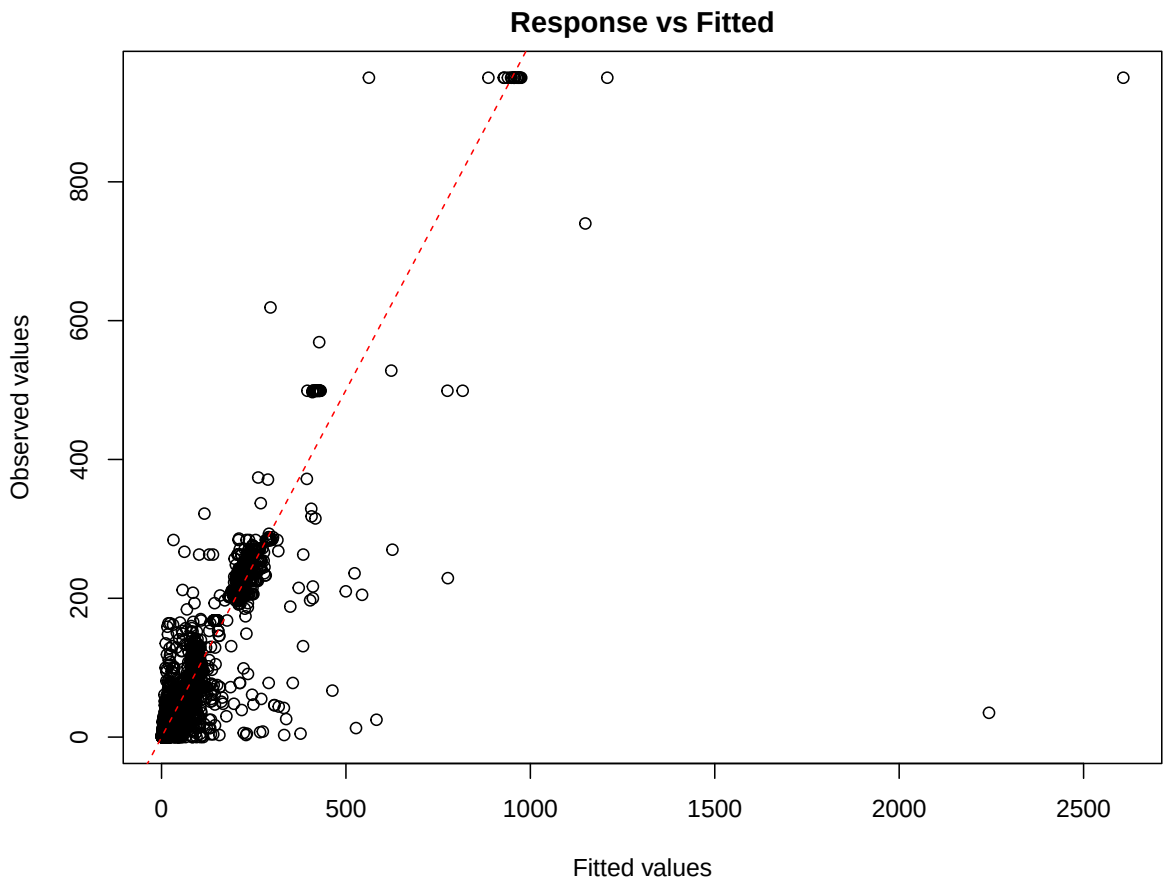

```
### Model 12 : 5a. Wind x Temp + Wind x Sun

Model data dimensions after removing NAs: 5650 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(3.94)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + ti(mean_wind_speed, sunlight_proportion) + s(view_id,
    bs = "re")

Parametric coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3895     0.2095   16.18   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                          edf Ref.df   Chi.sq  p-value
s(abundance_t_minus_1)                  8.832  8.988 17315.74  < 2e-16 ***
s(mean_wind_speed)                      5.578  6.745    31.52 3.46e-05 ***
s(mean_temp)                            8.361  8.855   472.56  < 2e-16 ***
s(sunlight_proportion)                  7.755  8.588   332.88  < 2e-16 ***
ti(mean_wind_speed,mean_temp)           6.374  7.746    34.41 2.16e-05 ***
ti(mean_wind_speed,sunlight_proportion) 10.919 12.737    83.43  < 2e-16 ***
s(view_id)                              7.737  8.000   258.79  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.735   Deviance explained = 80.5%
-REML =  24215  Scale est. = 1         n = 5650


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
6
4
2
0
−2
−4
0
500
1000
1500
2000
2500

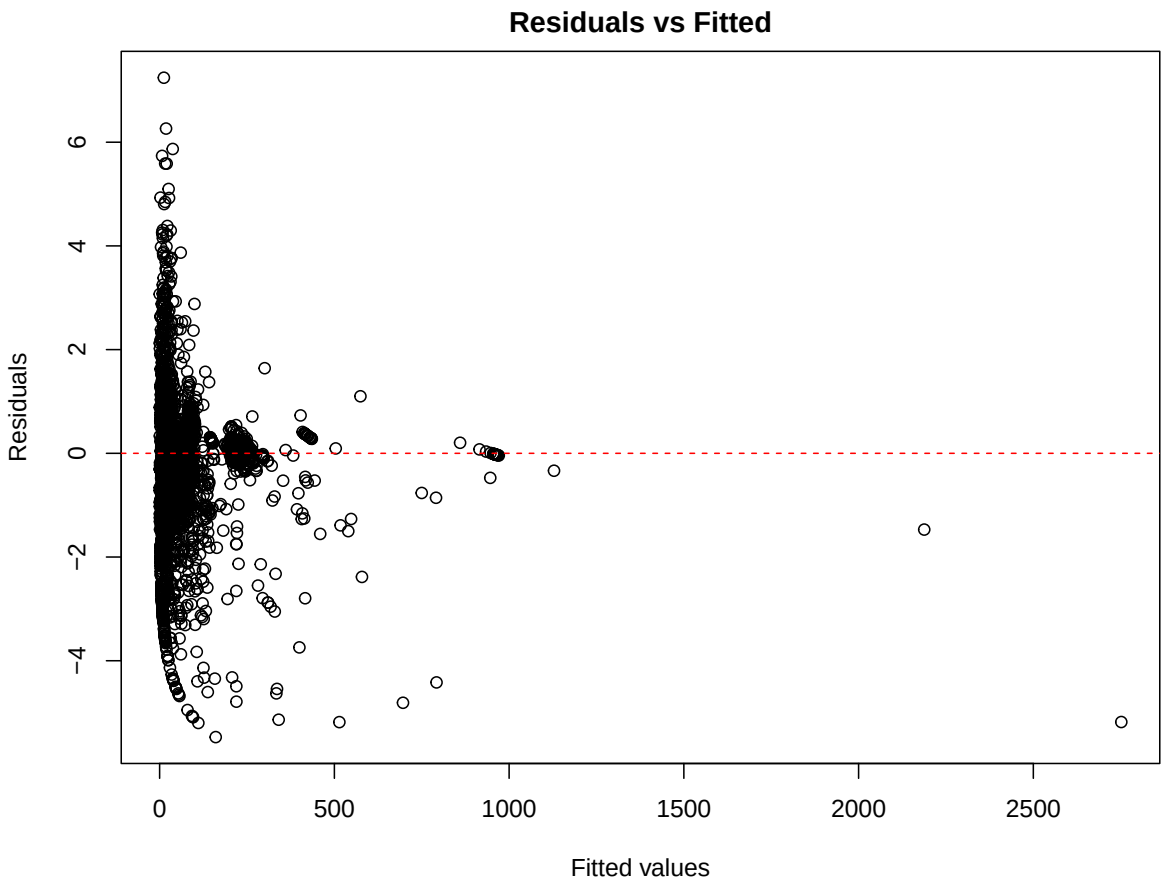

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6
−4
−2
0
2
4

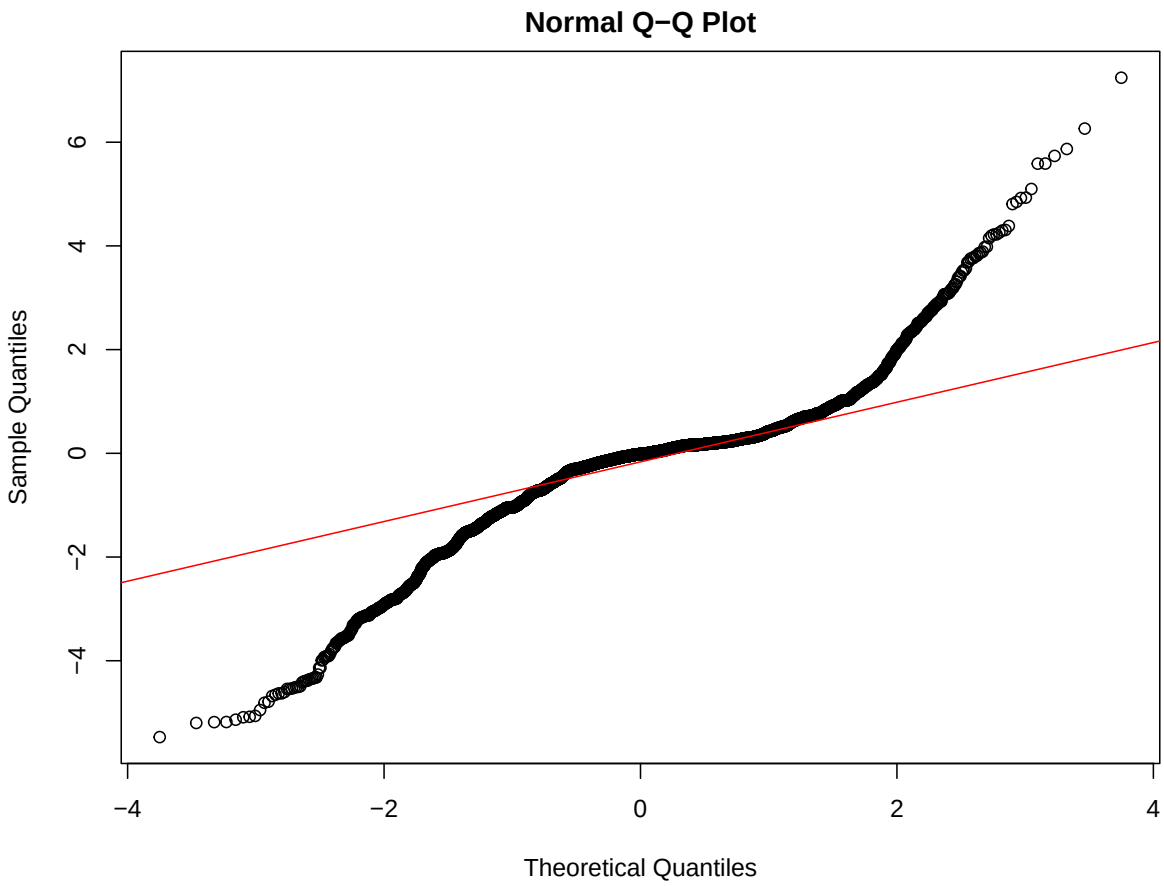

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
2500
Fitted values

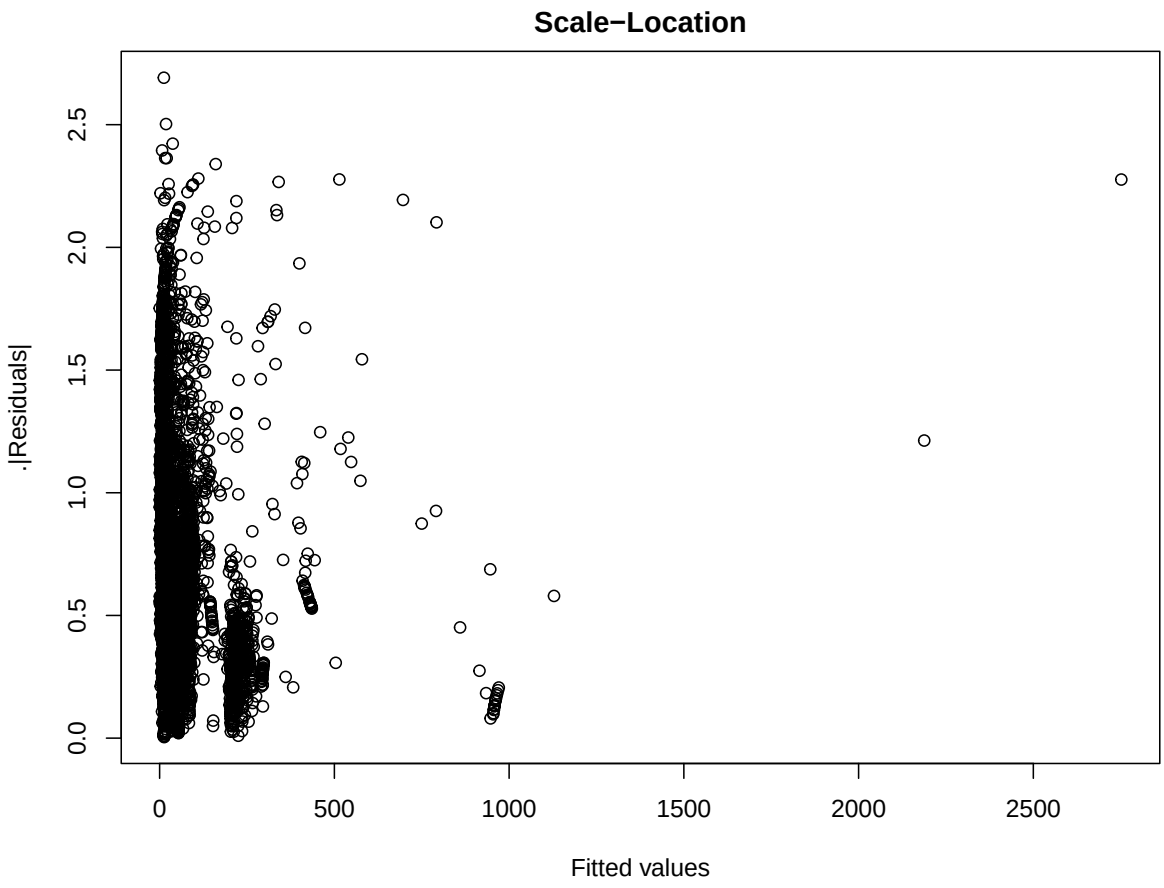


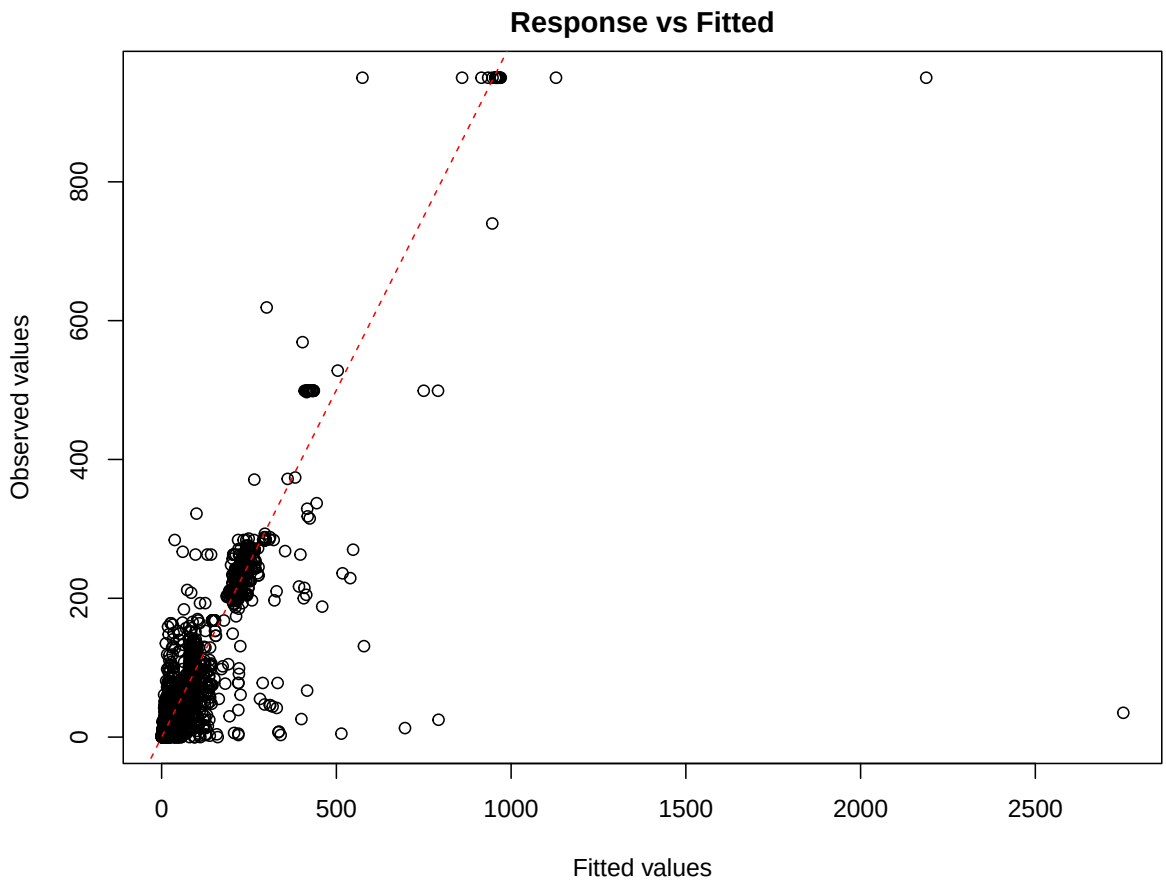

```
### Model 13 : 5b. Wind x Temp + Temp x Sun

Model data dimensions after removing NAs: 5650 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(4.04)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + ti(mean_temp, sunlight_proportion) + s(view_id,
    bs = "re")

Parametric coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3941     0.2069   16.41   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                    edf Ref.df   Chi.sq  p-value
s(abundance_t_minus_1)            8.833  8.988 17923.55  < 2e-16 ***
s(mean_wind_speed)                4.810  5.906    34.74 4.31e-06 ***
s(mean_temp)                      8.424  8.872   310.43  < 2e-16 ***
s(sunlight_proportion)            6.063  7.196    64.96  < 2e-16 ***
ti(mean_wind_speed,mean_temp)     3.706  4.345    38.20 2.04e-07 ***
ti(mean_temp,sunlight_proportion) 12.562 13.935   211.53  < 2e-16 ***
s(view_id)                        7.736  8.000   267.41  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.706   Deviance explained = 80.9%
-REML =  24154  Scale est. = 1         n = 5650


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
6
4
2
0
−2
−4
0
500
1000
1500
2000
2500

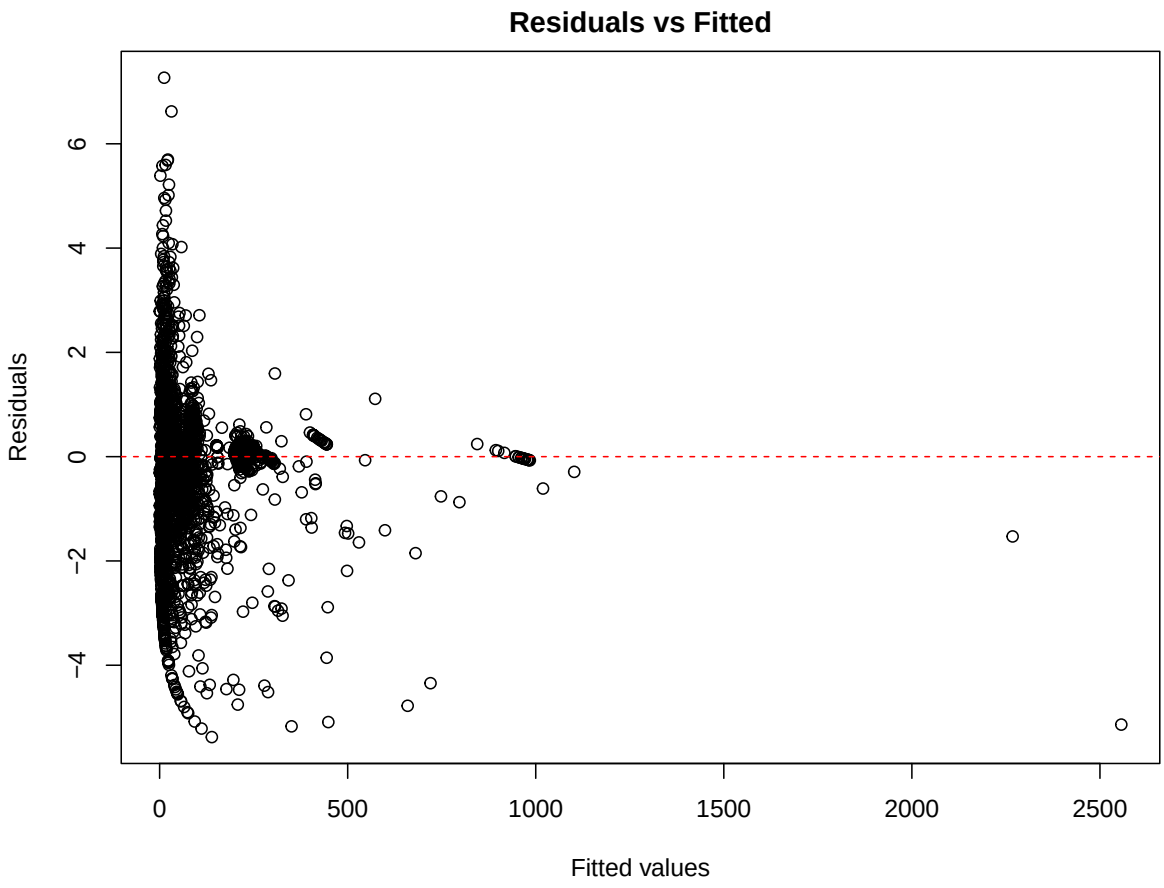

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6
−4
−2
0
2
4

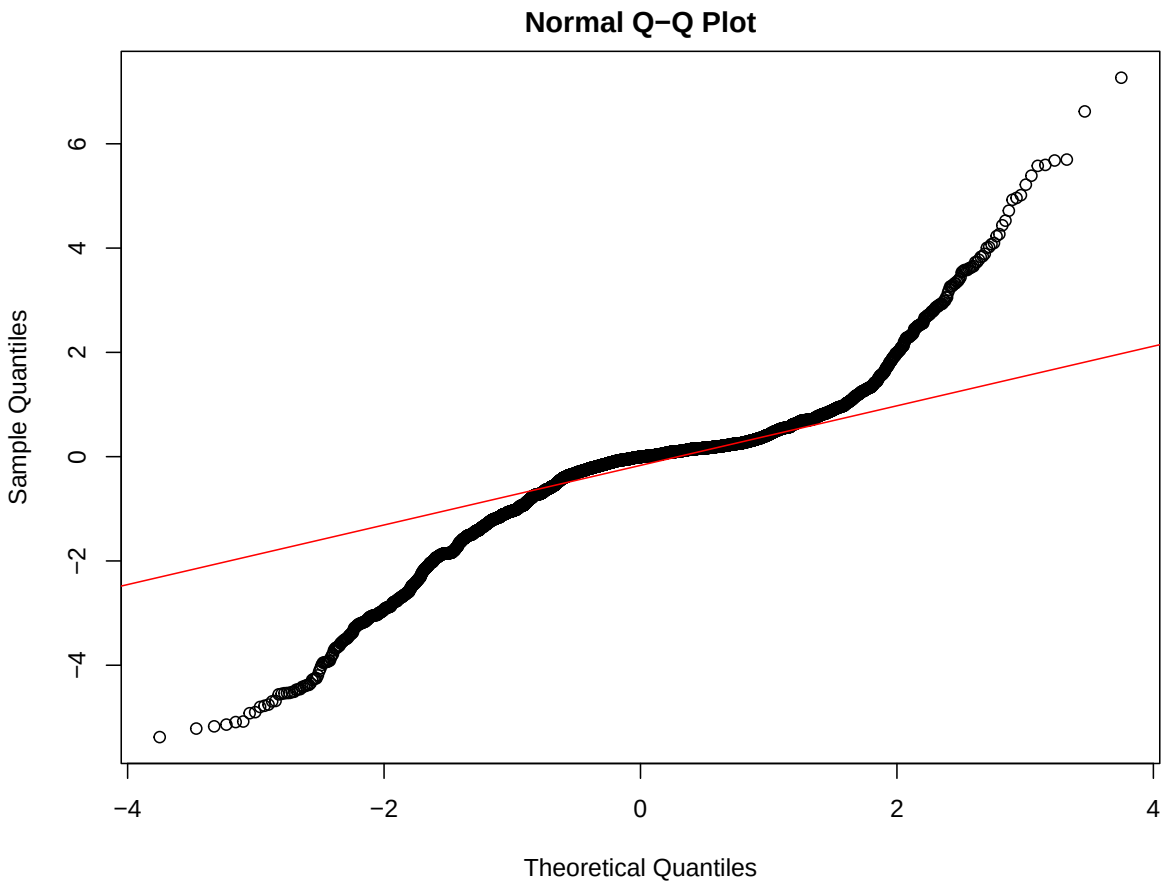

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
2500
Fitted values

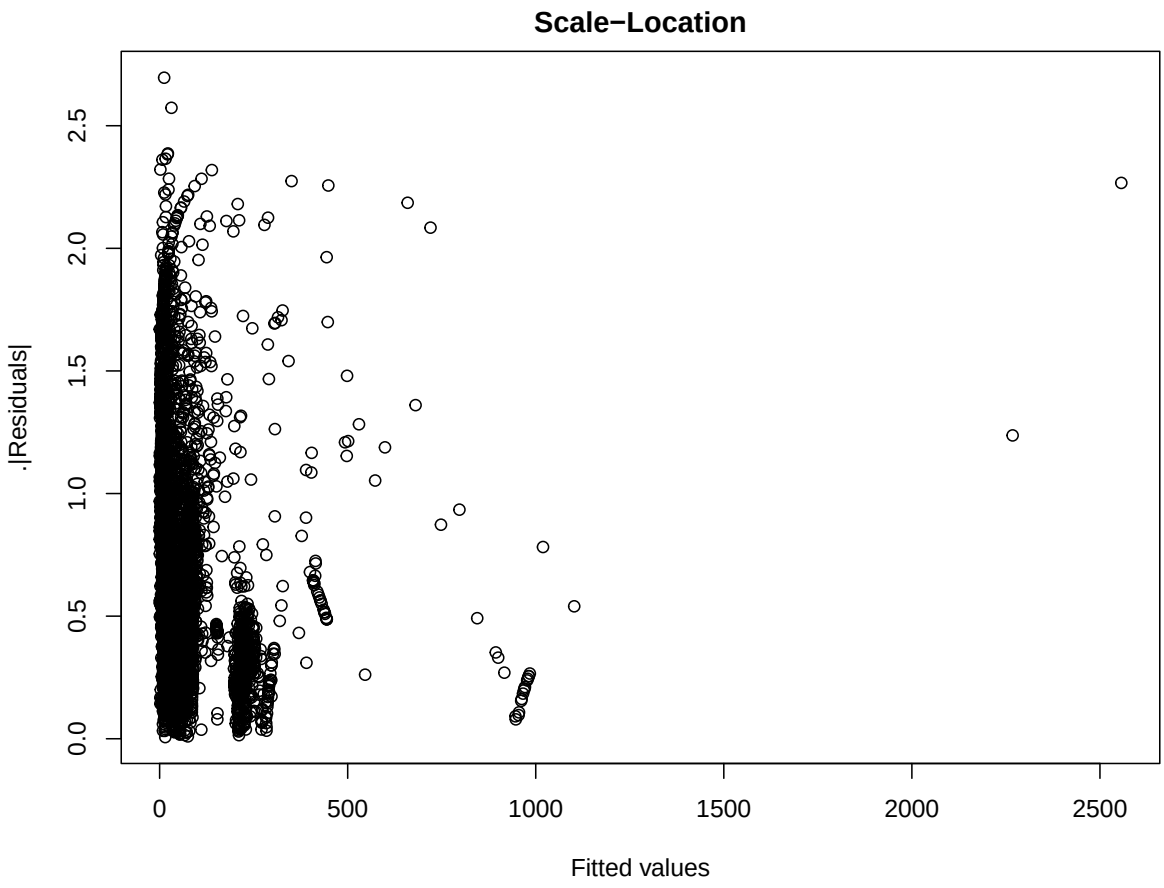


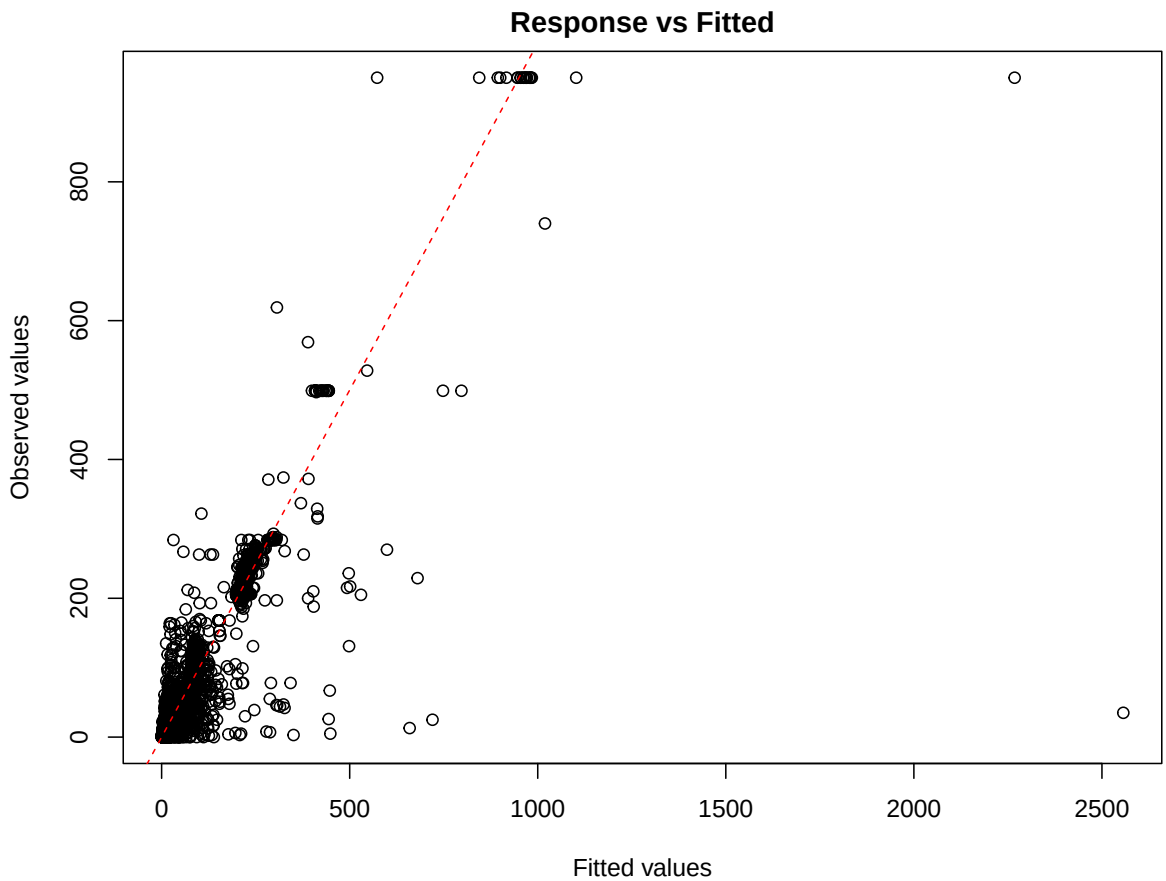

```
### Model 14 : 5c. Wind x Sun + Temp x Sun

Model data dimensions after removing NAs: 5650 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(4.062)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    sunlight_proportion) + ti(mean_temp, sunlight_proportion) +
    s(view_id, bs = "re")

Parametric coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3815     0.2026   16.69   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                          edf Ref.df   Chi.sq p-value
s(abundance_t_minus_1)                  8.832  8.988 19034.78  <2e-16 ***
s(mean_wind_speed)                      5.678  6.849    41.82  <2e-16 ***
s(mean_temp)                            8.371  8.852   330.67  <2e-16 ***
s(sunlight_proportion)                  5.662  6.785    54.40  <2e-16 ***
ti(mean_wind_speed,sunlight_proportion) 10.323 12.167    61.98  <2e-16 ***
ti(mean_temp,sunlight_proportion)       12.699 14.049   187.52  <2e-16 ***
s(view_id)                              7.727  8.000   256.03  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.721   Deviance explained =   81%
-REML =  24152  Scale est. = 1         n = 5650


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
6
4
2
0
−2
−4
0
500
1000
1500
2000
2500

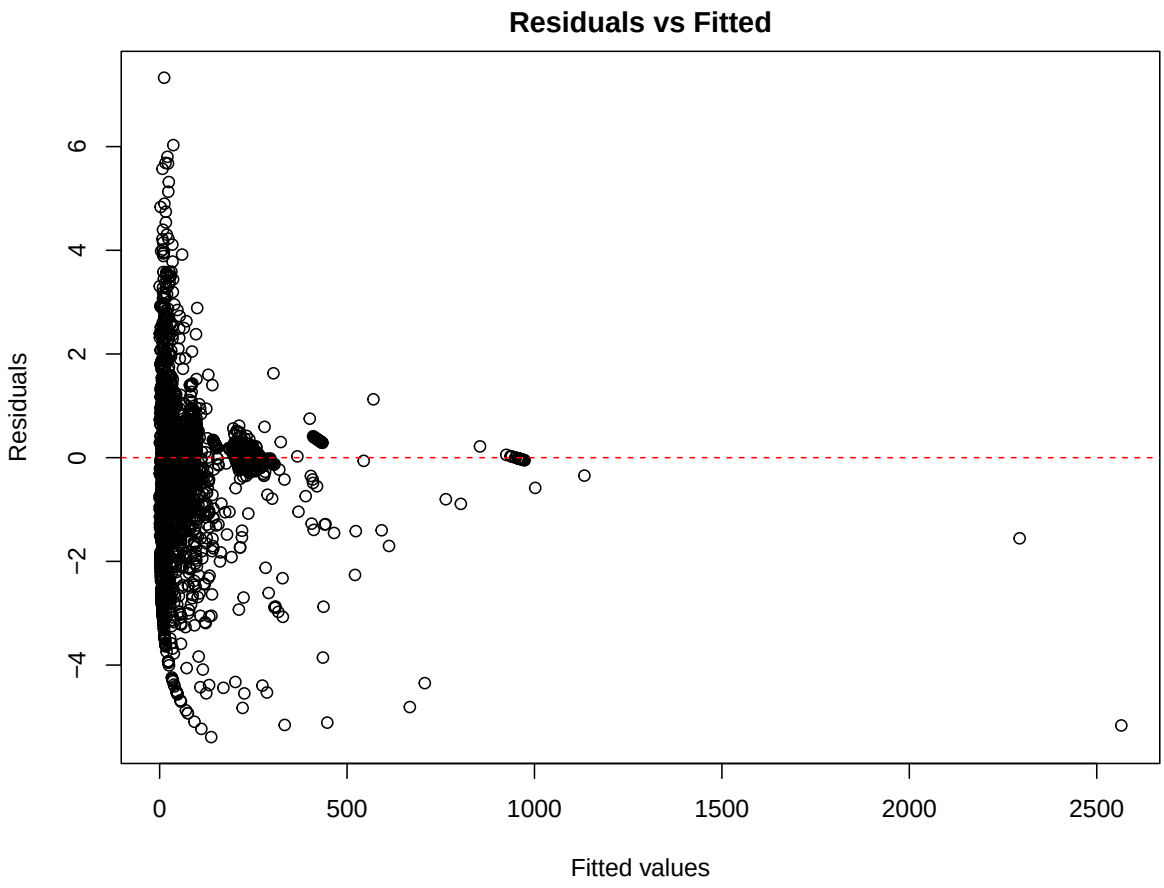

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6
−4
−2
0
2
4

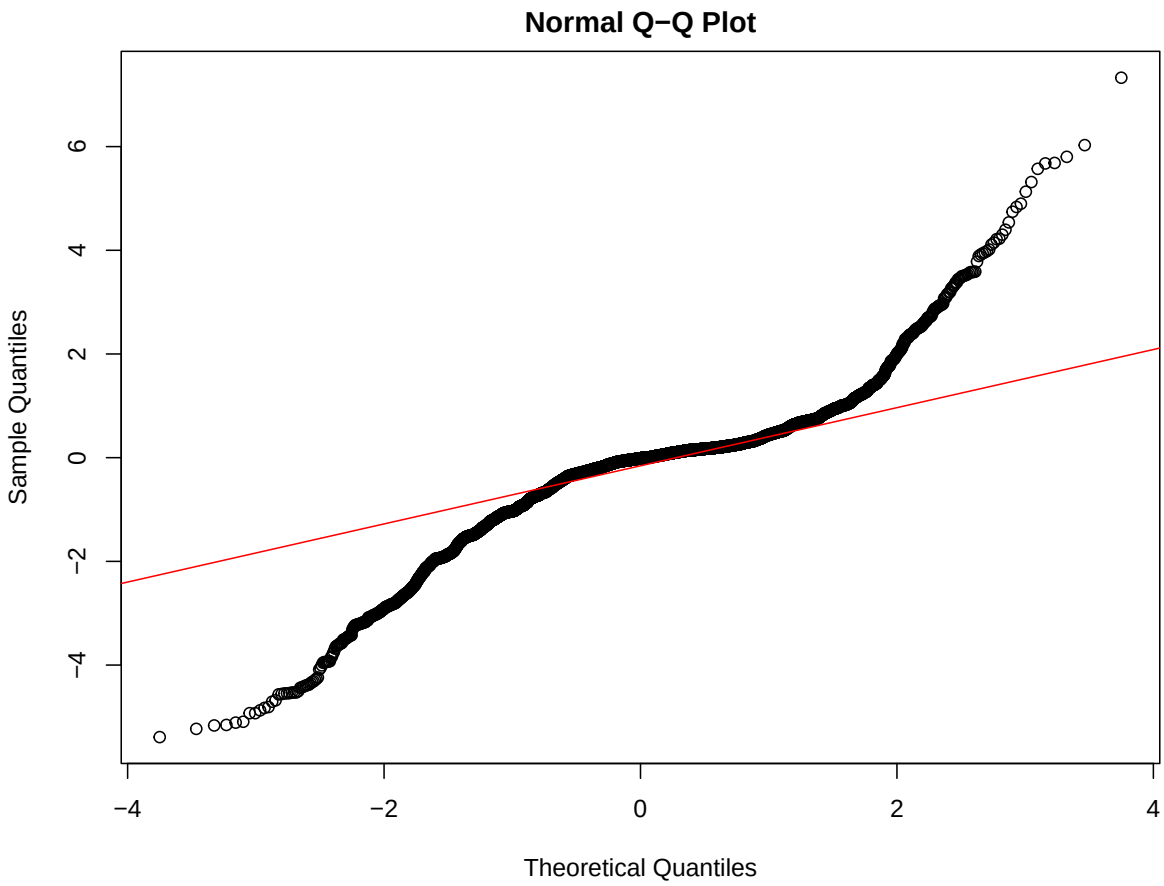

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
2500
Fitted values

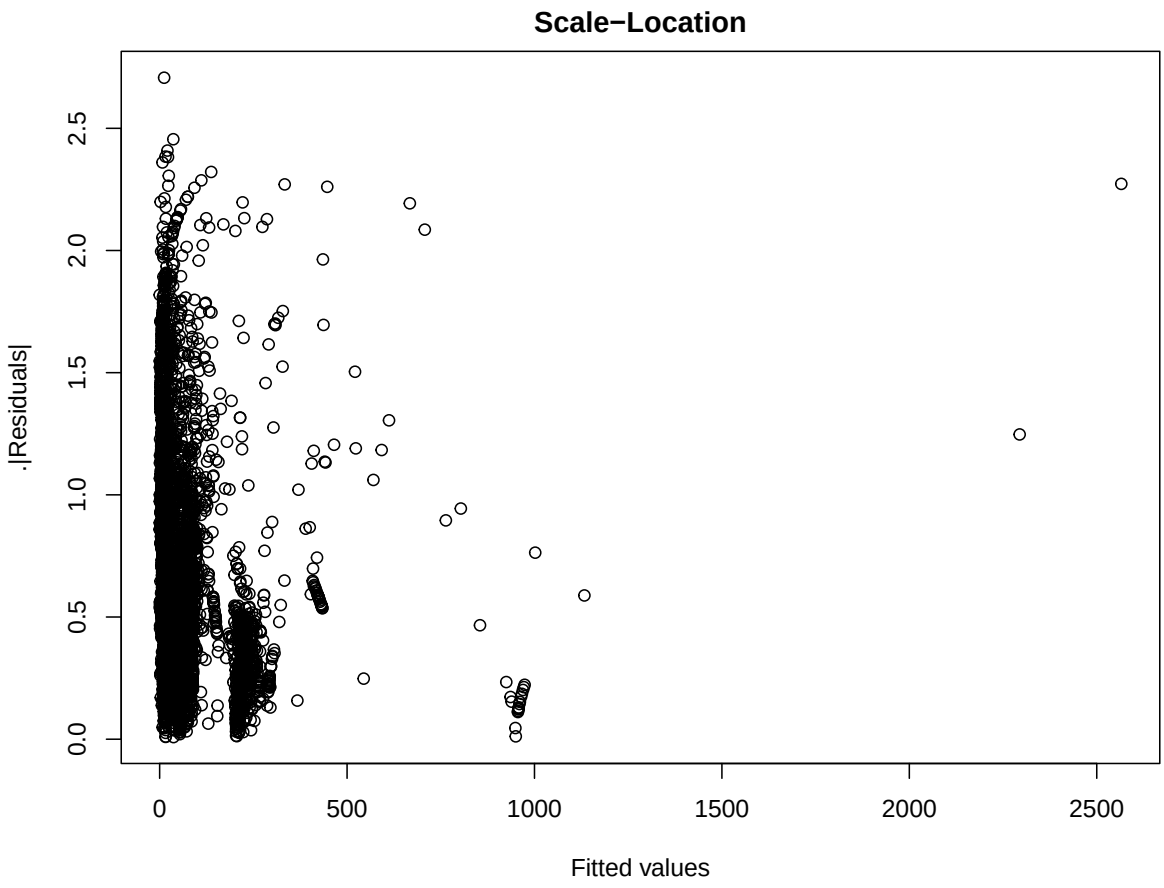


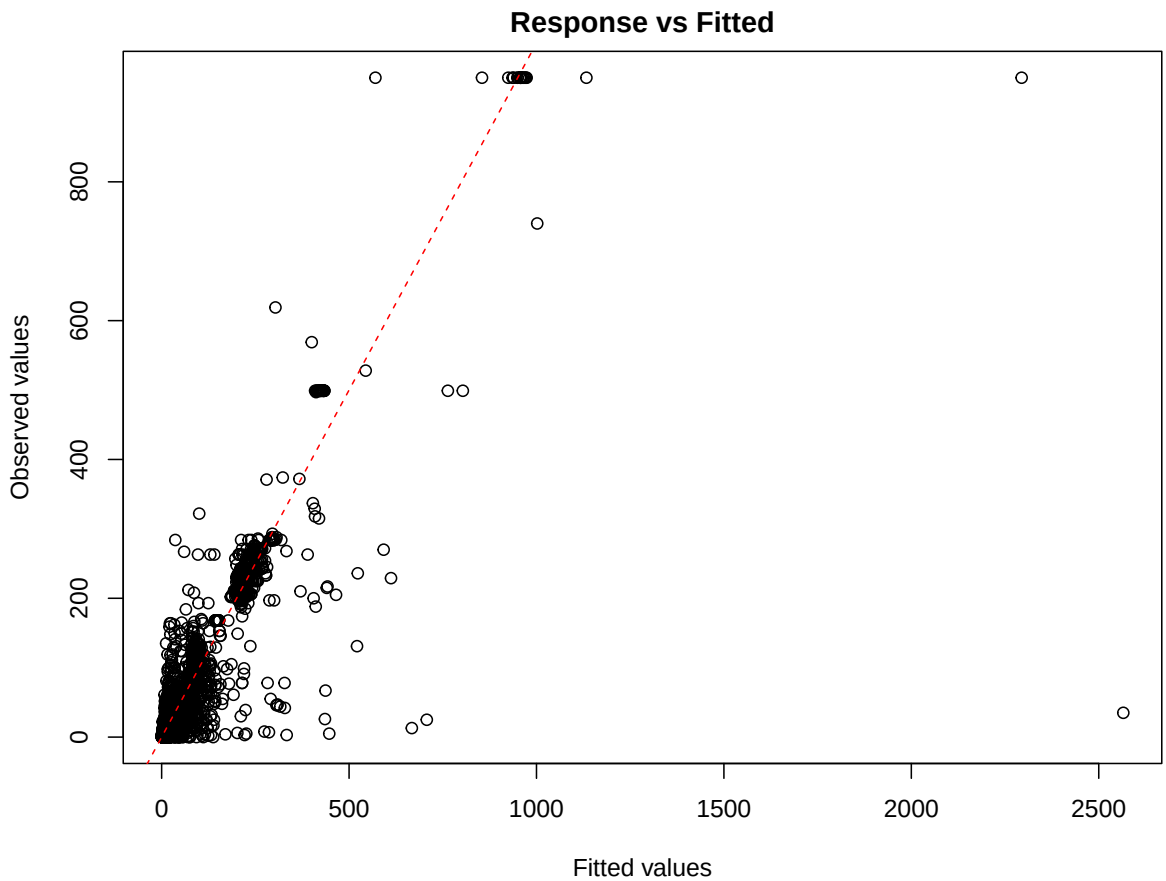

```
### Model 15 : 6. Global model

Model data dimensions after removing NAs: 5650 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(4.1)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + ti(mean_wind_speed, sunlight_proportion) + ti(mean_temp,
    sunlight_proportion) + s(view_id, bs = "re")

Parametric coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3886     0.2049   16.54   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                          edf Ref.df   Chi.sq  p-value
s(abundance_t_minus_1)                  8.837  8.989 17808.07  < 2e-16 ***
s(mean_wind_speed)                      5.770  6.939    41.54  < 2e-16 ***
s(mean_temp)                            8.420  8.870   311.15  < 2e-16 ***
s(sunlight_proportion)                  5.815  6.941    55.24  < 2e-16 ***
ti(mean_wind_speed,mean_temp)           5.424  6.641    35.30 3.79e-06 ***
ti(mean_wind_speed,sunlight_proportion) 10.468 12.291    54.46 5.16e-06 ***
ti(mean_temp,sunlight_proportion)      12.252 13.641   183.86  < 2e-16 ***
s(view_id)                              7.731  8.000   262.74  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.722   Deviance explained = 81.2%
-REML =  24139  Scale est. = 1         n = 5650


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
6
4
2
0
−2
−4
0
500
1000
1500
2000
2500


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6
−4
−2
0
2
4


Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
2500
Fitted values




```r
# Model comparison table
comparison_120m <- data.frame(
  Model = names(aic_values_120m),
  AIC = round(aic_values_120m, 2)
) %>%
  arrange(AIC)

cat("\n### Model Comparison (2-hour lag)\n")
```

```
### Model Comparison (2-hour lag)
```

```r
kable(comparison_120m)
```

	Model	AIC
6. Global model	6. Global model	48104.98
5c. Wind x Sun + Temp x Sun	5c. Wind x Sun + Temp x Sun	48132.33
5b. Wind x Temp + Temp x Sun	5b. Wind x Temp + Temp x Sun	48148.72
4c. Temp x Sunlight interaction	4c. Temp x Sunlight interaction	48178.96
5a. Wind x Temp + Wind x Sun	5a. Wind x Temp + Wind x Sun	48269.90
4b. Wind x Sunlight interaction	4b. Wind x Sunlight interaction	48294.28
4a. Wind x Temp interaction	4a. Wind x Temp interaction	48342.66
3d. All additive	3d. All additive	48365.21
3c. Temp + Sunlight	3c. Temp + Sunlight	48400.93
3a. Wind + Temp	3a. Wind + Temp	48761.57
2b. Temperature only	2b. Temperature only	48785.36
3b. Wind + Sunlight	3b. Wind + Sunlight	48906.77
2c. Sunlight only	2c. Sunlight only	48958.71
2a. Wind only	2a. Wind only	49266.06
1. Baseline	1. Baseline	49308.57

4-Hour Lag Analysis

```r
data_240m <- prepare_lag_data(master_df, wind, lag_minutes = 240)
```

```
Preparing data for 240 minute lag...
Valid pairs after filtering: 6894
Calculating interval predictors efficiently...
Processing 6894 unique intervals...
Progress: 10 %
Progress: 20 %
Progress: 30 %
Progress: 40 %
Progress: 50 %
Progress: 60 %
Progress: 70 %
Progress: 80 %
Progress: 89.9 %
Progress: 99.9 %
Combining results...
Final dataset rows: 5663
Time delta range: 240 to 240 minutes
```

Exploratory Data Analysis (4 hours)

```r
cat("Data structure:\n")
```

```
Data structure:
```

```r
glimpse(data_240m)
```

```
Rows: 5,663
Columns: 30
$ deployment_id.x           <chr> "SC1", "SC1", "SC1", "SC1", "SC1", "SC1", "S~
$ image_filename            <chr> "SC1_20231117184501.JPG", "SC1_2023111718500~
$ total_butterflies         <dbl> 54, 54, 54, 54, 54, 54, 54, 54, 54, 54, 54, ~
$ butterflies_direct_sun    <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ datetime                  <dttm> 2023-11-17 18:45:01, 2023-11-17 18:50:01, 2~
$ view_id                   <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ wind_meter_name           <chr> "JazzPlay", "JazzPlay", "JazzPlay", "JazzPla~
$ deployment_id.y           <chr> "SC1", "SC1", "SC1", "SC1", "SC1", "SC1", "S~
$ timestamp                 <dbl> 2.023112e+13, 2.023112e+13, 2.023112e+13, 2.~
$ temperature               <dbl> 15, 16, 15, 16, 15, 16, 16, 15, 16, 15, 16, ~
$ confidence                <dbl> 0.4041873, 0.5208792, 0.6782328, 0.5311629, ~
$ extraction_status         <chr> "success", "success", "success", "success", ~
$ datetime_t_minus_1        <dttm> 2023-11-17 14:45:01, 2023-11-17 14:50:01, 2~
$ abundance_t_minus_1       <dbl> 90, 80, 95, 58, 56, 56, 64, 40, 48, 59, 79, ~
$ butterflies_sun_t_minus_1 <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ temperature_t_minus_1     <dbl> 16, 16, 16, 16, 15, 16, 15, 15, 15, 15, 15, ~
$ time_delta_mins           <dbl> 240, 240, 240, 240, 240, 240, 240, 240, 240,~
$ mean_wind_speed           <dbl> 0.400000, 0.900000, 1.030769, 1.177778, 1.24~
$ max_wind_speed            <dbl> 0.4, 1.9, 1.9, 1.9, 1.9, 2.3, 2.7, 2.9, 2.9,~
$ sd_wind_speed             <dbl> 0.0000000, 0.6524678, 0.5297556, 0.5241863, ~
$ gust_factor               <dbl> 2.250000, 1.722222, 1.656716, 1.622642, 1.58~
$ minutes_above_2mps        <dbl> 0, 0, 0, 0, 0, 1, 3, 6, 7, 12, 16, 20, 25, 3~
$ mean_temp                 <dbl> 15.12245, 15.12245, 15.10204, 15.10204, 15.0~
$ max_temp                  <dbl> 16, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16, ~
$ min_temp                  <dbl> 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, ~
$ sd_temp                   <dbl> 0.3312007, 0.3312007, 0.3058389, 0.3058389, ~
$ sunlight_proportion       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ max_prop_sunlight         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ min_prop_sunlight         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ sd_prop_sunlight          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
```

```r
cat("\n\nData summary:\n")
```

```
Data summary:
```

```r
summary(data_240m)
```

```
 deployment_id.x    image_filename     total_butterflies butterflies_direct_sun
 Length:5663        Length:5663        Min.   :  0.00    Min.   :  0.000
 Class :character   Class :character   1st Qu.: 12.00    1st Qu.:  0.000
 Mode  :character   Mode  :character   Median : 34.00    Median :  0.000
                                       Mean   : 64.02    Mean   :  1.585
                                       3rd Qu.: 72.00    3rd Qu.:  0.000
                                       Max.   :950.00    Max.   :150.000

    datetime                          view_id     wind_meter_name
 Min.   :2023-11-17 18:45:01.00   4      :2895   Length:5663
 1st Qu.:2023-12-10 17:30:01.00   2      :1115   Class :character
 Median :2023-12-28 22:50:01.00   1      : 674   Mode  :character
 Mean   :2023-12-28 21:36:27.57   5      : 464
 3rd Qu.:2024-01-23 05:05:01.00   11     : 287
 Max.   :2024-02-04 10:50:01.00   9      : 165
                                  (Other):  63
 deployment_id.y      timestamp          temperature     confidence
 Length:5663        Min.   :2.023e+13   Min.   : 4.0   Min.   :0.1222
 Class :character   1st Qu.:2.023e+13   1st Qu.:11.0   1st Qu.:0.4971
 Mode  :character   Median :2.023e+13   Median :13.0   Median :0.5940
                    Mean   :2.024e+13   Mean   :13.3   Mean   :0.5959
                    3rd Qu.:2.024e+13   3rd Qu.:15.0   3rd Qu.:0.6945
                    Max.   :2.024e+13   Max.   :33.0   Max.   :1.0000

 extraction_status  datetime_t_minus_1               abundance_t_minus_1
 Length:5663        Min.   :2023-11-17 14:45:01.00   Min.   :  0.0
 Class :character   1st Qu.:2023-12-10 13:30:01.00   1st Qu.: 14.0
 Mode  :character   Median :2023-12-28 18:50:01.00   Median : 35.0
                    Mean   :2023-12-28 17:36:27.57   Mean   : 65.2
                    3rd Qu.:2024-01-23 01:05:01.00   3rd Qu.: 75.0
                    Max.   :2024-02-04 06:50:01.00   Max.   :950.0

 butterflies_sun_t_minus_1 temperature_t_minus_1 time_delta_mins
```

```
 Min.   :  0.000           Min.   : 4.00         Min.   :240
 1st Qu.:  0.000           1st Qu.:11.00         1st Qu.:240
 Median :  0.000           Median :13.00         Median :240
 Mean   :  1.668           Mean   :13.32         Mean   :240
 3rd Qu.:  0.000           3rd Qu.:15.00         3rd Qu.:240
 Max.   :150.000           Max.   :31.00         Max.   :240

 mean_wind_speed  max_wind_speed  sd_wind_speed     gust_factor
 Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :  1.205
 1st Qu.:0.3140   1st Qu.:1.200   1st Qu.:0.2626   1st Qu.:  1.363
 Median :0.7368   Median :1.800   Median :0.3848   Median :  1.442
 Mean   :0.8833   Mean   :1.952   Mean   :0.3932   Mean   :  1.633
 3rd Qu.:1.2775   3rd Qu.:2.500   3rd Qu.:0.5087   3rd Qu.:  1.559
 Max.   :4.4979   Max.   :6.700   Max.   :1.5738   Max.   :102.000
                                  NA's   :1        NA's   :318
 minutes_above_2mps   mean_temp         max_temp     min_temp
 Min.   :  0.00     Min.   : 4.556   Min.   : 5   Min.   : 4.00
 1st Qu.:  0.00     1st Qu.:11.160   1st Qu.:12   1st Qu.:10.00
 Median :  0.00     Median :13.000   Median :15   Median :12.00
 Mean   : 22.85     Mean   :13.300   Mean   :15   Mean   :11.76
 3rd Qu.: 12.00     3rd Qu.:15.444   3rd Qu.:17   3rd Qu.:14.00
 Max.   :240.00     Max.   :28.333   Max.   :33   Max.   :26.00

    sd_temp       sunlight_proportion max_prop_sunlight min_prop_sunlight
 Min.   :0.0000   Min.   :0.00000     Min.   :0.0000    Min.   :0.00000
 1st Qu.:0.5000   1st Qu.:0.00000     1st Qu.:0.0000    1st Qu.:0.00000
 Median :0.8327   Median :0.00000     Median :0.0000    Median :0.00000
 Mean   :1.1158   Mean   :0.06910     Mean   :0.2060    Mean   :0.01889
 3rd Qu.:1.3944   3rd Qu.:0.03934     3rd Qu.:0.2308    3rd Qu.:0.00000
 Max.   :7.3314   Max.   :1.00000     Max.   :1.0000    Max.   :1.00000

 sd_prop_sunlight
 Min.   :0.00000
 1st Qu.:0.00000
 Median :0.00000
 Mean   :0.06631
 3rd Qu.:0.05261
 Max.   :0.70711
 NA's   :21
```

```r
# Check for data quality issues
cat("\n\nData quality checks:\n")
```

```
Data quality checks:
```

```r
cat("Total rows:", nrow(data_240m), "\n")
```

```
Total rows: 5663
```

```r
cat("Complete cases:", sum(complete.cases(data_240m)), "\n")
```

```
Complete cases: 5324
```

```r
cat("Rows with abundance_t_minus_1 > 0:", sum(data_240m$abundance_t_minus_1 > 0, na.rm = TRU
```

```
Rows with abundance_t_minus_1 > 0: 5508
```

```r
cat("Rows with total_butterflies > 0:", sum(data_240m$total_butterflies > 0, na.rm = TRUE), "
```

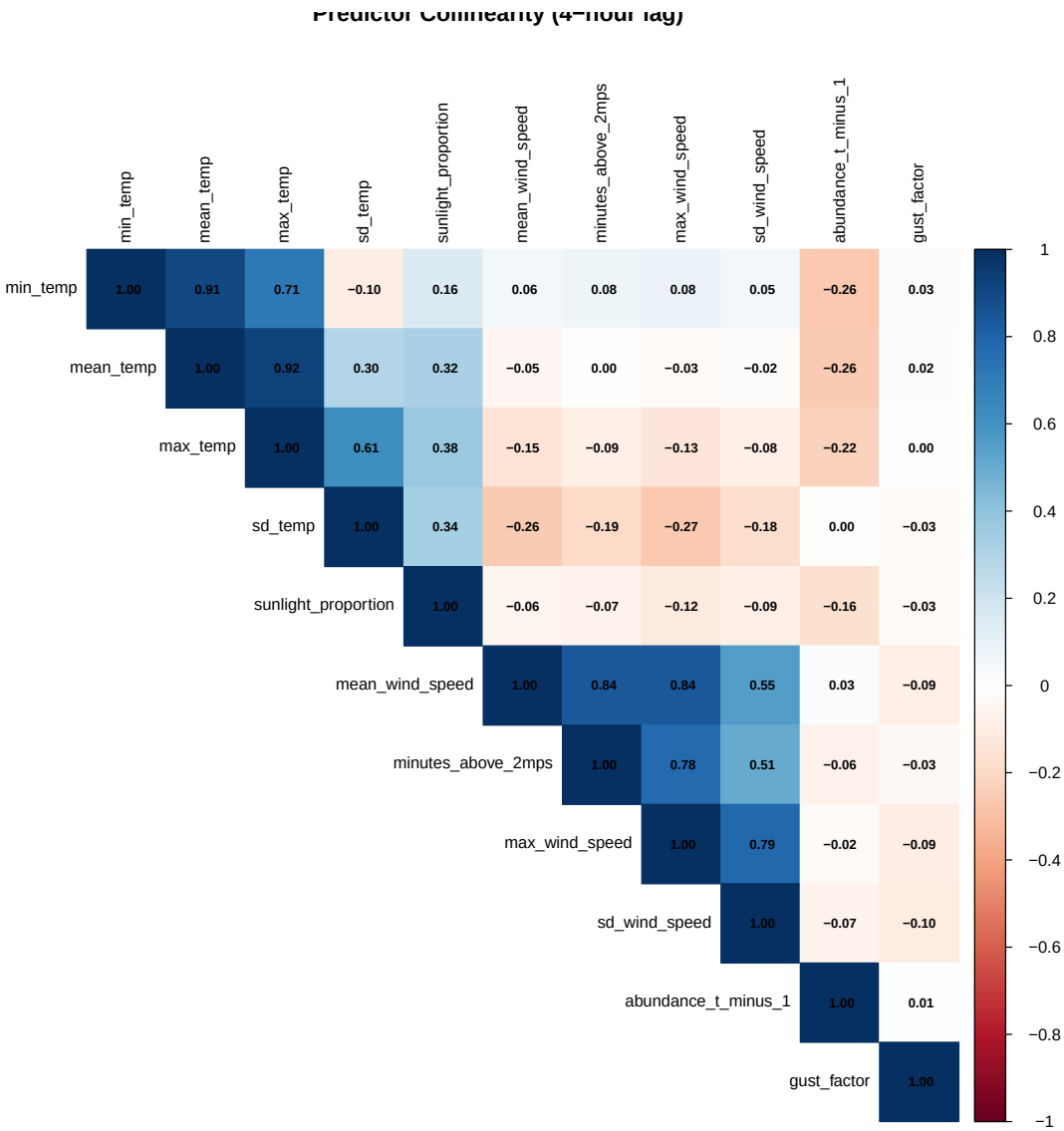
```
Rows with total_butterflies > 0: 5437
```

```r
cat("Unique view_ids:", length(unique(data_240m$view_id)), "\n")
```

```
Unique view_ids: 9
```

```r
# Select numeric predictors for correlation analysis
predictors_240m <- data_240m %>%
  select(abundance_t_minus_1, mean_wind_speed, max_wind_speed, sd_wind_speed,
         gust_factor, minutes_above_2mps, mean_temp, max_temp, min_temp, sd_temp,
         sunlight_proportion) %>%
  na.omit()

if (nrow(predictors_240m) > 1) {
  cor_matrix_240m <- cor(predictors_240m, use = "complete.obs")
  if (!any(is.na(cor_matrix_240m) | is.infinite(cor_matrix_240m))) {
    corrplot(cor_matrix_240m, method = "color", type = "upper",
             order = "hclust", tl.cex = 0.8, tl.col = "black",
             addCoef.col = "black", number.cex = 0.7,
             title = "Predictor Collinearity (4-hour lag)")
```

```r
  } else {
    corrplot(cor_matrix_240m, method = "color", type = "upper",
             order = "original", tl.cex = 0.8, tl.col = "black",
             addCoef.col = "black", number.cex = 0.7,
             title = "Predictor Collinearity (4-hour lag)")
  }
} else {
  cat("Not enough data for correlation analysis\n")
}
```



Model Fitting (4 hours)

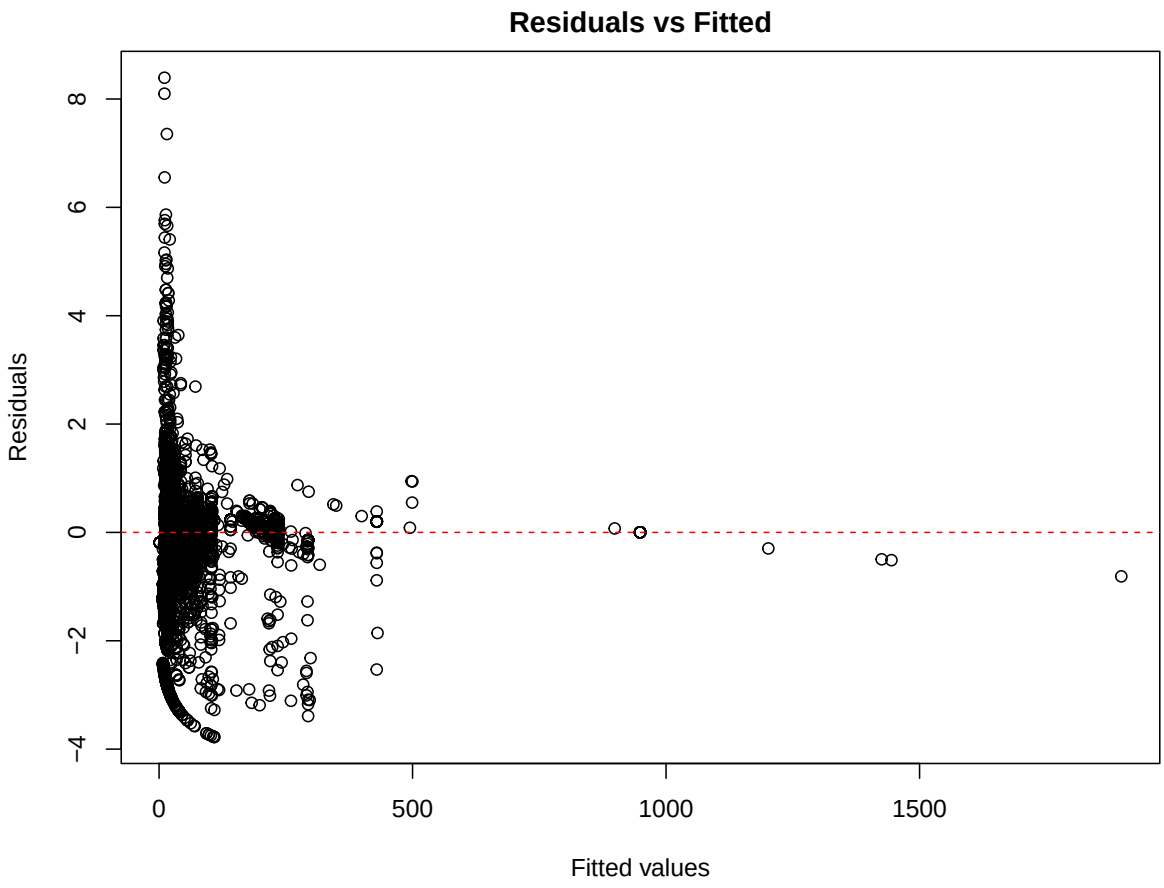
```
# Define comprehensive model formulas
model_formulas_240m <- list(
  # 1. Baseline Model
  "1. Baseline" = total_butterflies ~ s(abundance_t_minus_1) + s(view_id, bs = "re"),
```

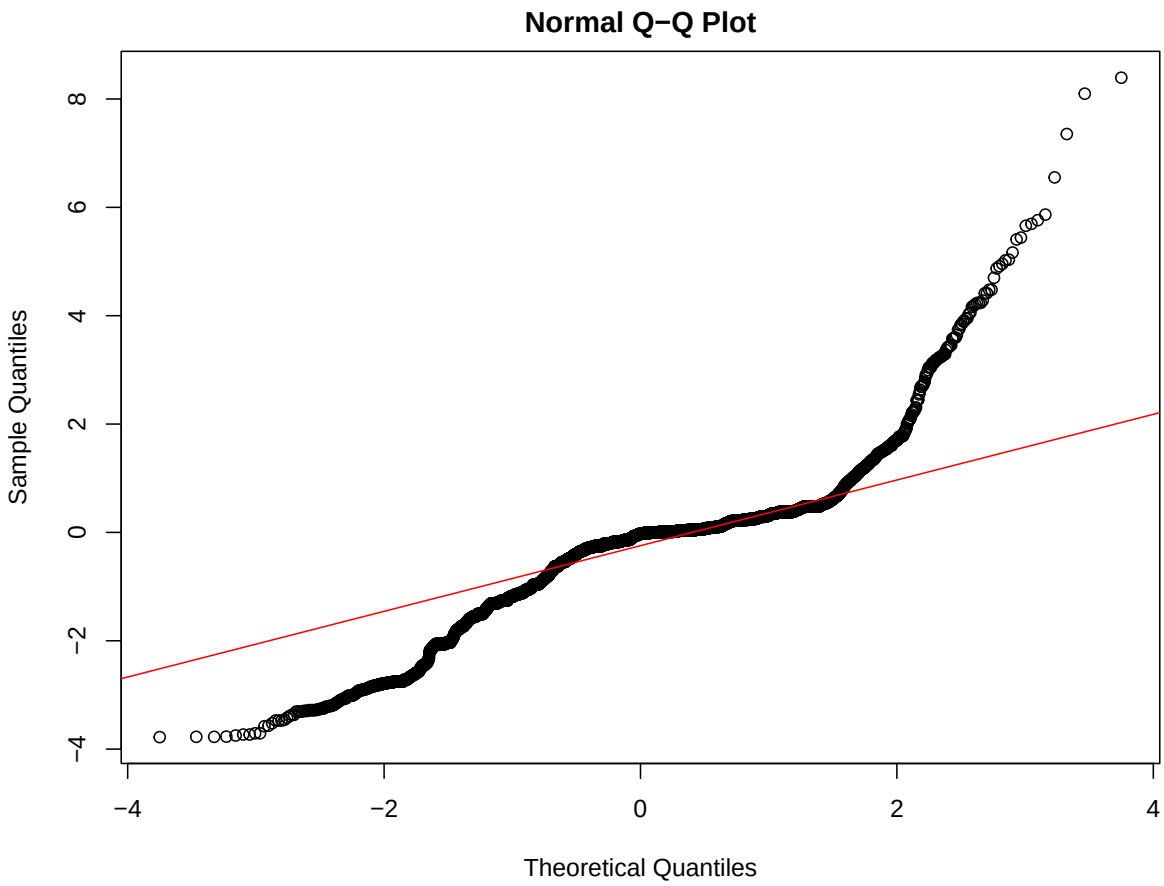
```
  # 2. Single Environmental Factor Models
  "2a. Wind only" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(view_
  "2b. Temperature only" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(vie
  "2c. Sunlight only" = total_butterflies ~ s(abundance_t_minus_1) + s(sunlight_proportion) +

  # 3. Additive Models (No Interactions)
  "3a. Wind + Temp" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(mea
  "3b. Wind + Sunlight" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s
  "3c. Temp + Sunlight" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(sunli
  "3d. All additive" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(me

  # 4. Single Interaction Models
  "4a. Wind x Temp interaction" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_sp
  "4b. Wind x Sunlight interaction" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wi
  "4c. Temp x Sunlight interaction" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wi

  # 5. Multiple Interaction Models
  "5a. Wind x Temp + Wind x Sun" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_s
  "5b. Wind x Temp + Temp x Sun" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_s
  "5c. Wind x Sun + Temp x Sun" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_sp

  # 6. Global Model with All Interactions
  "6. Global model" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(mea
)

# Storage for results
models_240m <- list()
aic_values_240m <- numeric()

# Fit each model
for (i in seq_along(model_formulas_240m)) {
  model_name <- names(model_formulas_240m)[i]
  formula <- model_formulas_240m[[i]]

  cat("\n### Model", i, ":", model_name, "\n\n")

  tryCatch({
    # Check data before fitting
    model_vars <- all.vars(formula)
    model_data <- data_240m %>%
      select(all_of(model_vars), view_id) %>%
      na.omit()
```

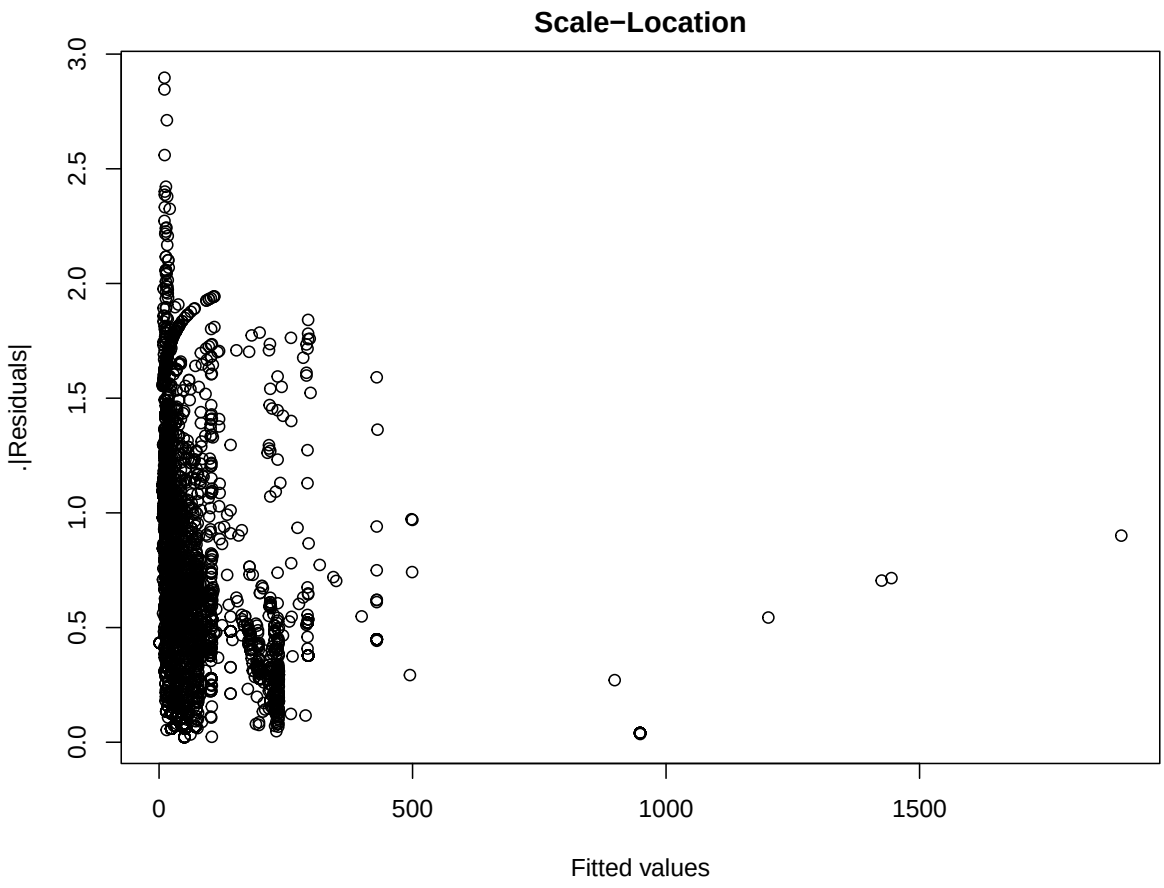
```r
cat("Model data dimensions after removing NAs:", nrow(model_data), "x", ncol(model_data)

# Fit GAM
gam_model <- gam(
  formula,
  family = nb(),
  data = model_data
)

cat("GAM fitted successfully\n")

# Store model
models_240m[[model_name]] <- gam_model
aic_values_240m[model_name] <- AIC(gam_model)

# Print summary
cat("Model Summary:\n")
print(summary(gam_model))

# Diagnostic plots - individual plots for better visibility in PDF
cat("\n\nDiagnostic plots:\n")

# Set up for individual plots
par(mfrow = c(1, 1), mar = c(4, 4, 2, 1))

# Plot 1: Residuals vs Fitted
plot(fitted(gam_model), residuals(gam_model),
     xlab = "Fitted values", ylab = "Residuals",
     main = "Residuals vs Fitted")
abline(h = 0, col = "red", lty = 2)

# Plot 2: Q-Q plot
qqnorm(residuals(gam_model))
qqline(residuals(gam_model), col = "red")

# Plot 3: Scale-Location (if appropriate)
plot(fitted(gam_model), sqrt(abs(residuals(gam_model))),
     xlab = "Fitted values", ylab = "√|Residuals|",
     main = "Scale-Location")

# Plot 4: Response vs Fitted
plot(fitted(gam_model), model_data$total_butterflies,
```

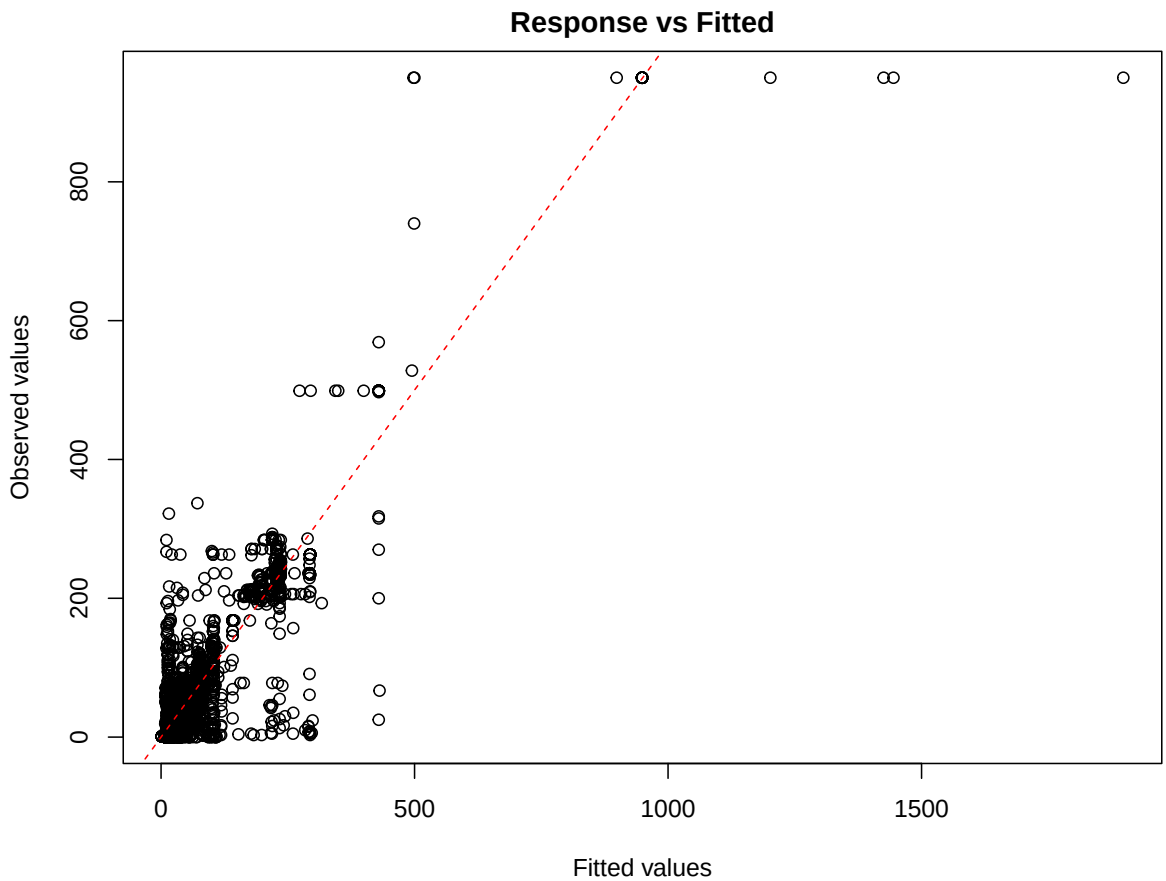
```r
        xlab = "Fitted values", ylab = "Observed values",
        main = "Response vs Fitted")
    abline(a = 0, b = 1, col = "red", lty = 2)

  }, error = function(e) {
    cat("Error fitting model:", e$message, "\n")
    aic_values_240m[model_name] <- NA
  })
}
```

```
### Model 1 : 1. Baseline

Model data dimensions after removing NAs: 5663 x 3
GAM fitted successfully
Model Summary:

Family: Negative Binomial(1.711)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.5412     0.3467   10.21   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                         edf Ref.df Chi.sq p-value
s(abundance_t_minus_1) 8.302  8.828 7340.7  <2e-16 ***
s(view_id)             7.833  8.000  305.7  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.841   Deviance explained = 60.7%
-REML =  25971  Scale est. = 1         n = 5663


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
8
6
4
2
0
−2
−4
0
500
1000
1500

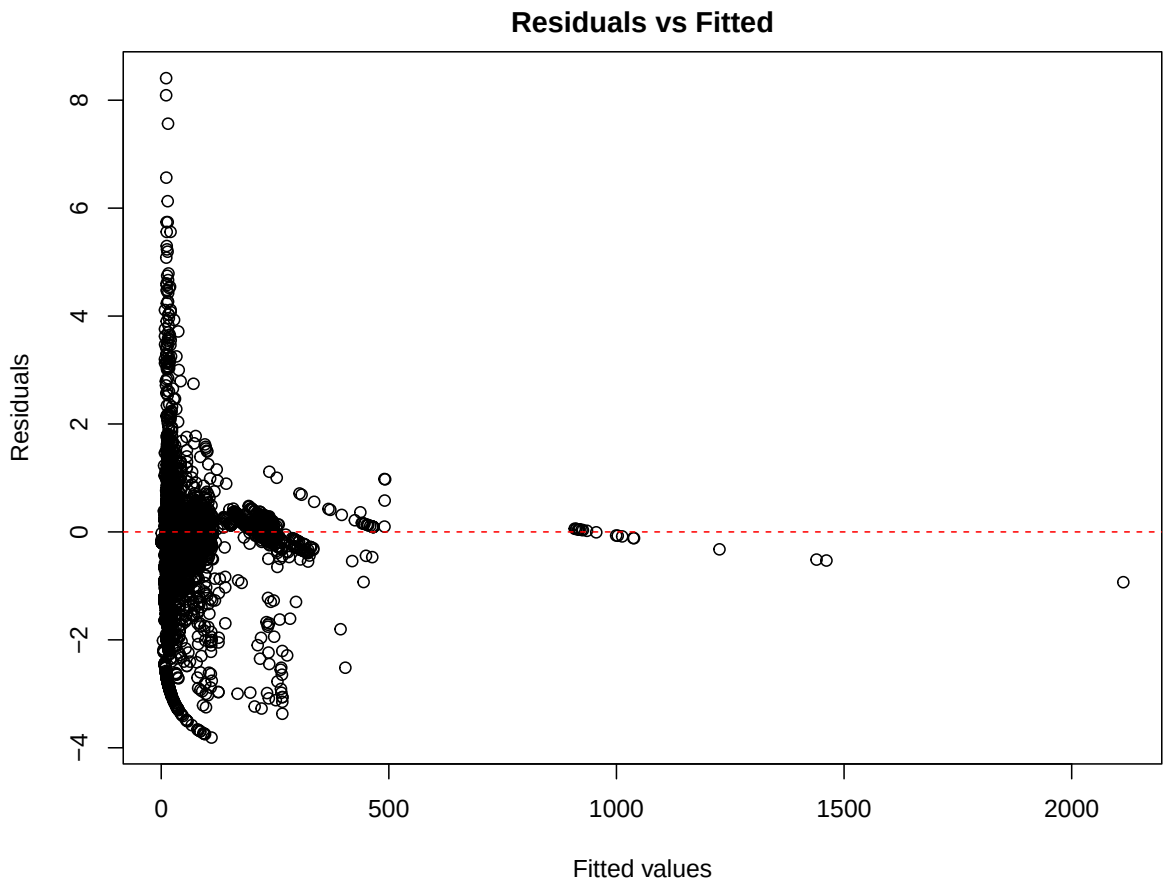

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6
8
−4
−2
0
2
4

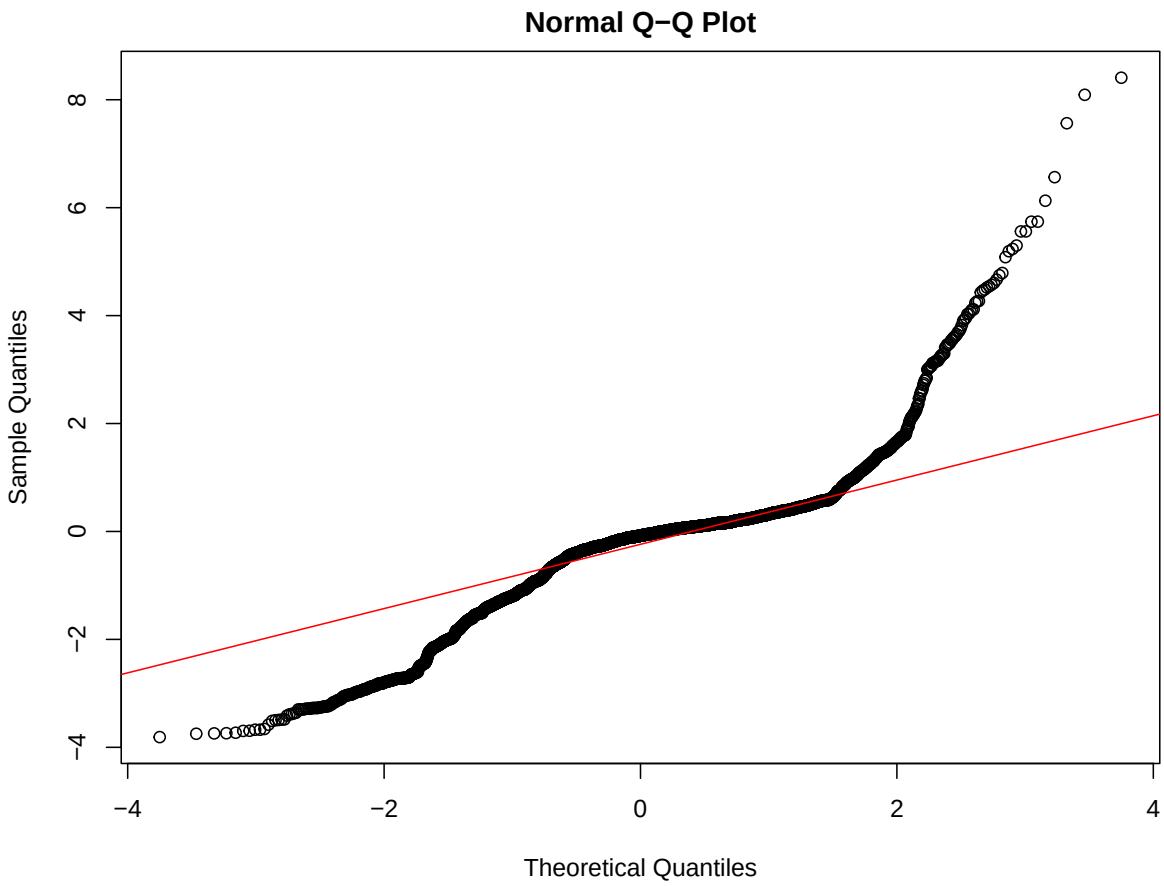

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
500
1000
1500
Fitted values

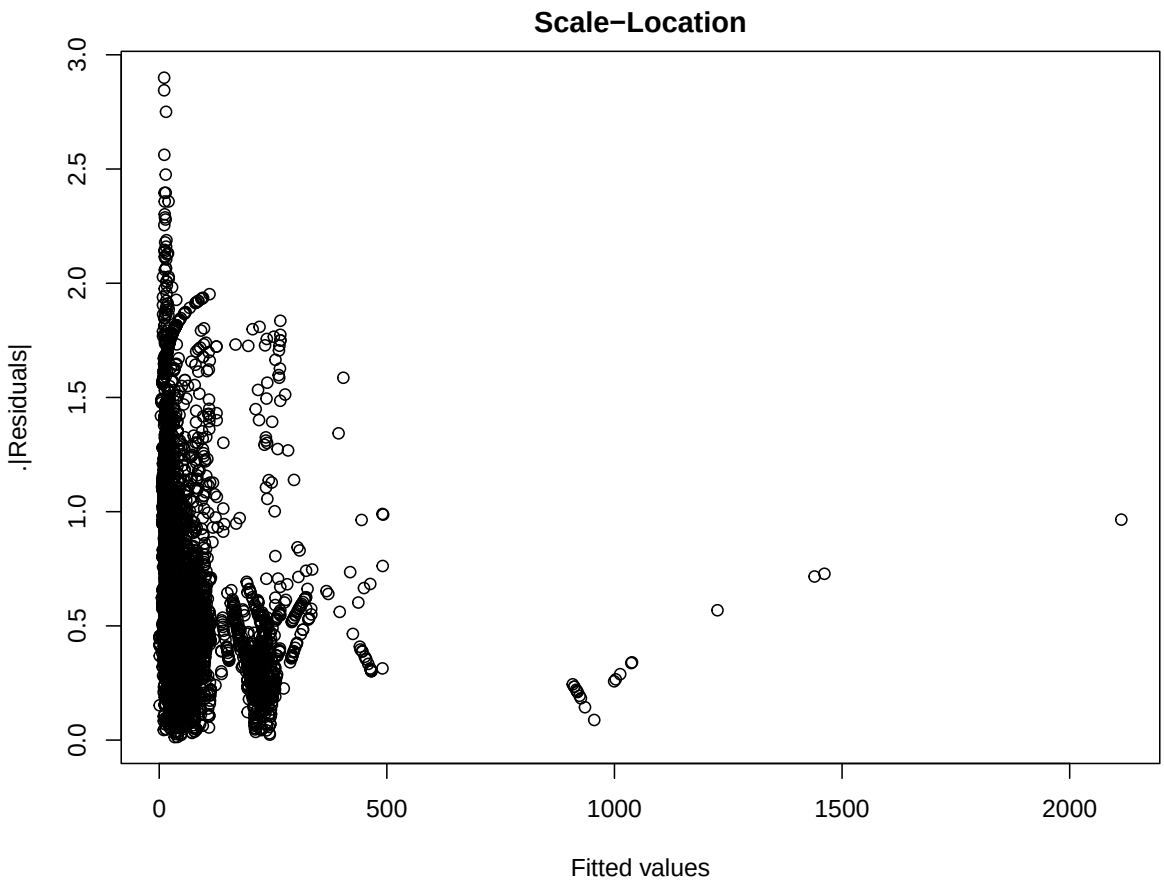


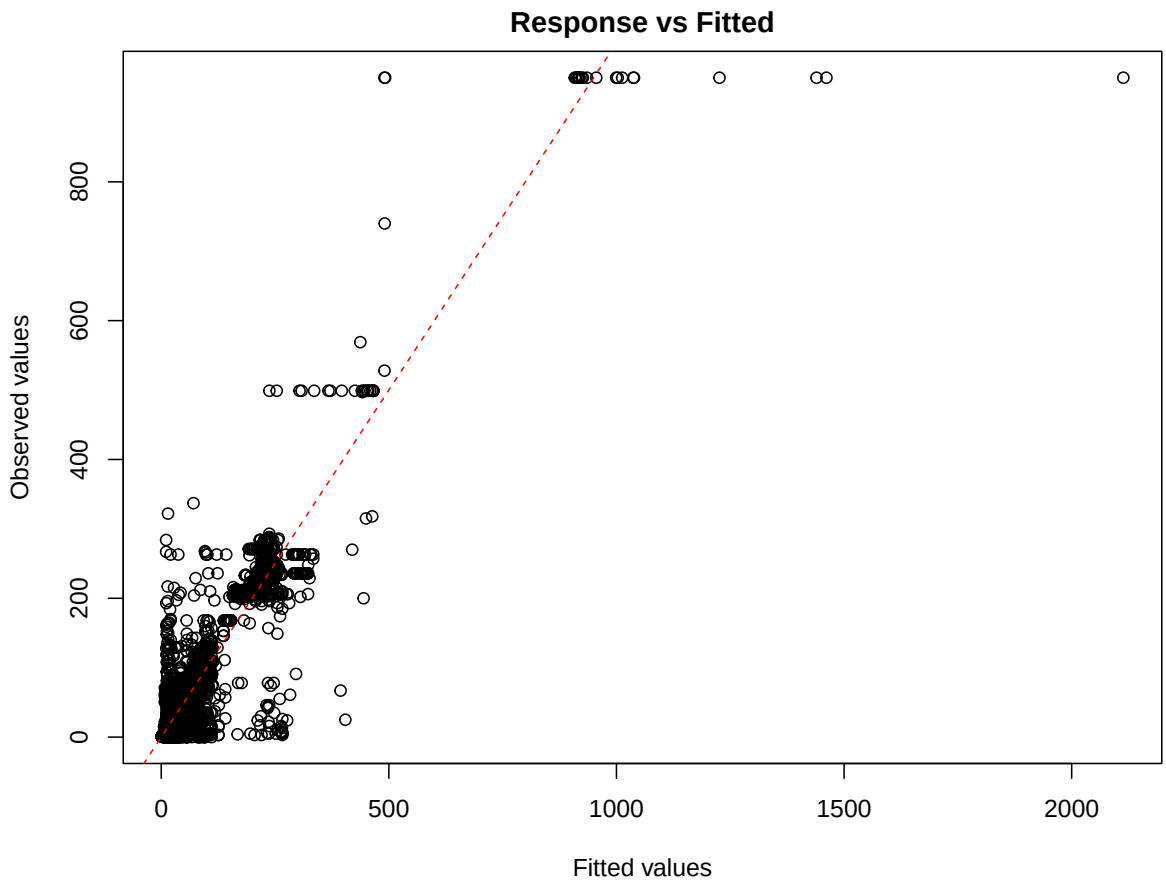

```
### Model 2 : 2a. Wind only

Model data dimensions after removing NAs: 5663 x 4
GAM fitted successfully
Model Summary:

Family: Negative Binomial(1.744)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.5671     0.3657   9.755   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df Chi.sq p-value
s(abundance_t_minus_1) 8.339  8.845 7236.5  <2e-16 ***
s(mean_wind_speed)     6.491  7.635  111.5  <2e-16 ***
s(view_id)             7.849  8.000  319.4  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.828   Deviance explained = 61.4%
-REML =  25933  Scale est. = 1         n = 5663


Diagnostic plots:
```



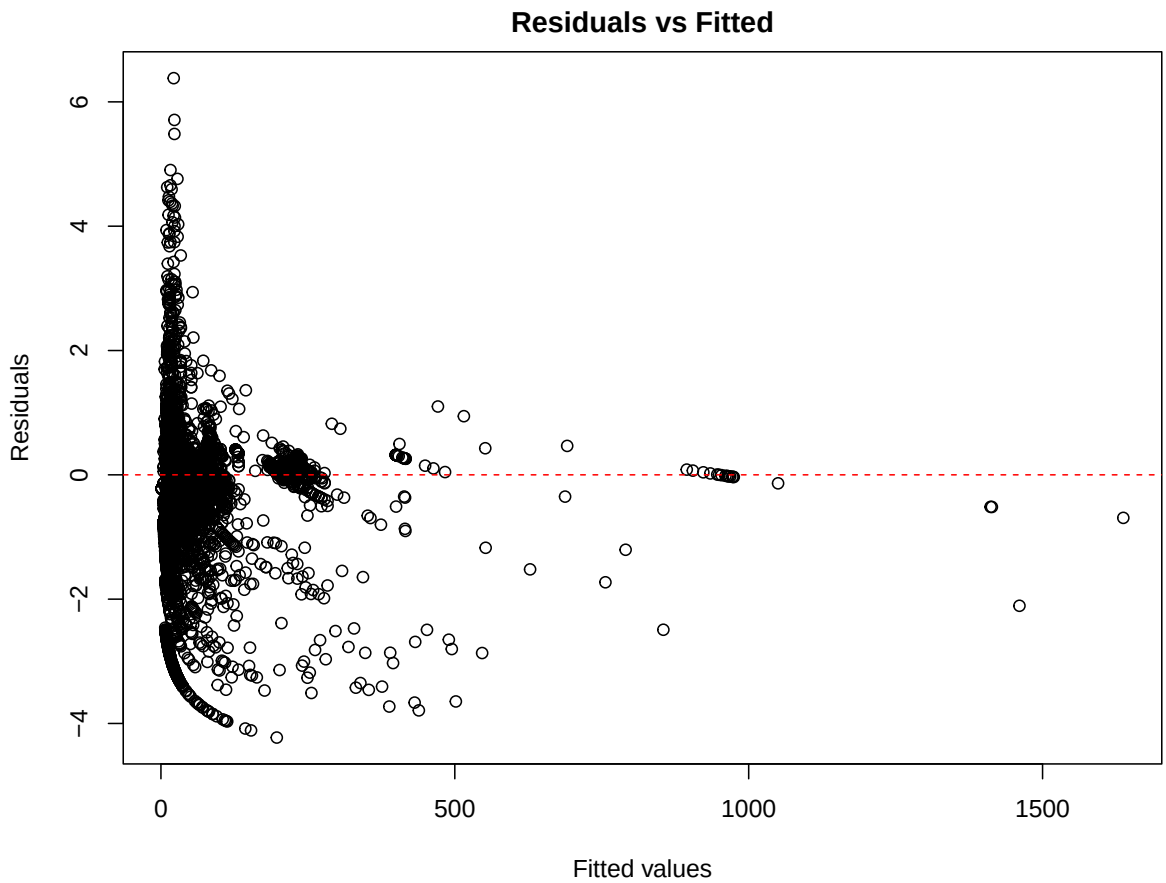

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6
8
−4
−2
0
2
4

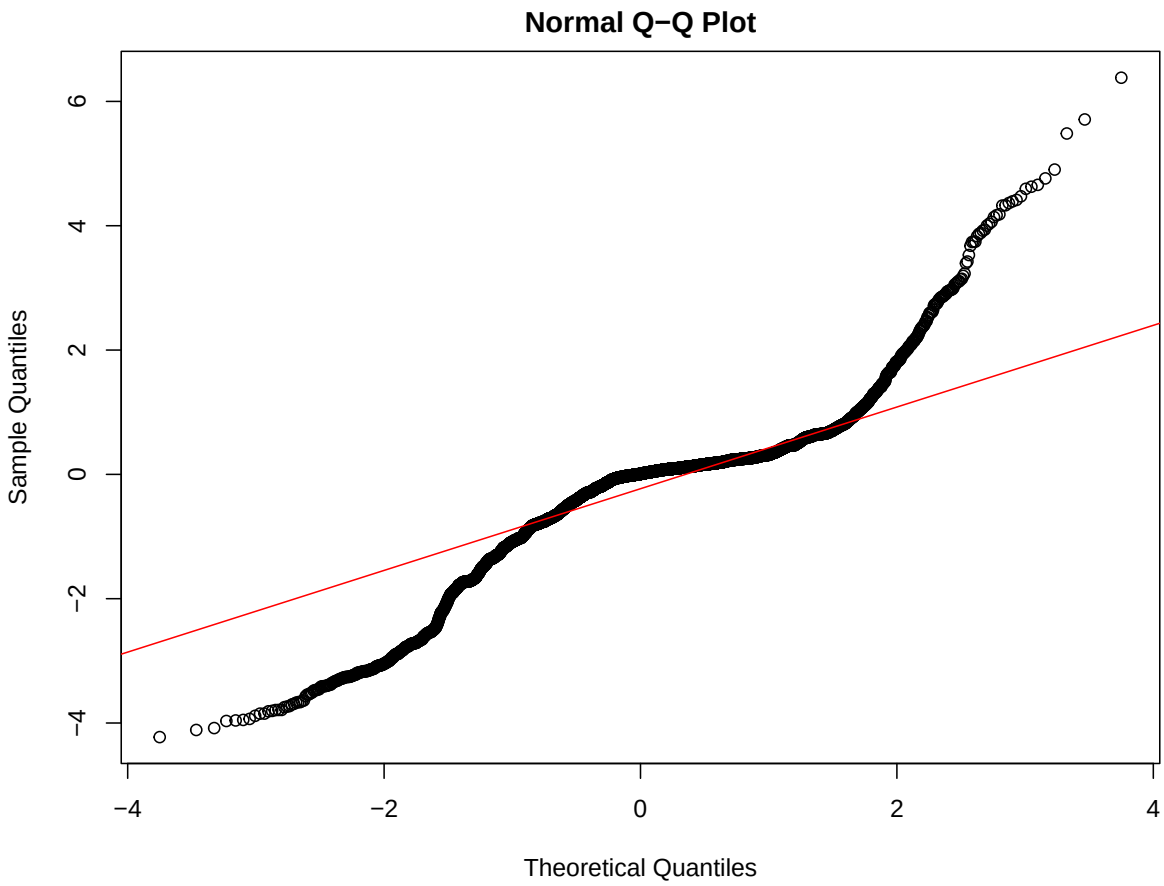

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
500
1000
1500
2000
Fitted values

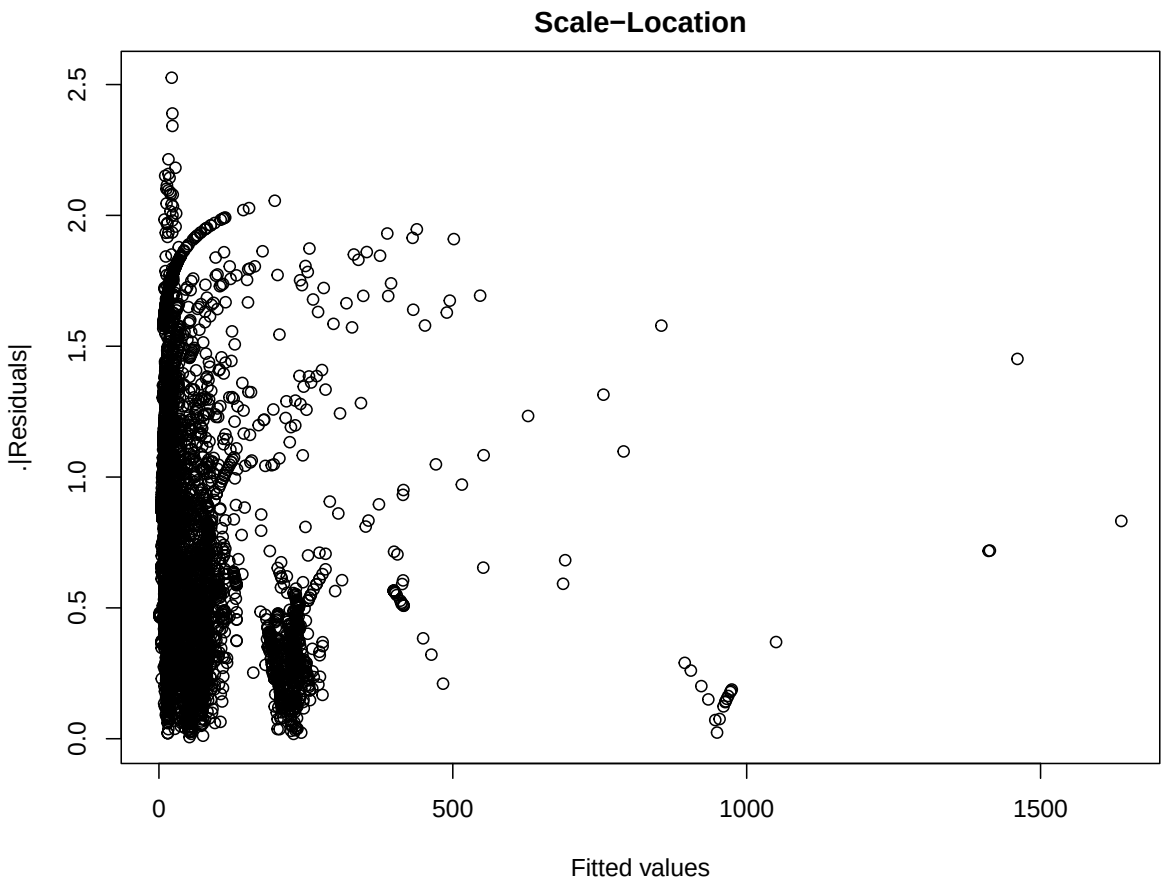


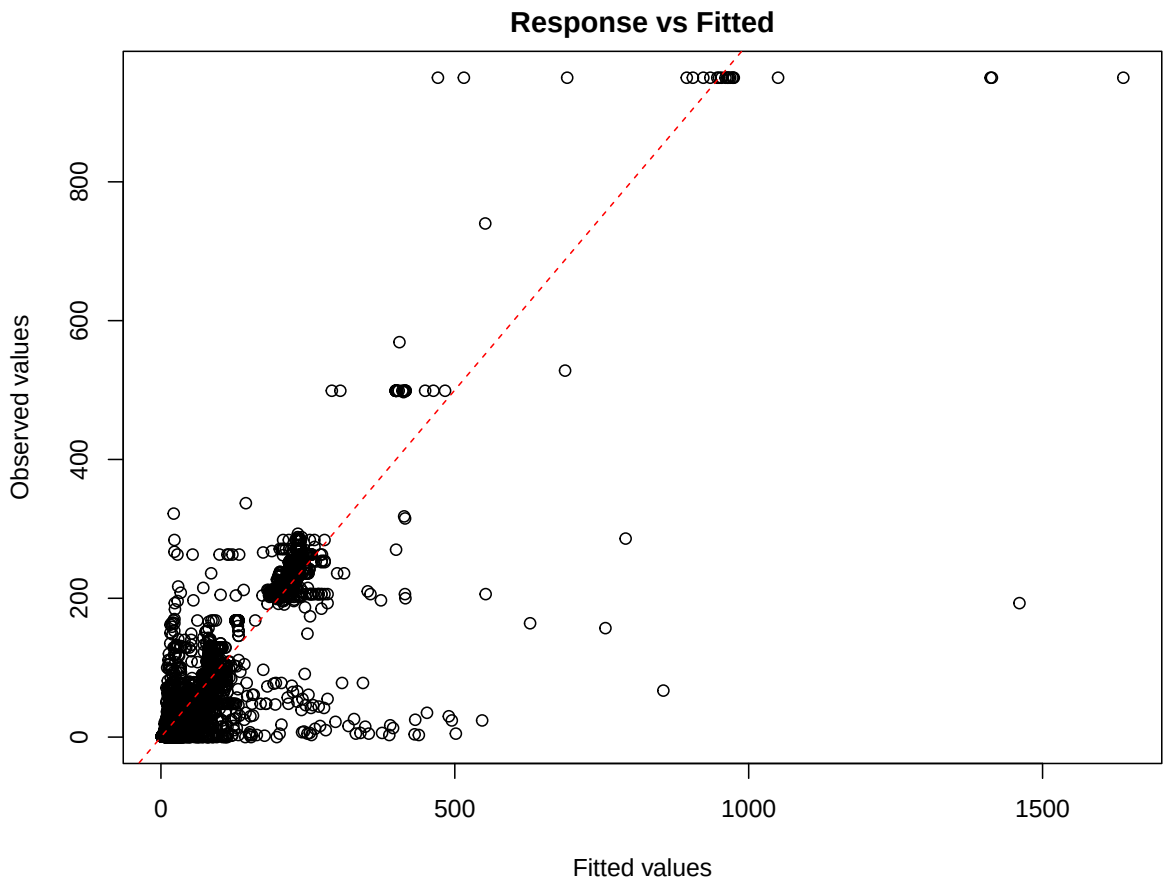

```
### Model 3 : 2b. Temperature only

Model data dimensions after removing NAs: 5663 x 4
GAM fitted successfully
Model Summary:

Family: Negative Binomial(1.926)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(view_id,
    bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.4828     0.2704   12.88   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df Chi.sq p-value
s(abundance_t_minus_1) 8.536  8.920 9584.2  <2e-16 ***
s(mean_temp)           7.489  8.406  741.7  <2e-16 ***
s(view_id)             7.751  8.000  409.9  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.738   Deviance explained = 64.9%
-REML =  25651  Scale est. = 1         n = 5663


Diagnostic plots:
```



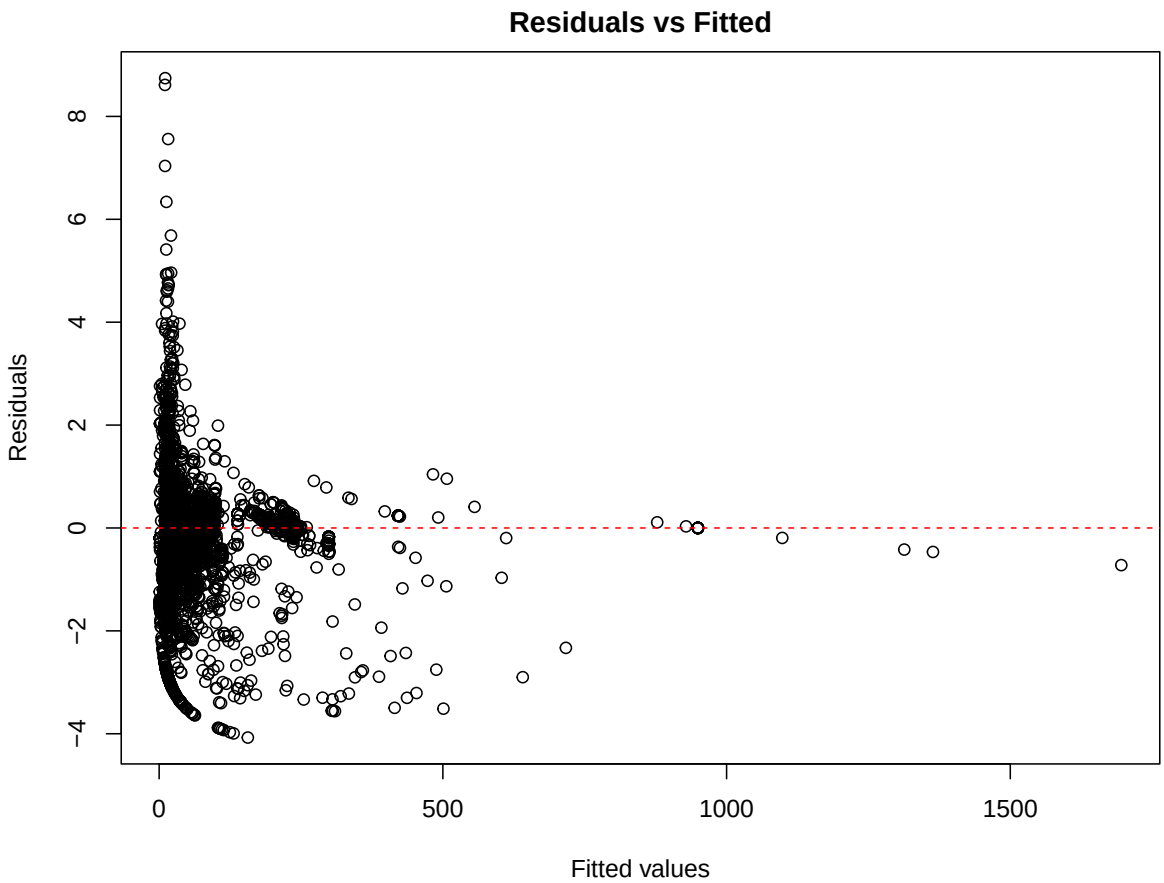

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6
−4
−2
0
2
4

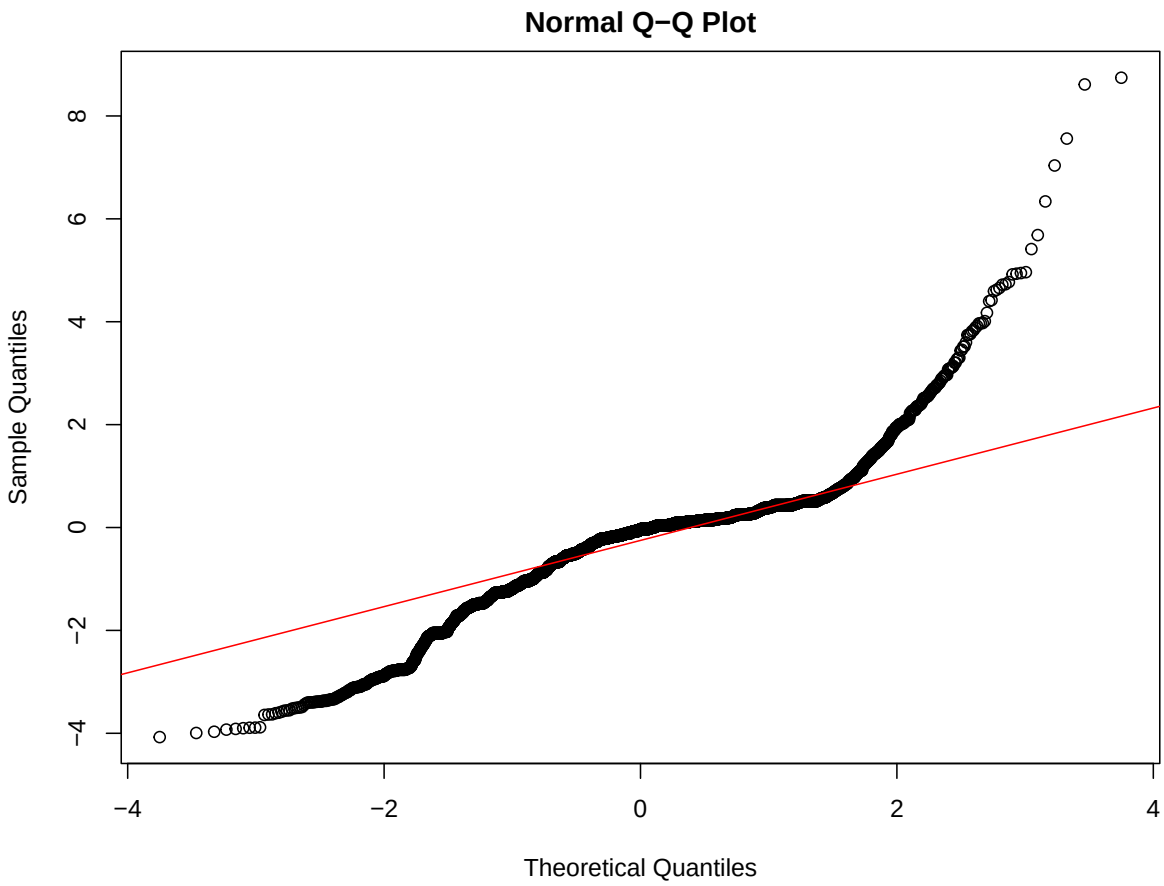

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
Fitted values

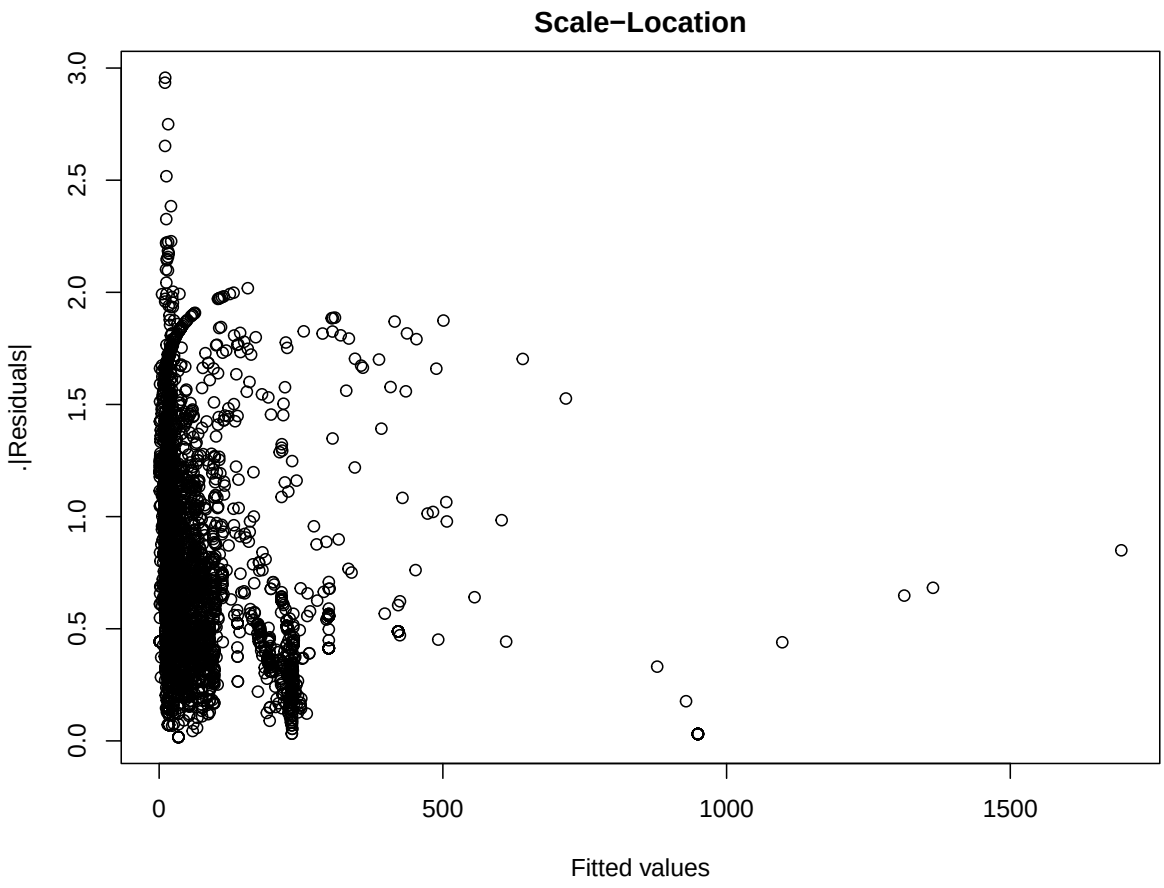


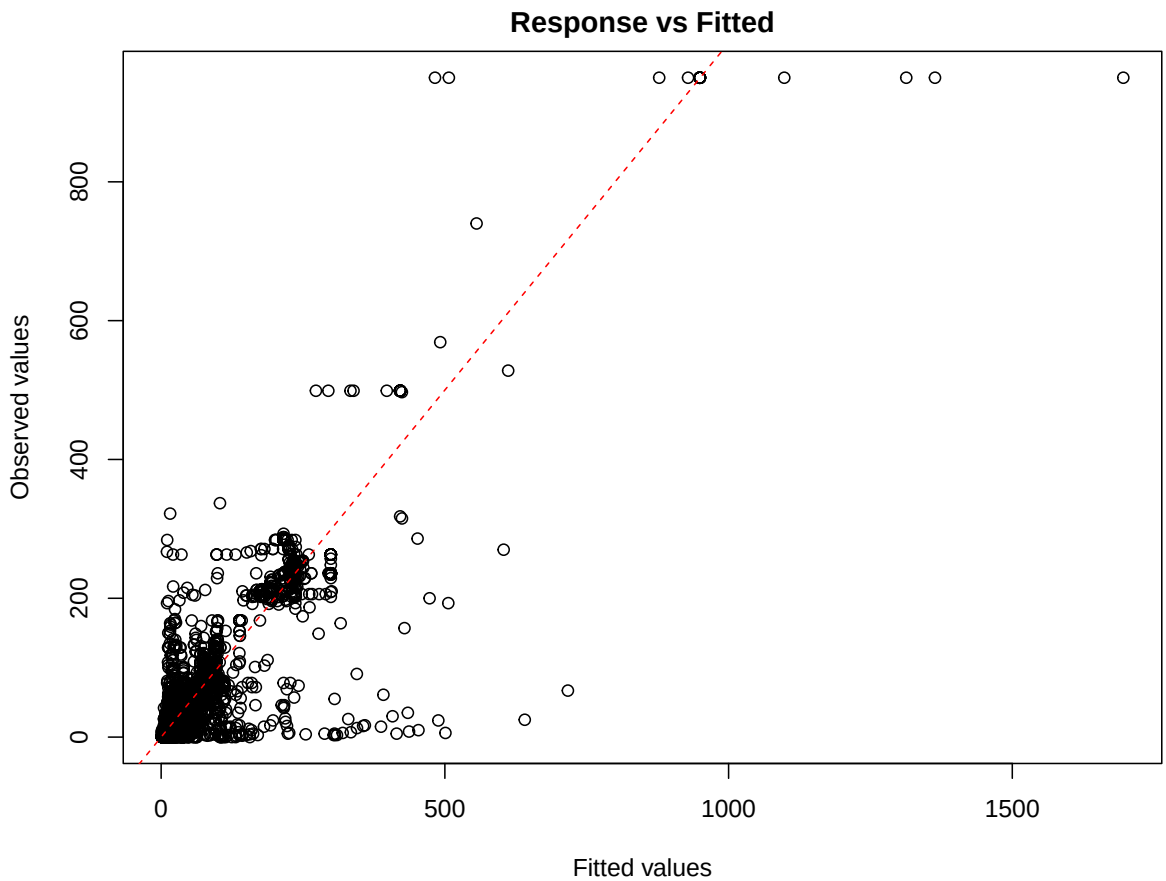

```
### Model 4 : 2c. Sunlight only

Model data dimensions after removing NAs: 5663 x 4
GAM fitted successfully
Model Summary:

Family: Negative Binomial(1.87)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(sunlight_proportion) +
    s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.5287     0.3282   10.75   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df Chi.sq p-value
s(abundance_t_minus_1) 8.371  8.858 8002.5  <2e-16 ***
s(sunlight_proportion) 7.752  8.586  555.7  <2e-16 ***
s(view_id)             7.825  8.000  277.8  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.796   Deviance explained = 63.9%
-REML =  25735  Scale est. = 1         n = 5663


Diagnostic plots:
```



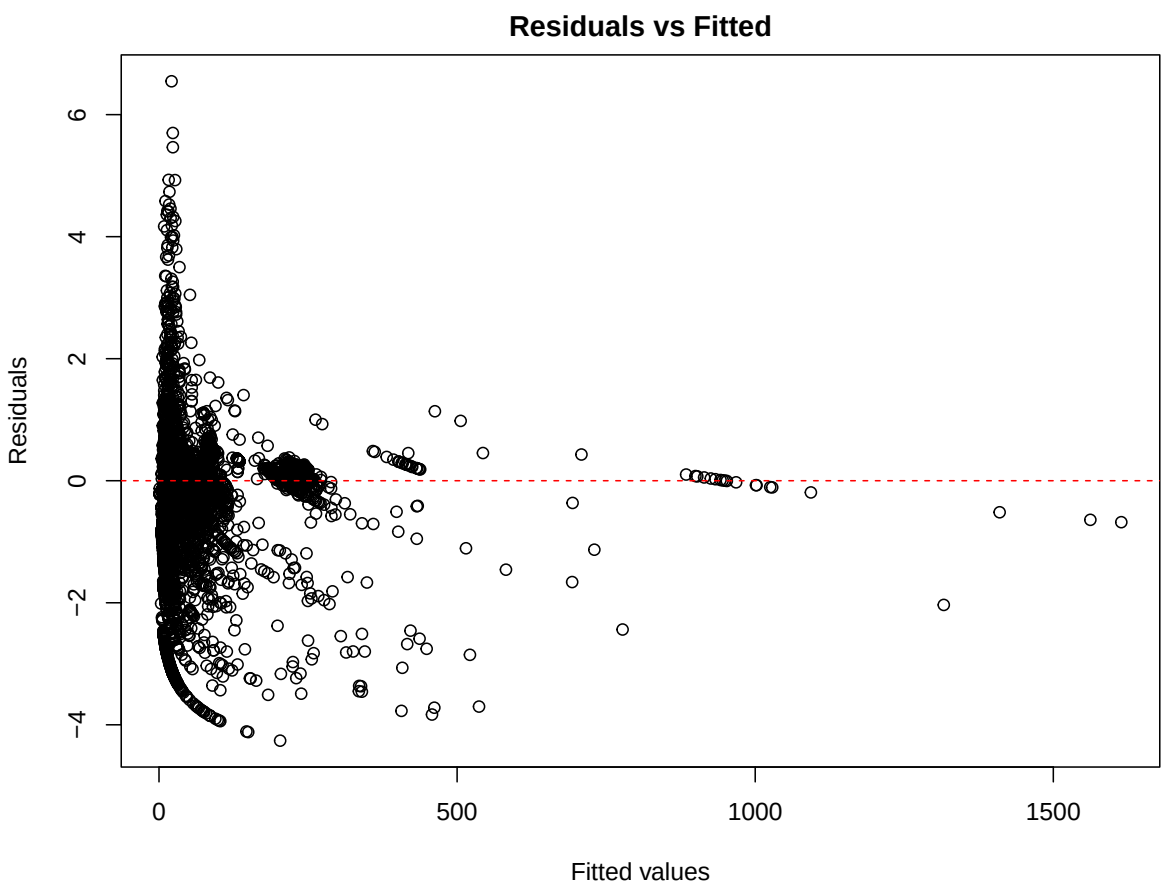

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6
8
−4
−2
0
2
4

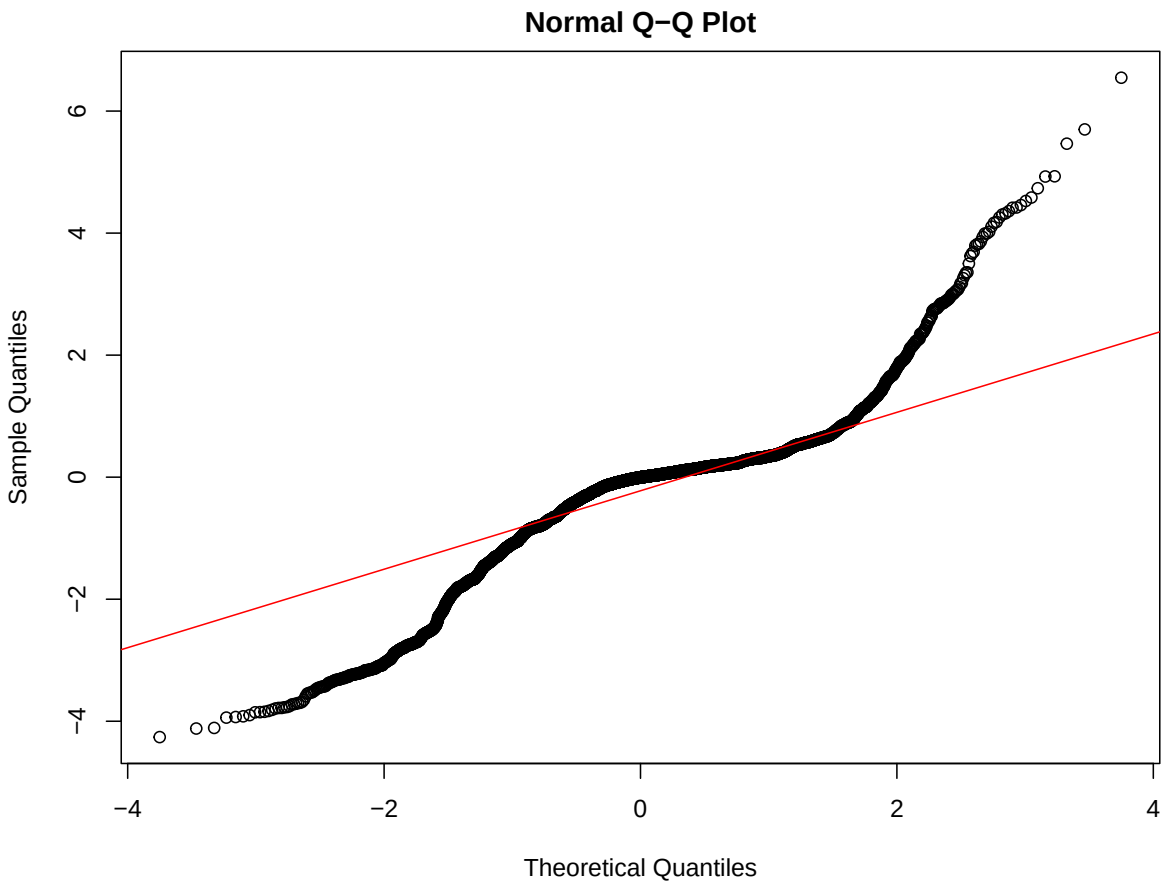

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
500
1000
1500
Fitted values

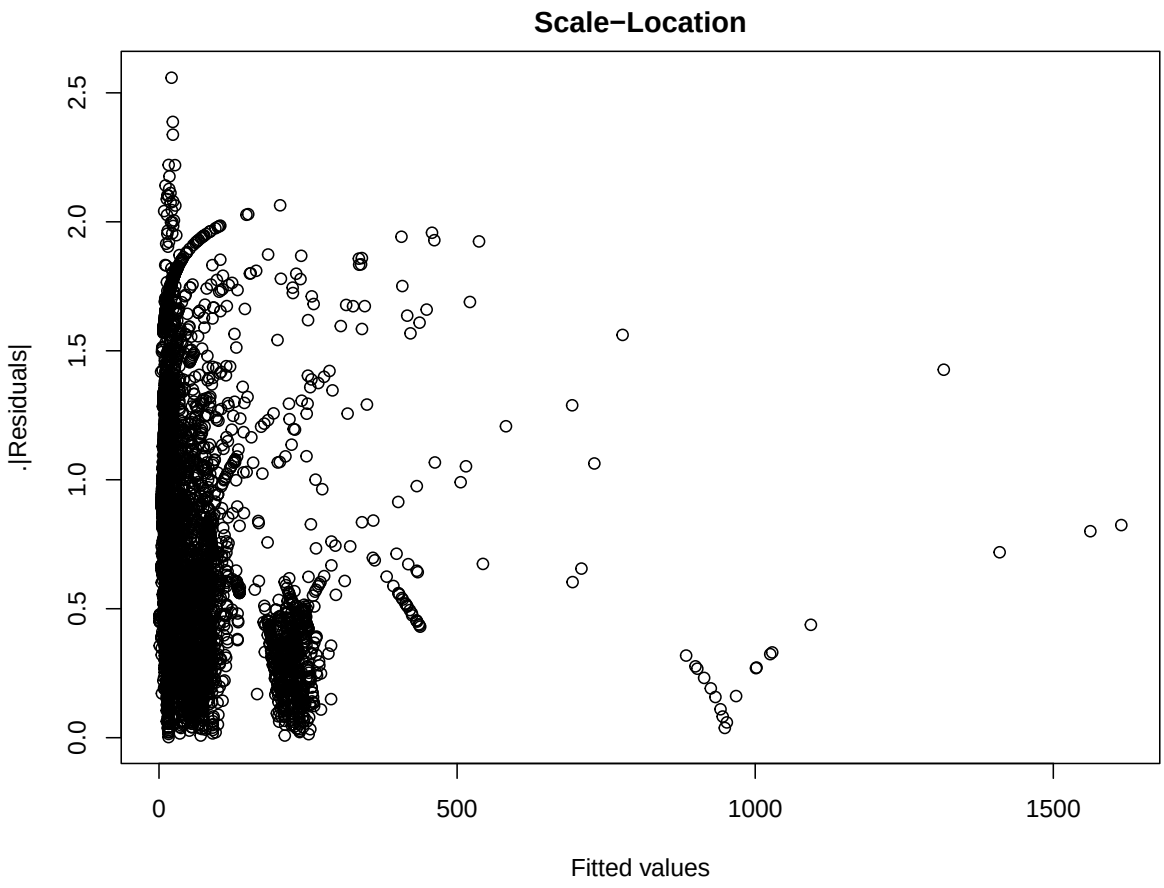


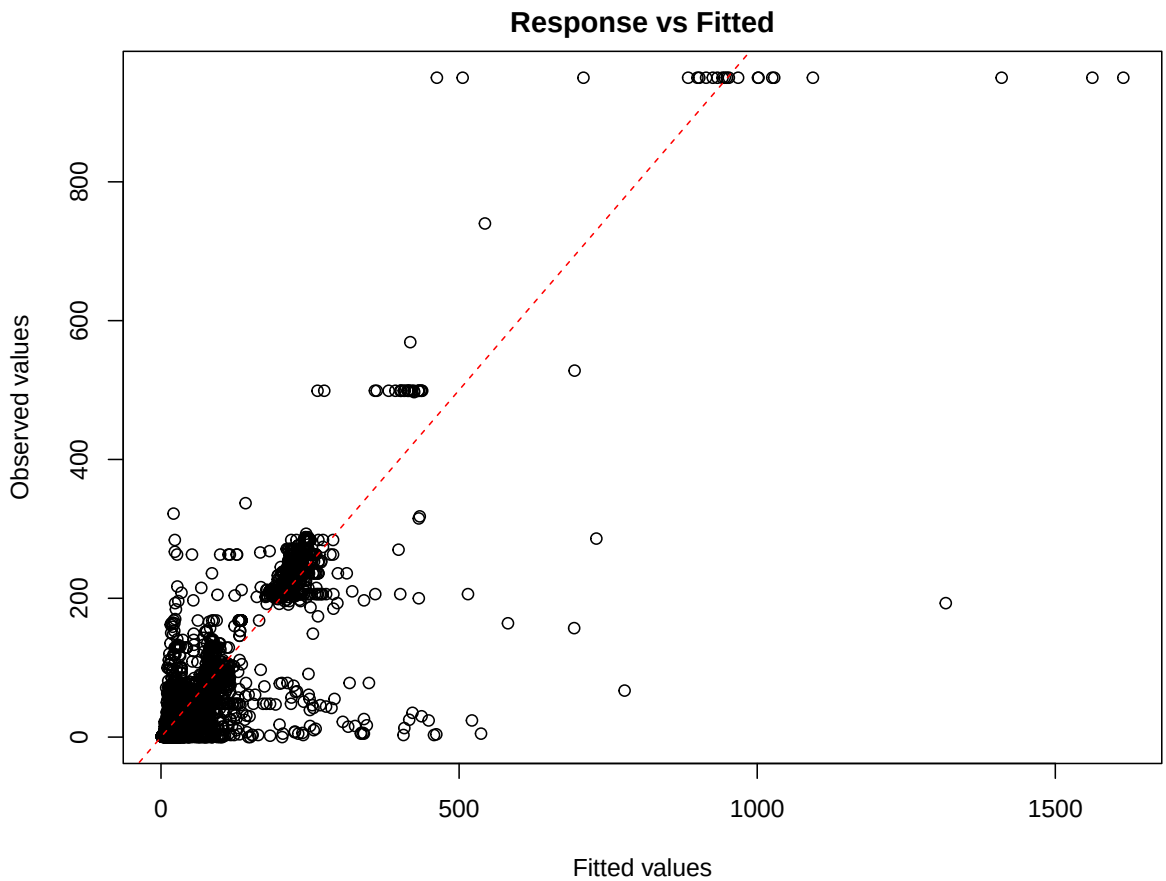

```
### Model 5 : 3a. Wind + Temp

Model data dimensions after removing NAs: 5663 x 5
GAM fitted successfully
Model Summary:

Family: Negative Binomial(1.949)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.5028     0.2837   12.35   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df  Chi.sq p-value
s(abundance_t_minus_1) 8.554  8.925 9363.46  <2e-16 ***
s(mean_wind_speed)     5.926  7.094   65.76  <2e-16 ***
s(mean_temp)           7.637  8.503  691.10  <2e-16 ***
s(view_id)             7.771  8.000  374.94  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.751   Deviance explained = 65.3%
-REML =  25633  Scale est. = 1         n = 5663


Diagnostic plots:
```



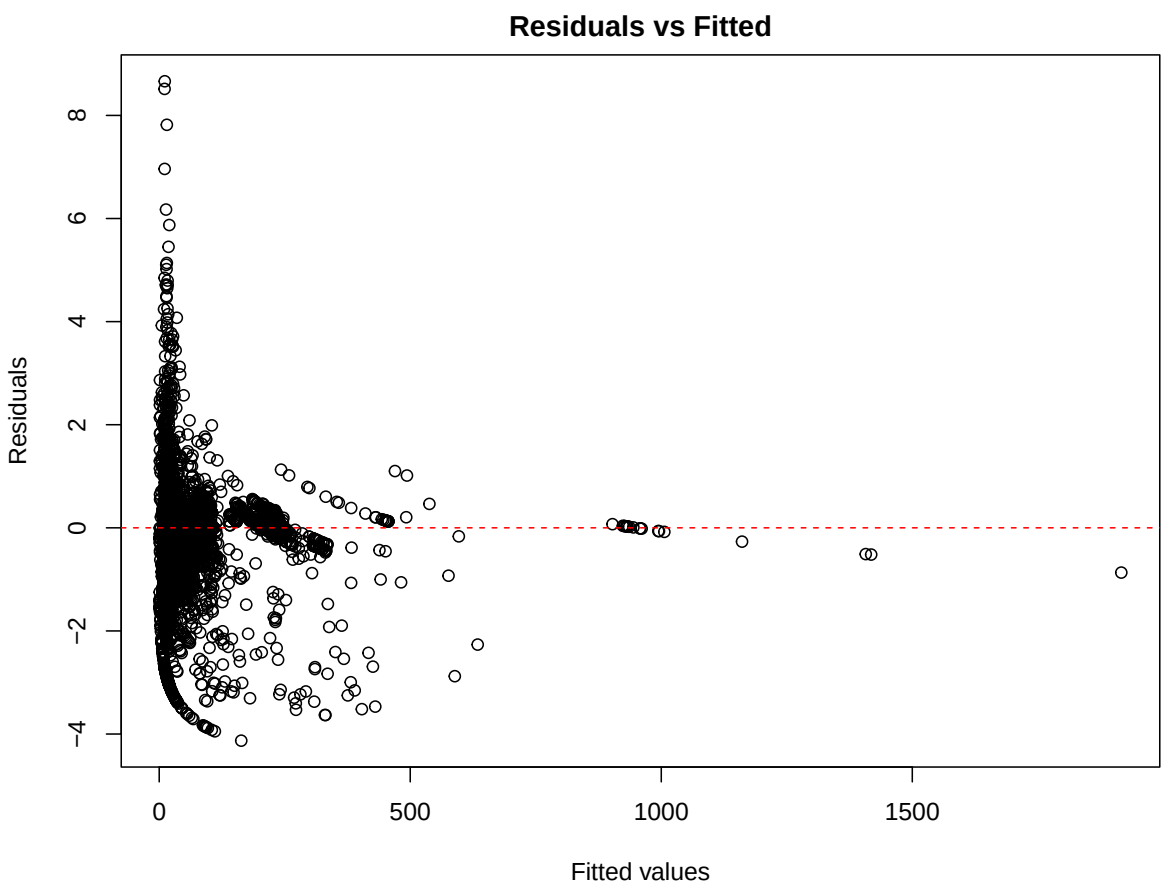

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6
−4
−2
0
2
4

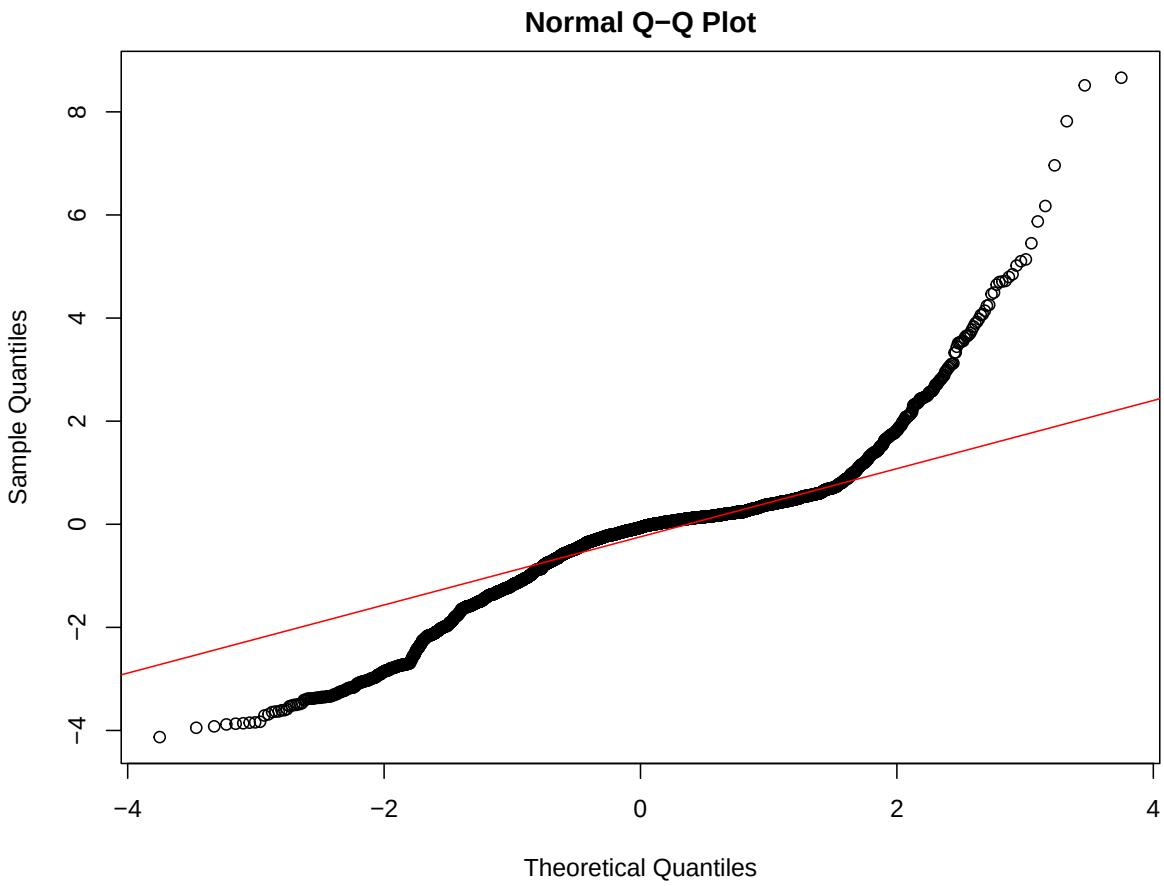

Scale−Location
.|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
Fitted values

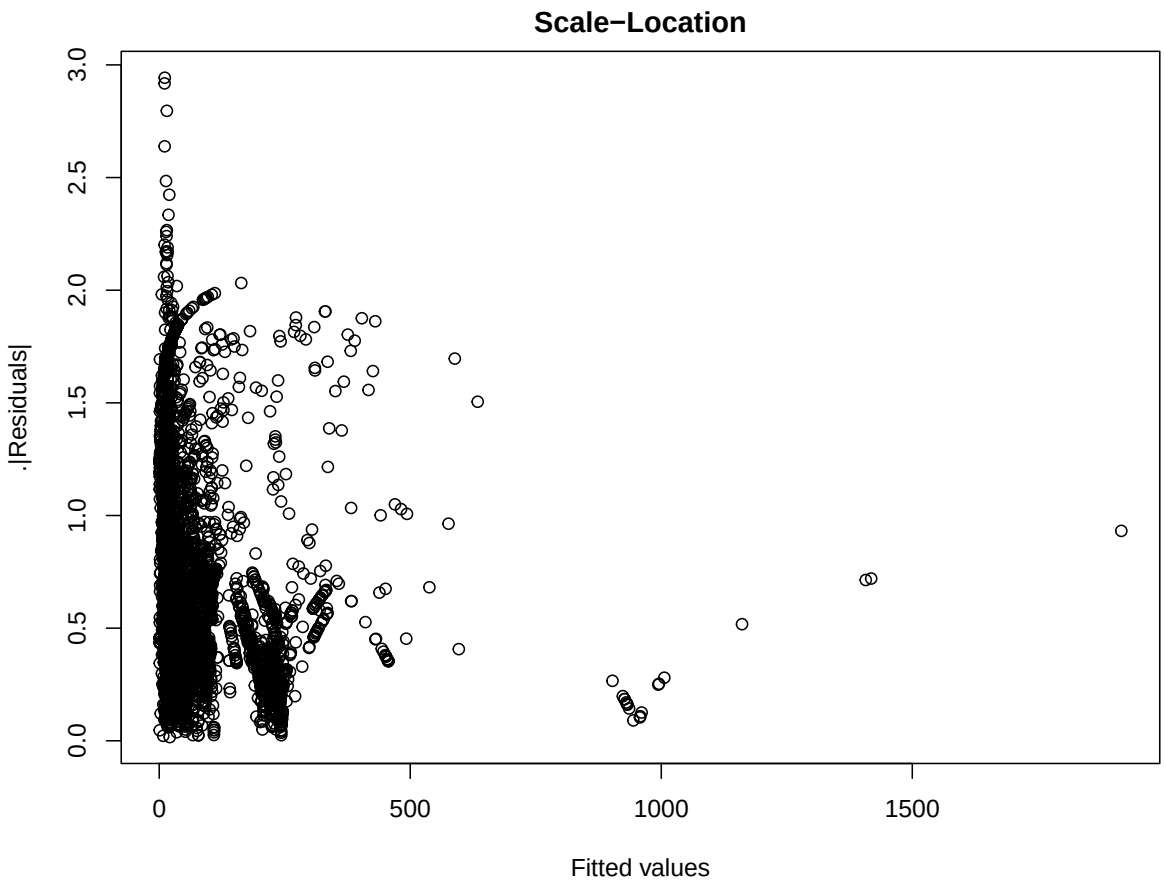


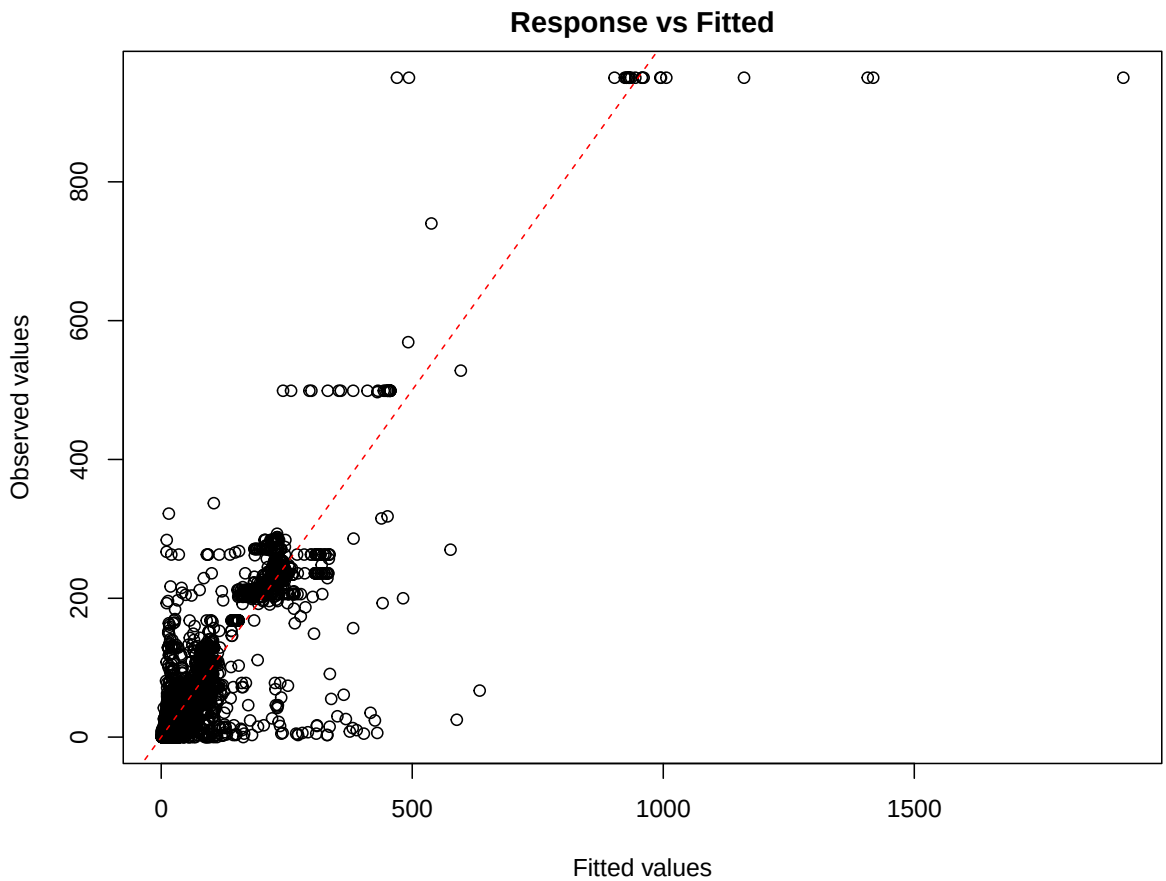

```
### Model 6 : 3b. Wind + Sunlight

Model data dimensions after removing NAs: 5663 x 5
GAM fitted successfully
Model Summary:

Family: Negative Binomial(1.911)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(sunlight_proportion) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.5636     0.3518   10.13   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df Chi.sq p-value
s(abundance_t_minus_1) 8.394  8.868 7756.6  <2e-16 ***
s(mean_wind_speed)     6.366  7.520  121.9  <2e-16 ***
s(sunlight_proportion) 7.708  8.559  568.9  <2e-16 ***
s(view_id)             7.846  8.000  316.5  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.794   Deviance explained = 64.6%
-REML =  25691  Scale est. = 1         n = 5663


Diagnostic plots:
```



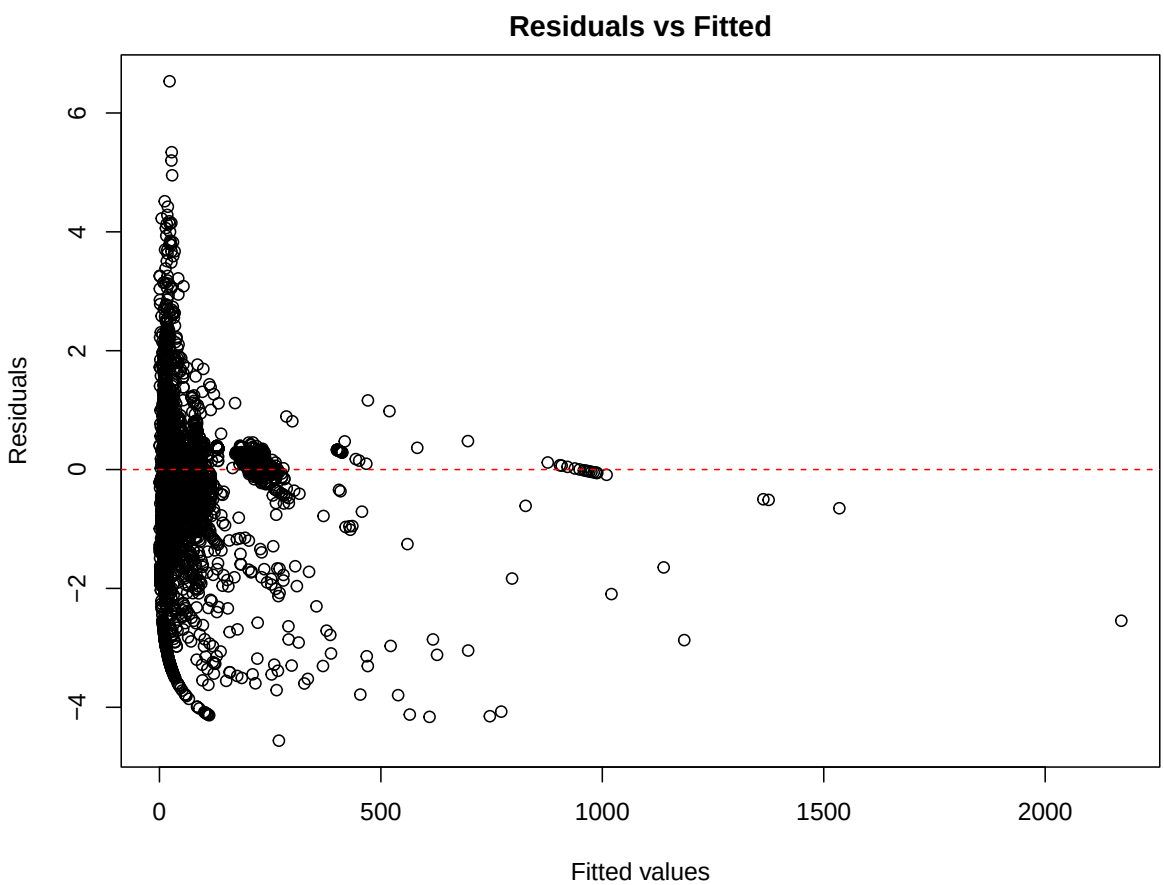

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6
8
−4
−2
0
2
4

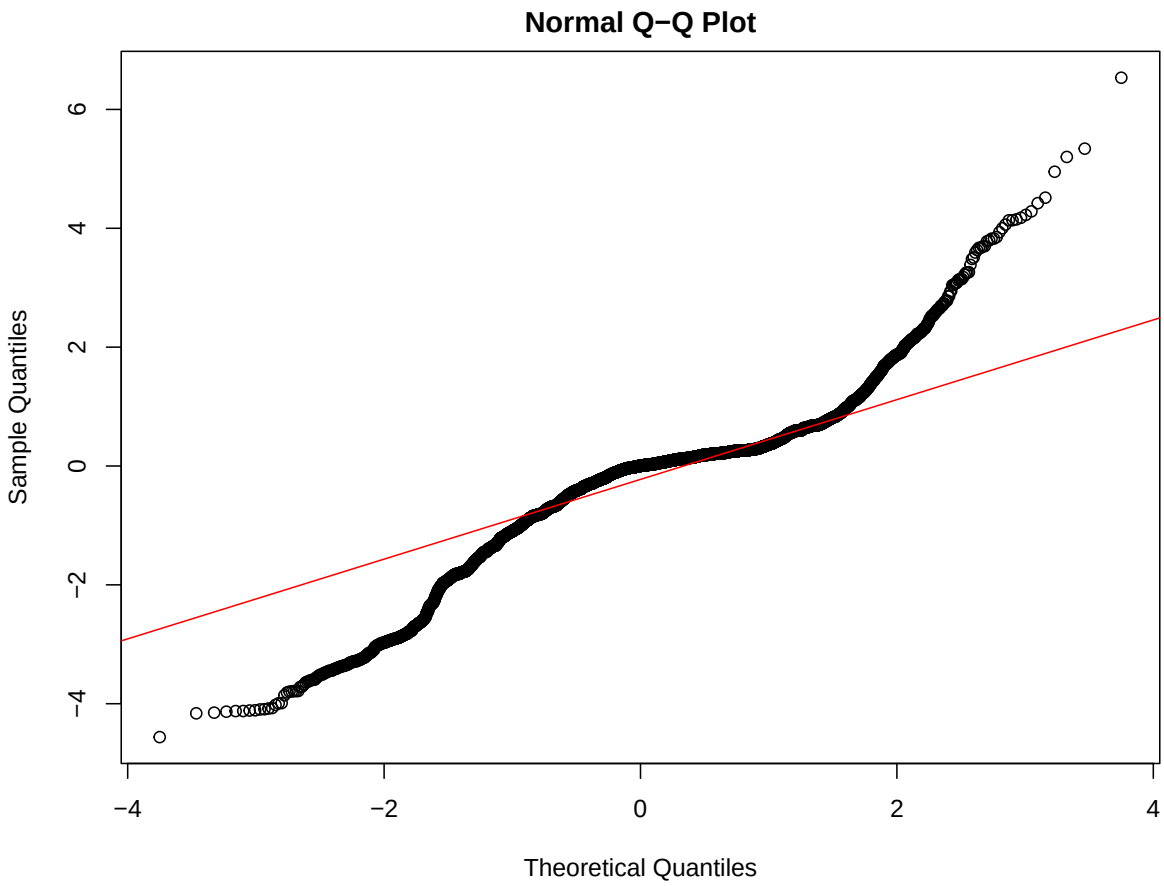

Scale−Location
.|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
500
1000
1500
Fitted values

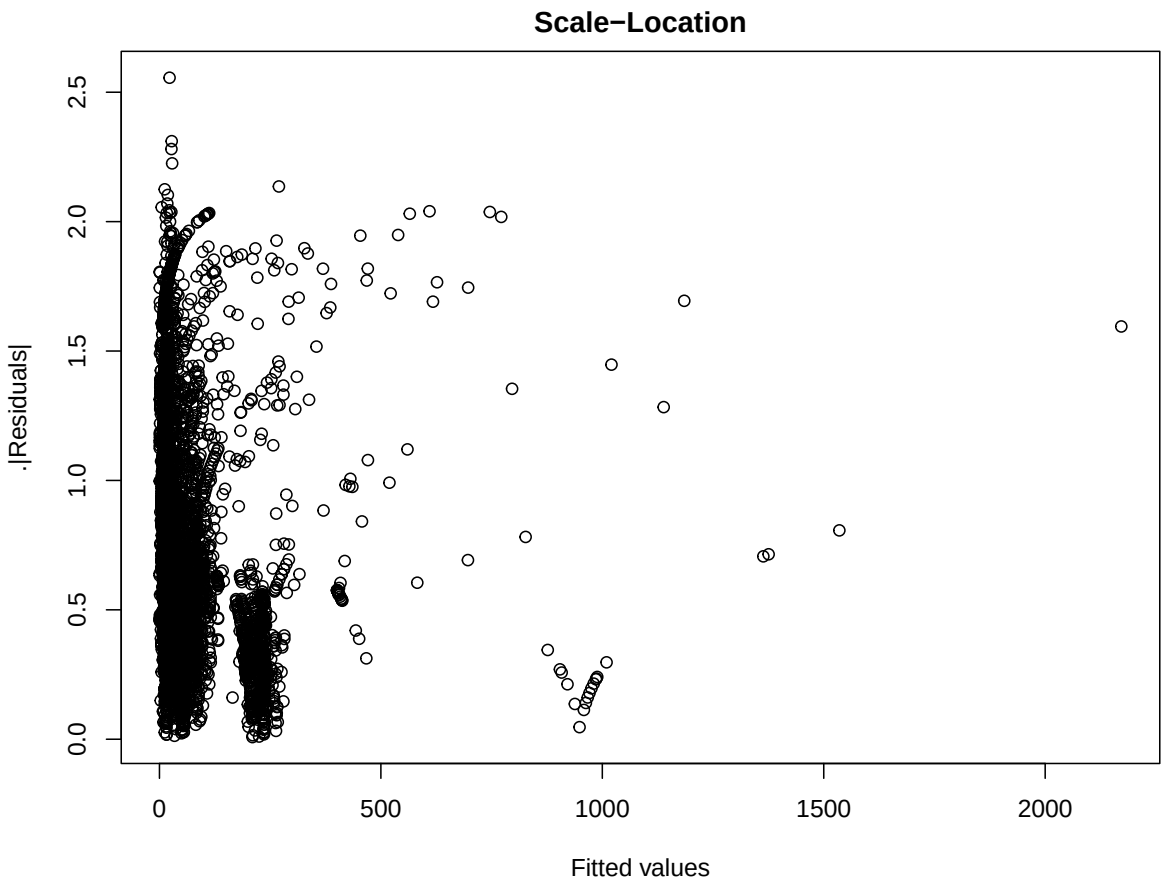


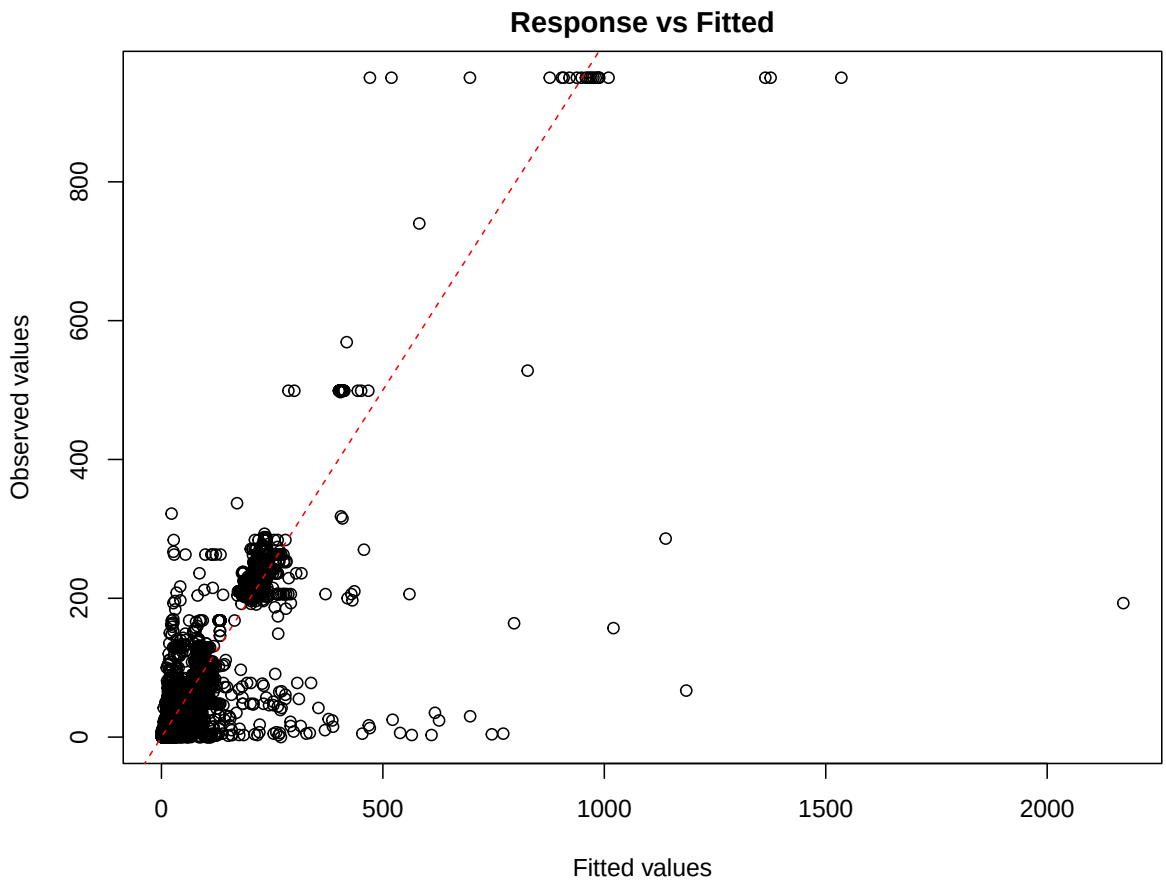

```
### Model 7 : 3c. Temp + Sunlight

Model data dimensions after removing NAs: 5663 x 5
GAM fitted successfully
Model Summary:

Family: Negative Binomial(2.15)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(sunlight_proportion) +
    s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.4591     0.2696   12.83   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df  Chi.sq p-value
s(abundance_t_minus_1) 8.577  8.933 10323.8  <2e-16 ***
s(mean_temp)           7.889  8.646   749.8  <2e-16 ***
s(sunlight_proportion) 7.764  8.594   633.9  <2e-16 ***
s(view_id)             7.766  8.000   422.0  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.616   Deviance explained = 68.2%
-REML =  25387  Scale est. = 1         n = 5663


Diagnostic plots:
```



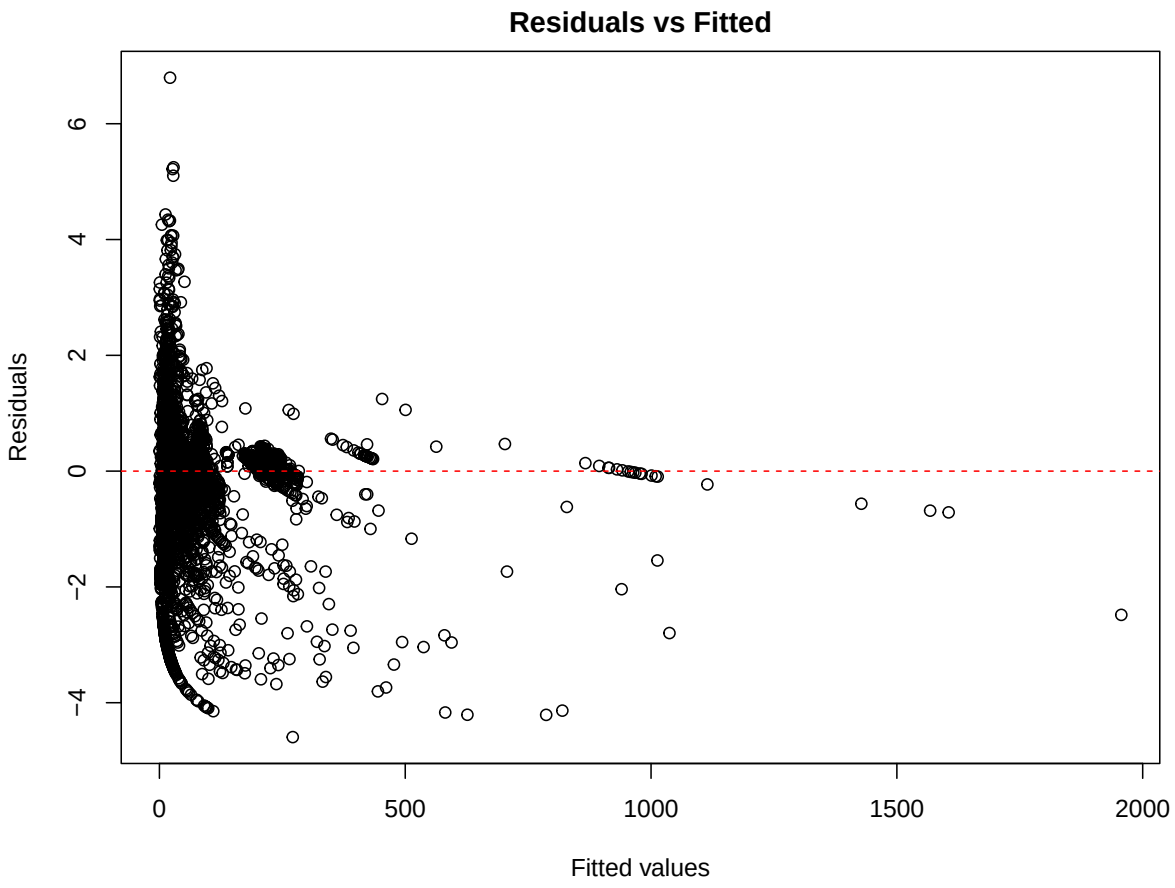

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6
−4
−2
0
2
4

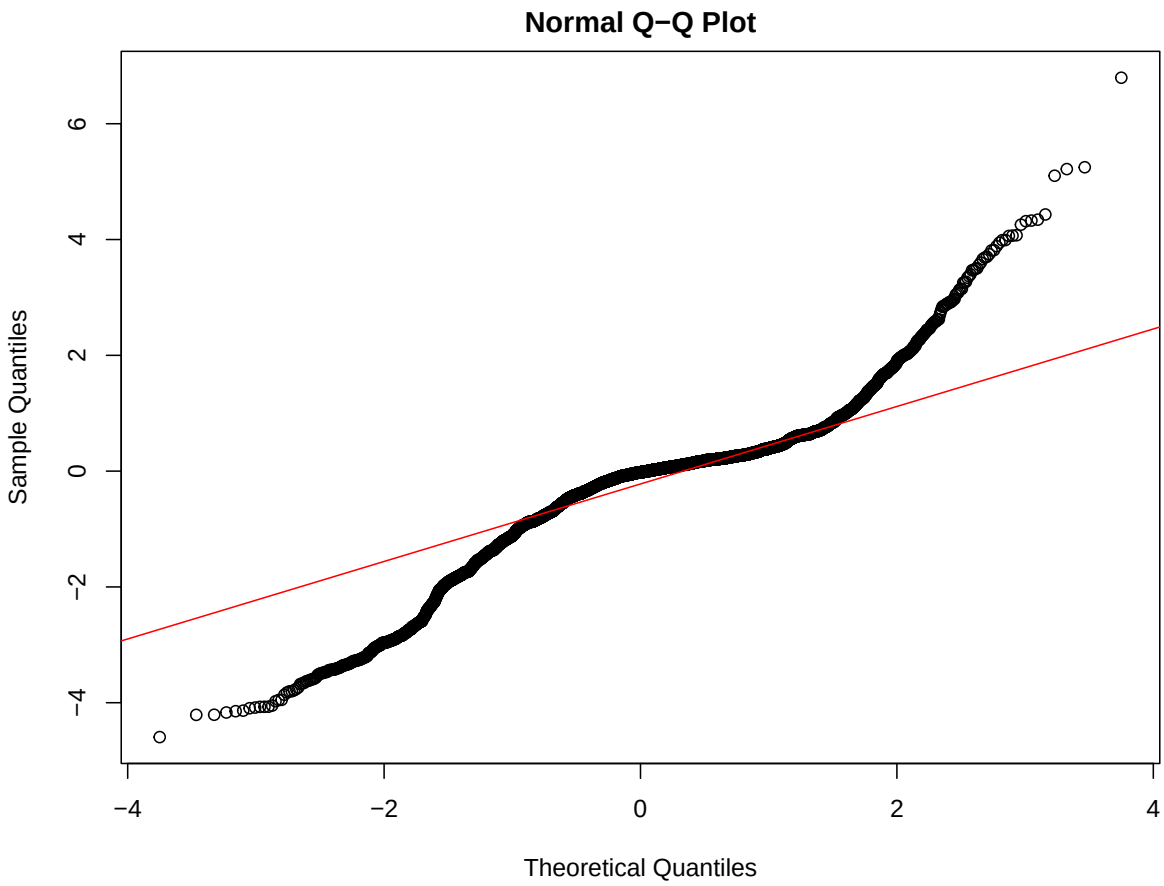

Scale−Location
.|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
Fitted values

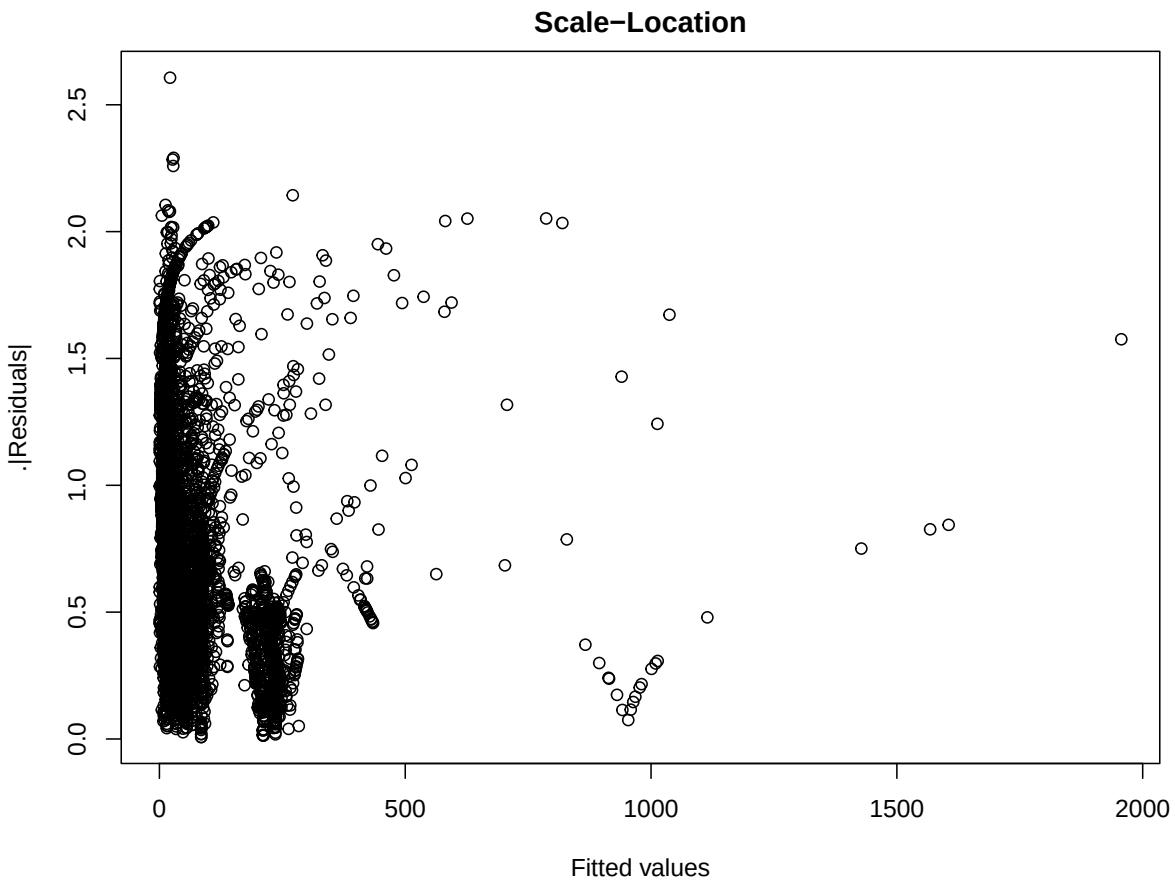


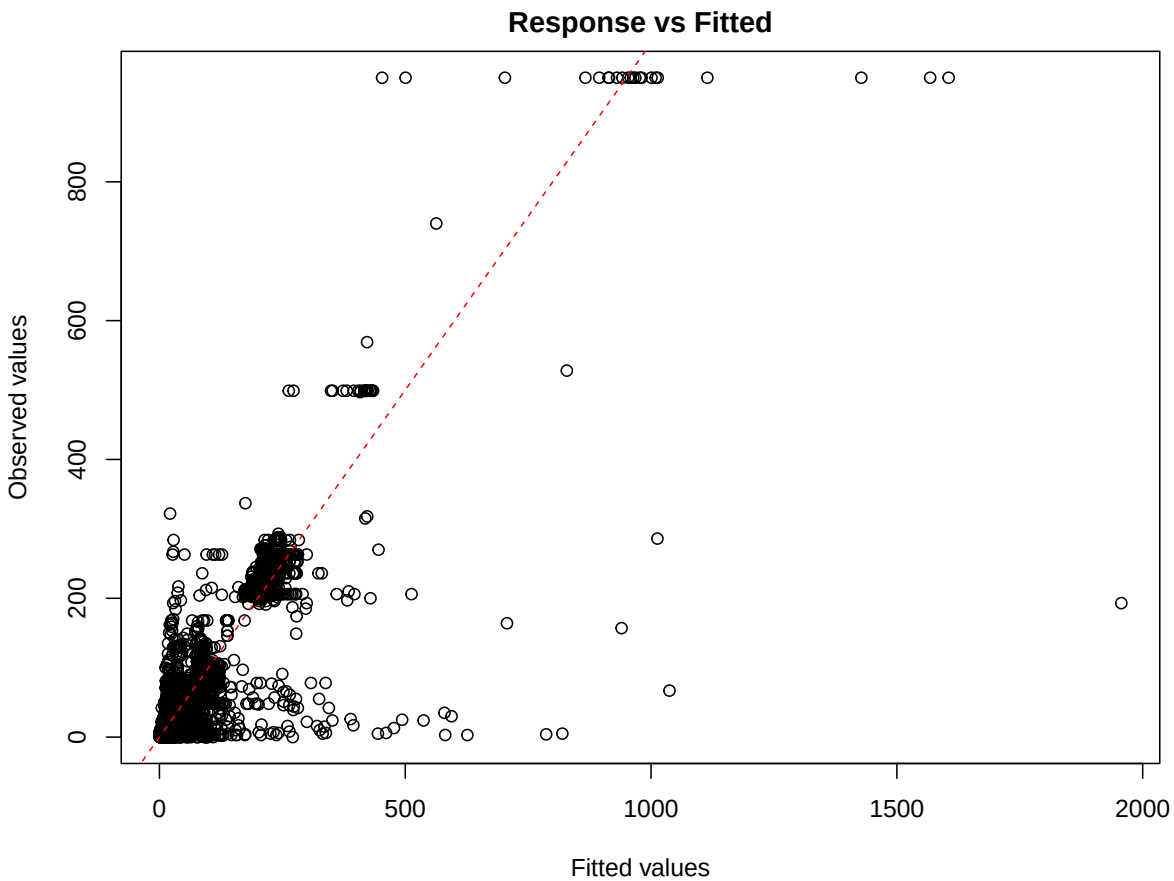

```
### Model 8 : 3d. All additive

Model data dimensions after removing NAs: 5663 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(2.185)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.4933     0.2884   12.11   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df   Chi.sq p-value
s(abundance_t_minus_1) 8.604  8.941 10021.49  <2e-16 ***
s(mean_wind_speed)     5.862  7.031    82.96  <2e-16 ***
s(mean_temp)           8.018  8.716   703.75  <2e-16 ***
s(sunlight_proportion) 7.741  8.580   652.89  <2e-16 ***
s(view_id)             7.792  8.000   407.81  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.648   Deviance explained = 68.7%
-REML =  25359  Scale est. = 1         n = 5663


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
6
4
2
0
−2
−4
0
500
1000
1500
2000

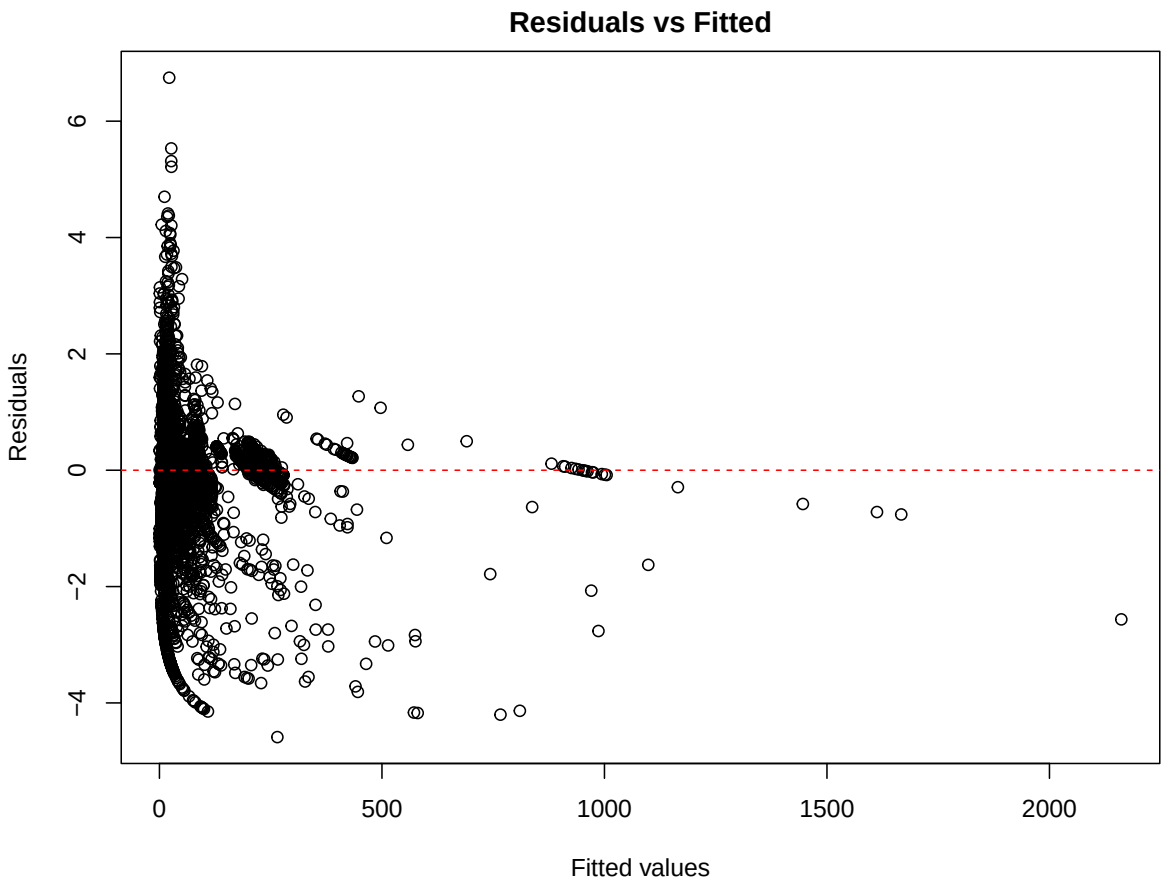

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6
−4
−2
0
2
4

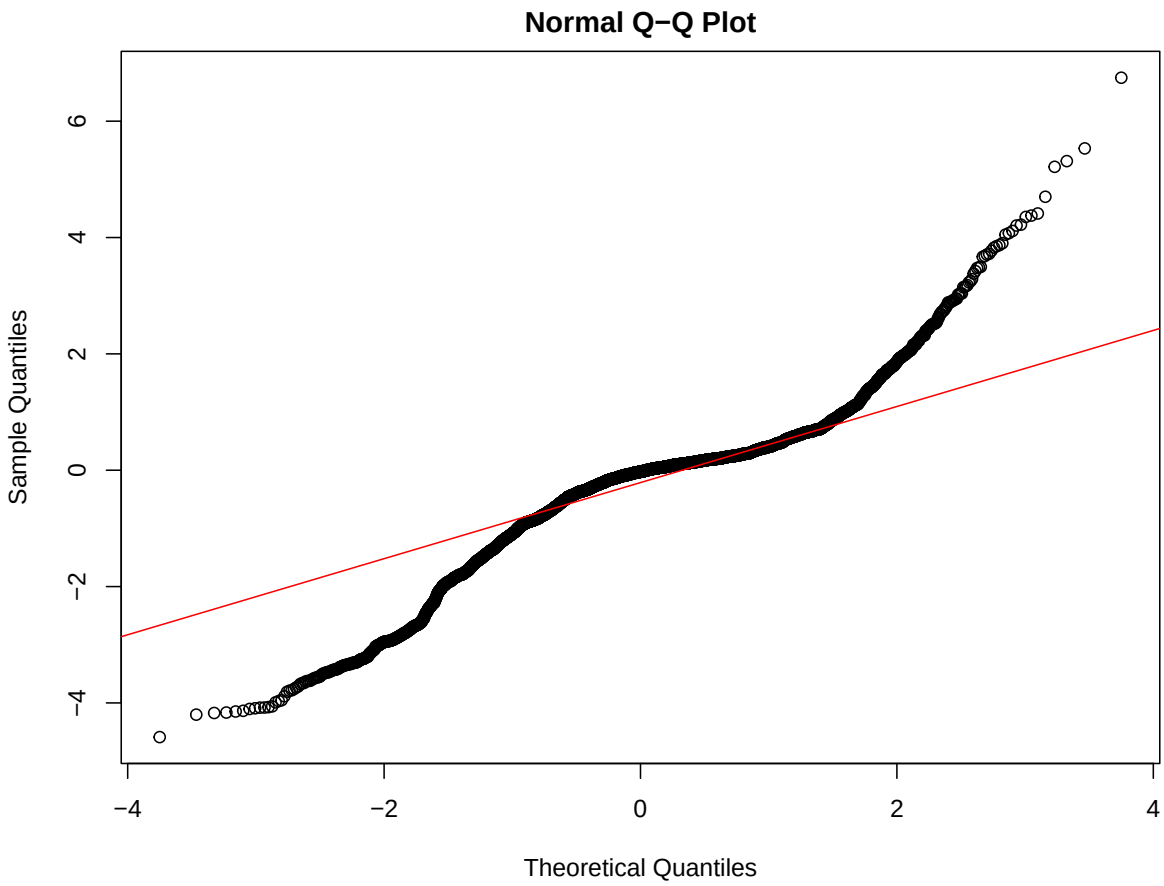

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
Fitted values

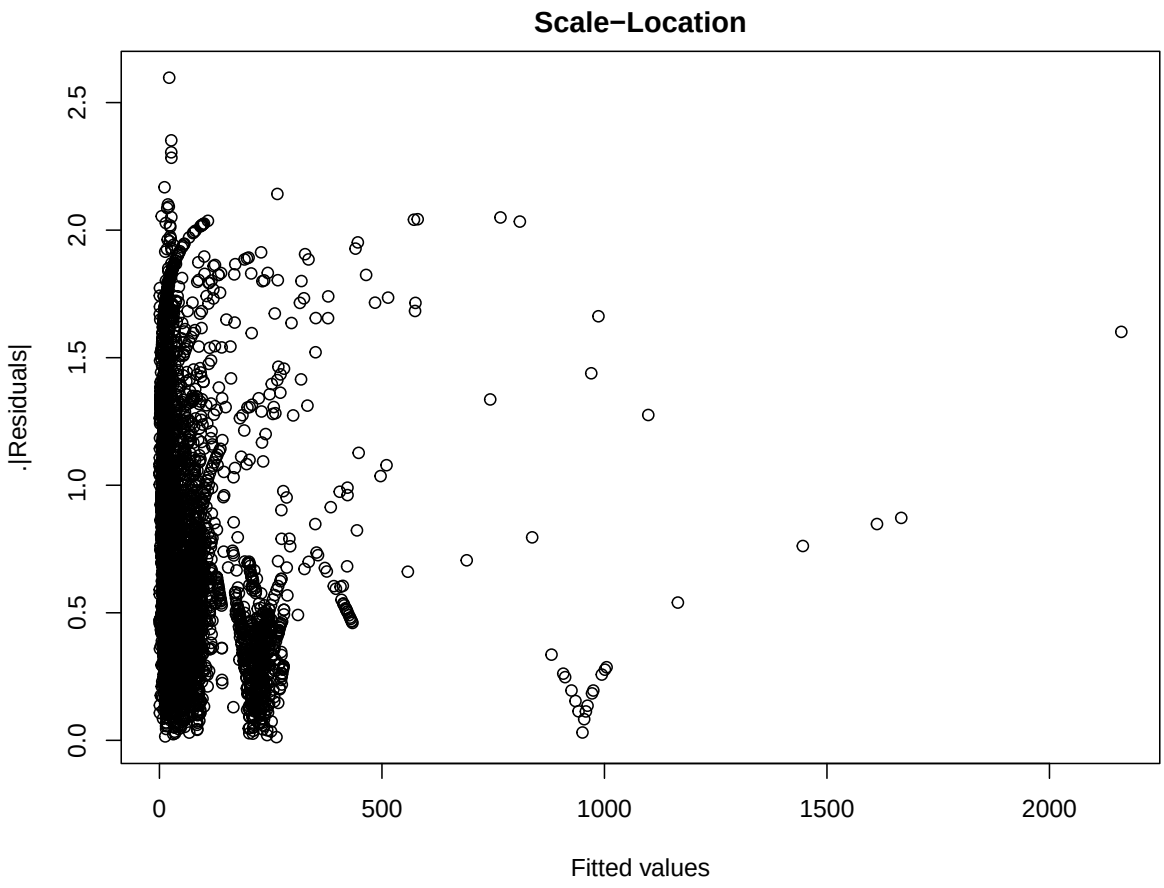


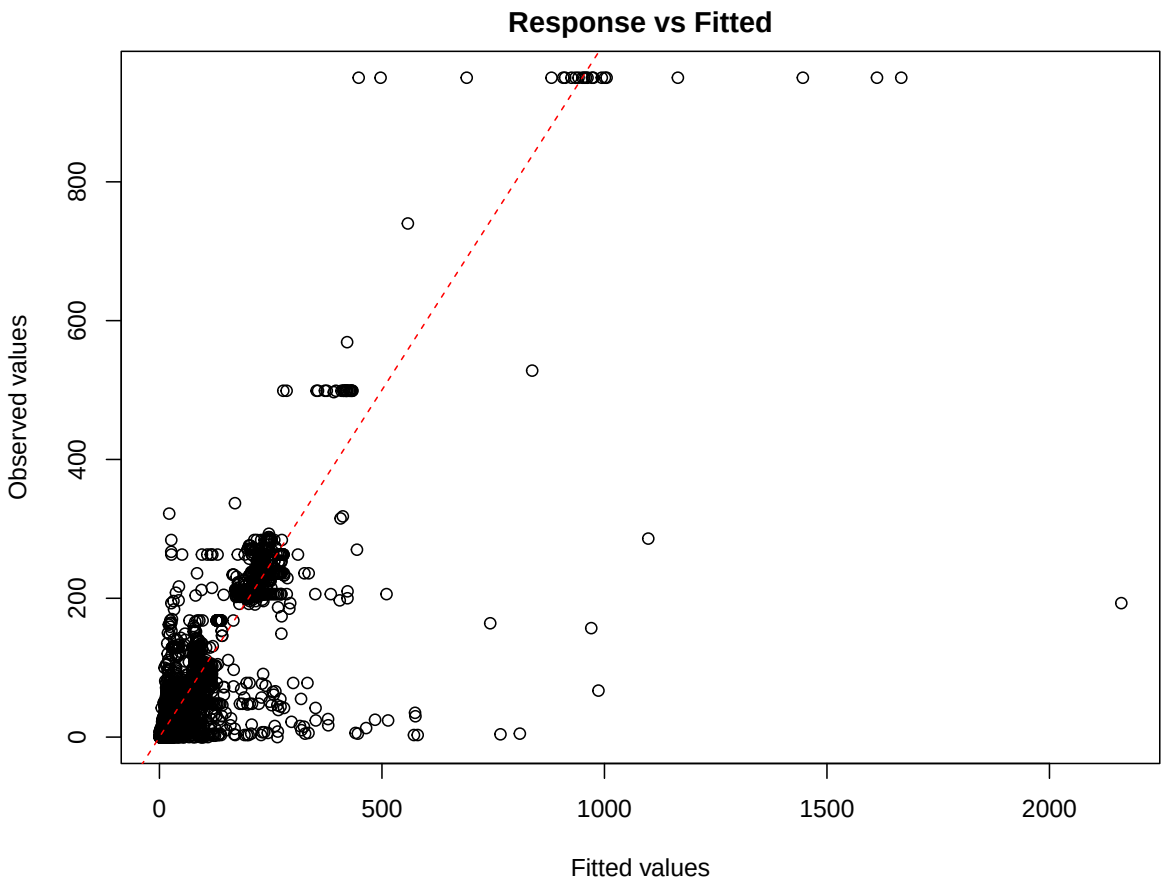

```
### Model 9 : 4a. Wind x Temp interaction

Model data dimensions after removing NAs: 5663 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(2.191)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)   3.4817     0.2919   11.93   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                edf Ref.df  Chi.sq  p-value
s(abundance_t_minus_1)        8.604  8.940 9259.68  < 2e-16 ***
s(mean_wind_speed)            5.521  6.676   42.98 2.52e-06 ***
s(mean_temp)                  8.036  8.720  562.92  < 2e-16 ***
s(sunlight_proportion)        7.682  8.543  596.03  < 2e-16 ***
ti(mean_wind_speed,mean_temp) 4.818  5.853   22.23 0.000968 ***
s(view_id)                    7.796  8.000  391.68  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.632   Deviance explained = 68.8%
-REML =  25352  Scale est. = 1         n = 5663


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
6
4
2
0
−2
−4
0
500
1000
1500
2000

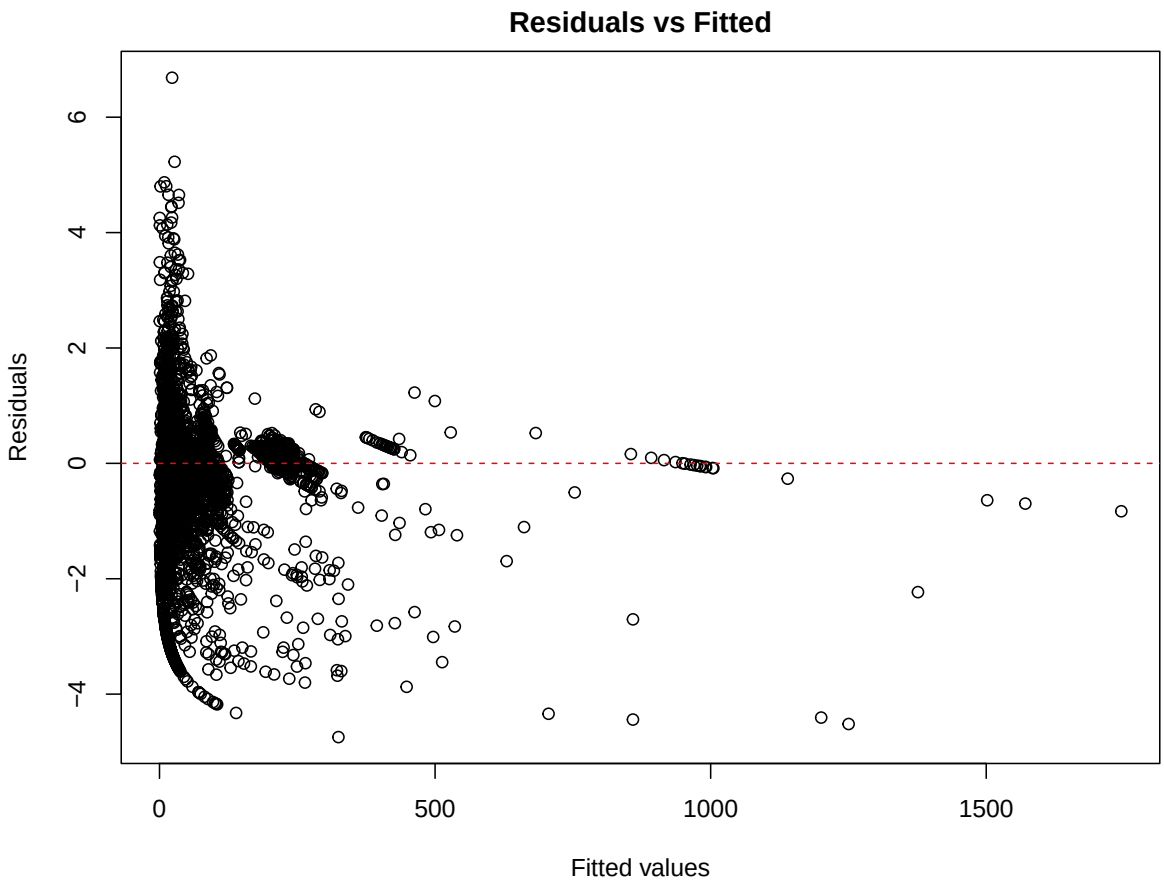

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6
−4
−2
0
2
4

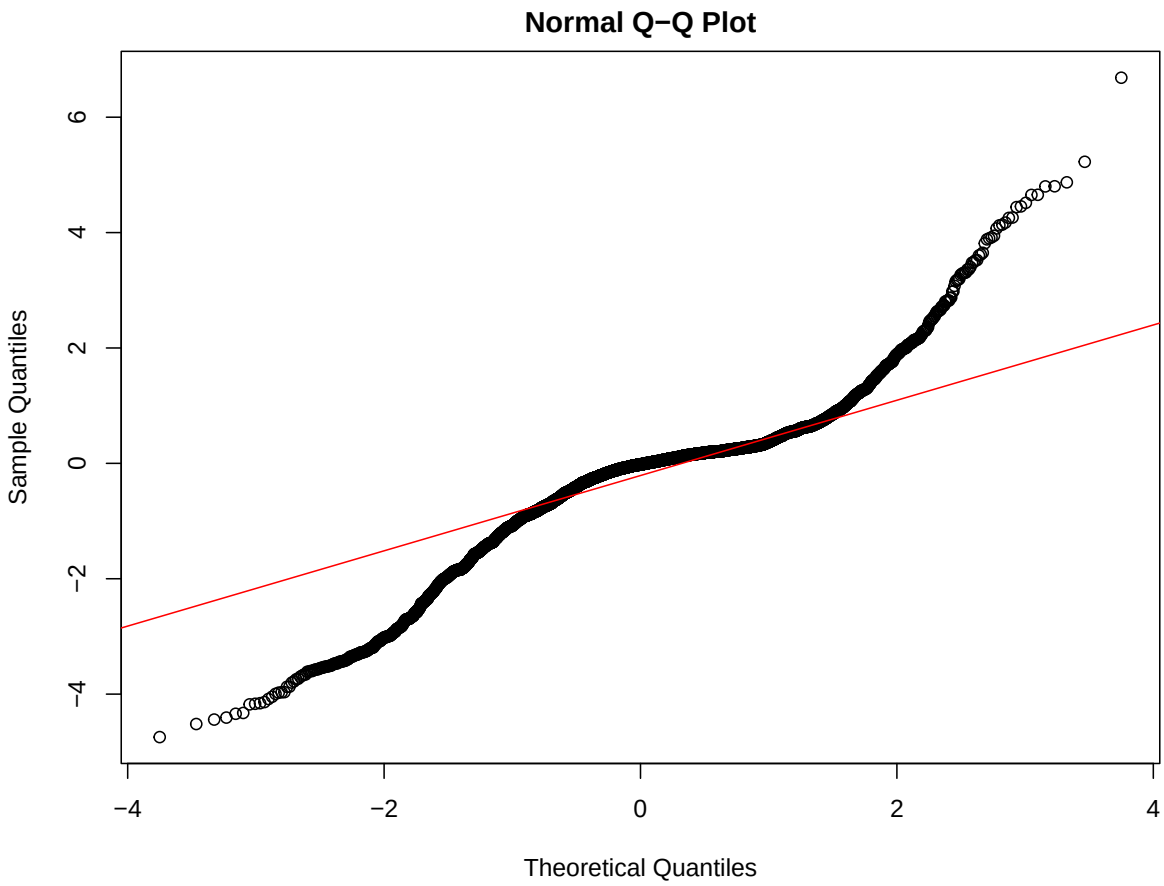

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
Fitted values

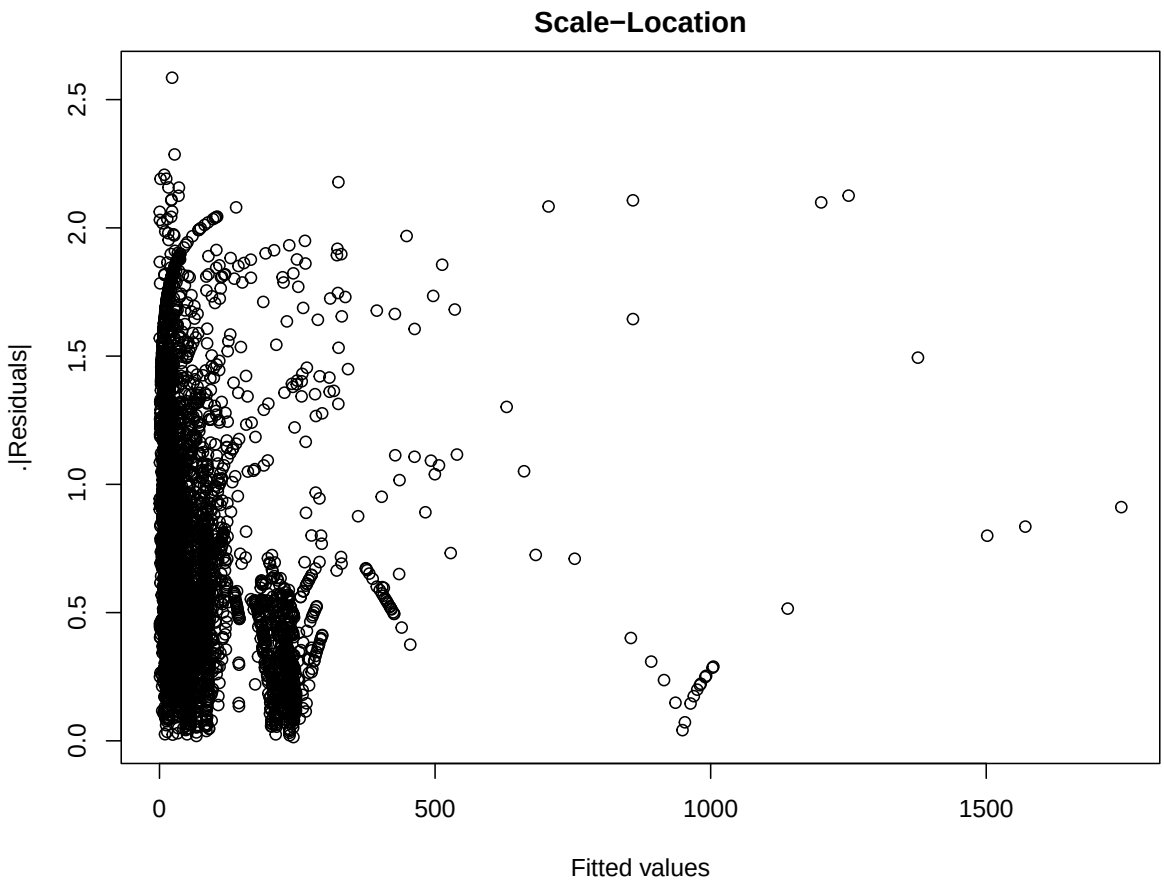


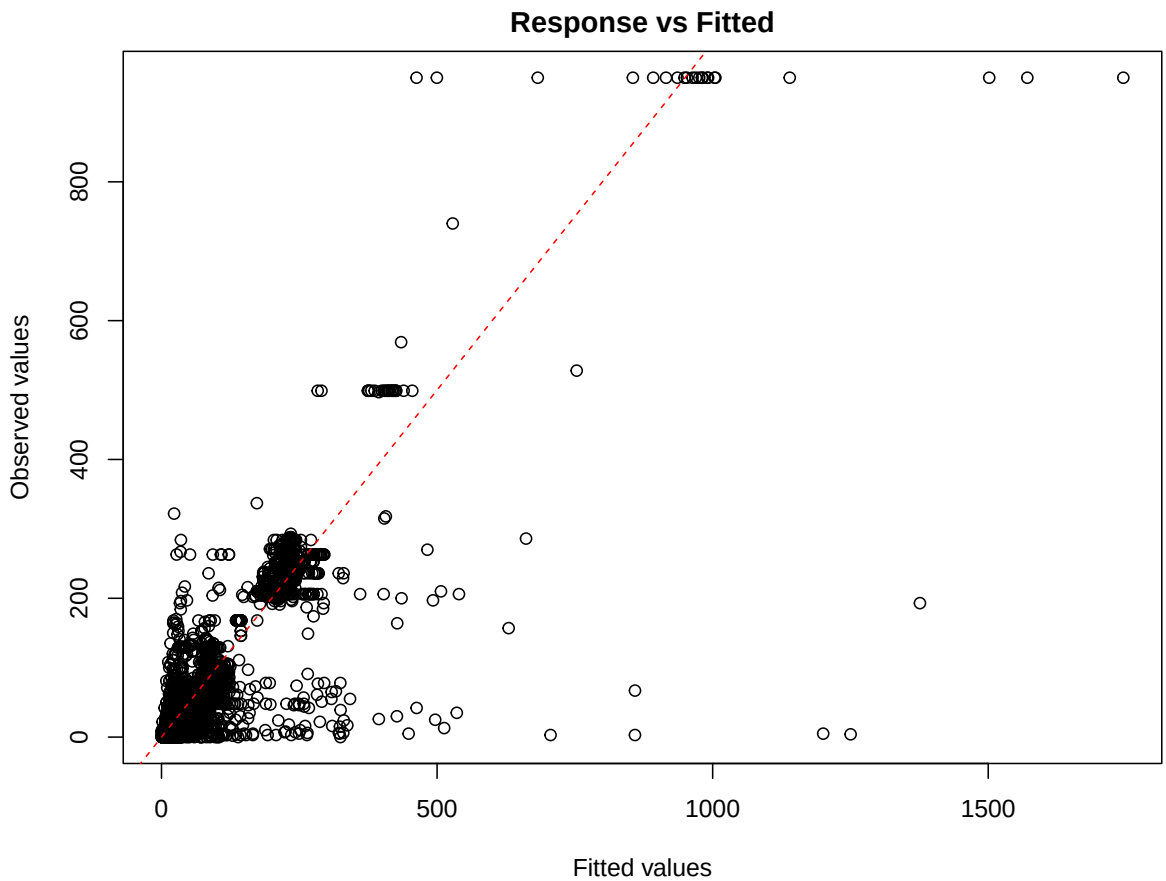

```
### Model 10 : 4b. Wind x Sunlight interaction

Model data dimensions after removing NAs: 5663 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(2.264)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    sunlight_proportion) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)   3.4876     0.2893   12.06   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                          edf Ref.df   Chi.sq p-value
s(abundance_t_minus_1)                  8.623  8.946 10249.44  <2e-16 ***
s(mean_wind_speed)                      6.122  7.292    71.91  <2e-16 ***
s(mean_temp)                            8.474  8.908   795.06  <2e-16 ***
s(sunlight_proportion)                  7.743  8.579   577.33  <2e-16 ***
ti(mean_wind_speed,sunlight_proportion) 9.294 10.603   185.09  <2e-16 ***
s(view_id)                              7.797  8.000   406.01  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.661   Deviance explained = 69.6%
-REML =  25289  Scale est. = 1         n = 5663


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
6
4
2
0
−2
−4
0
500
1000
1500

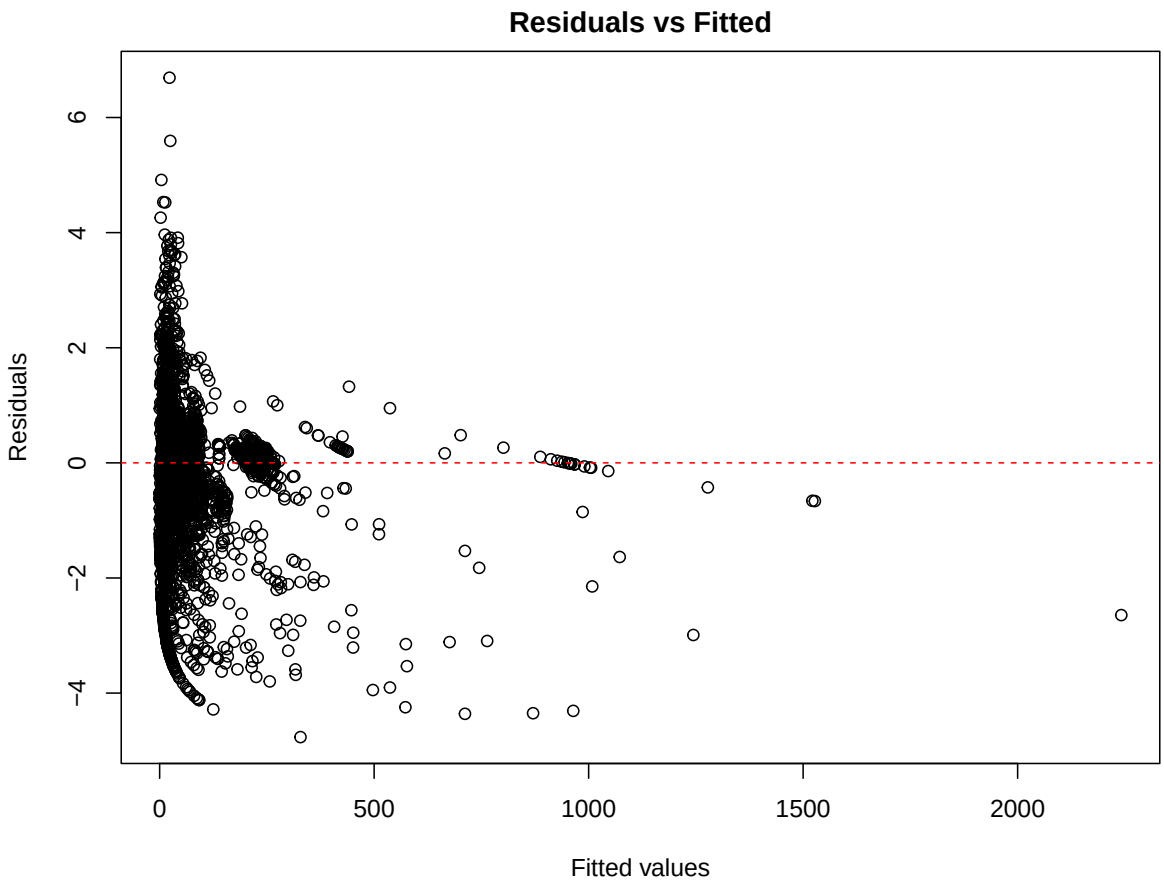

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6
−4
−2
0
2
4

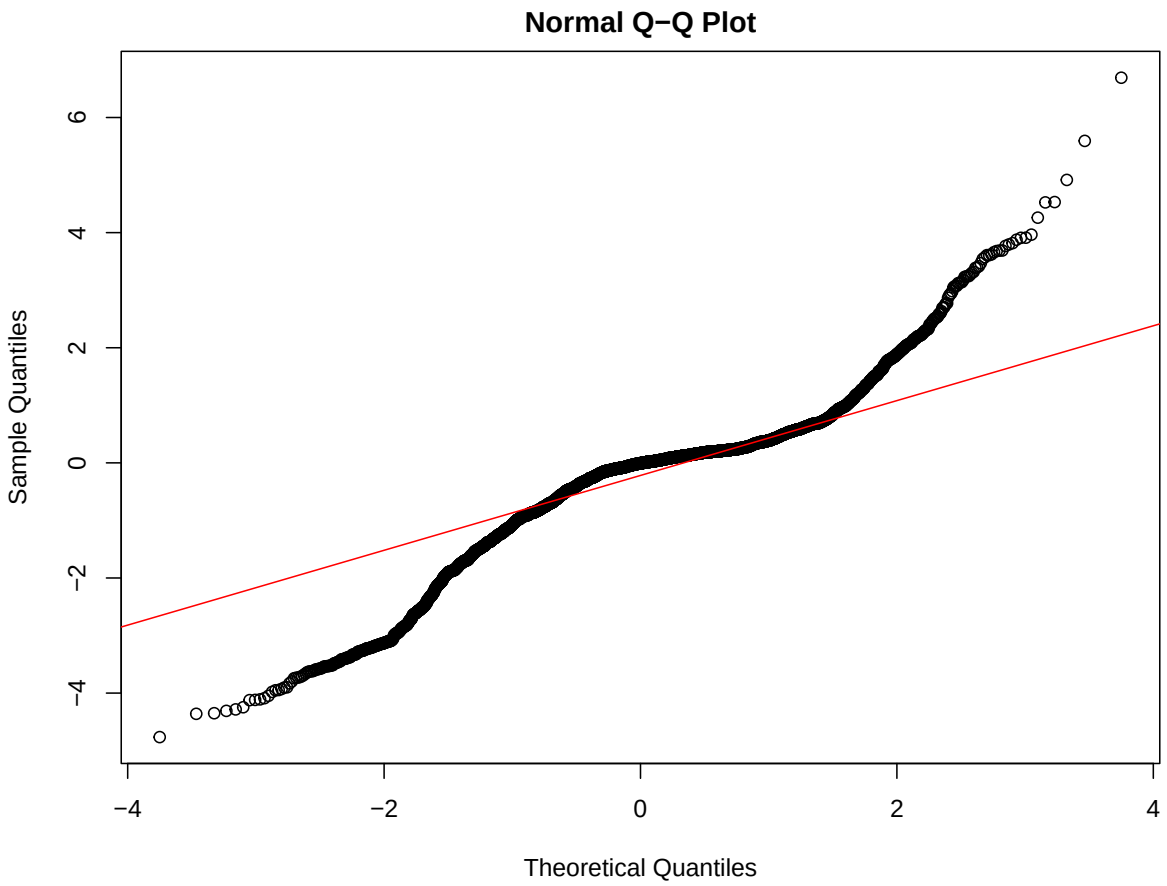

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
Fitted values

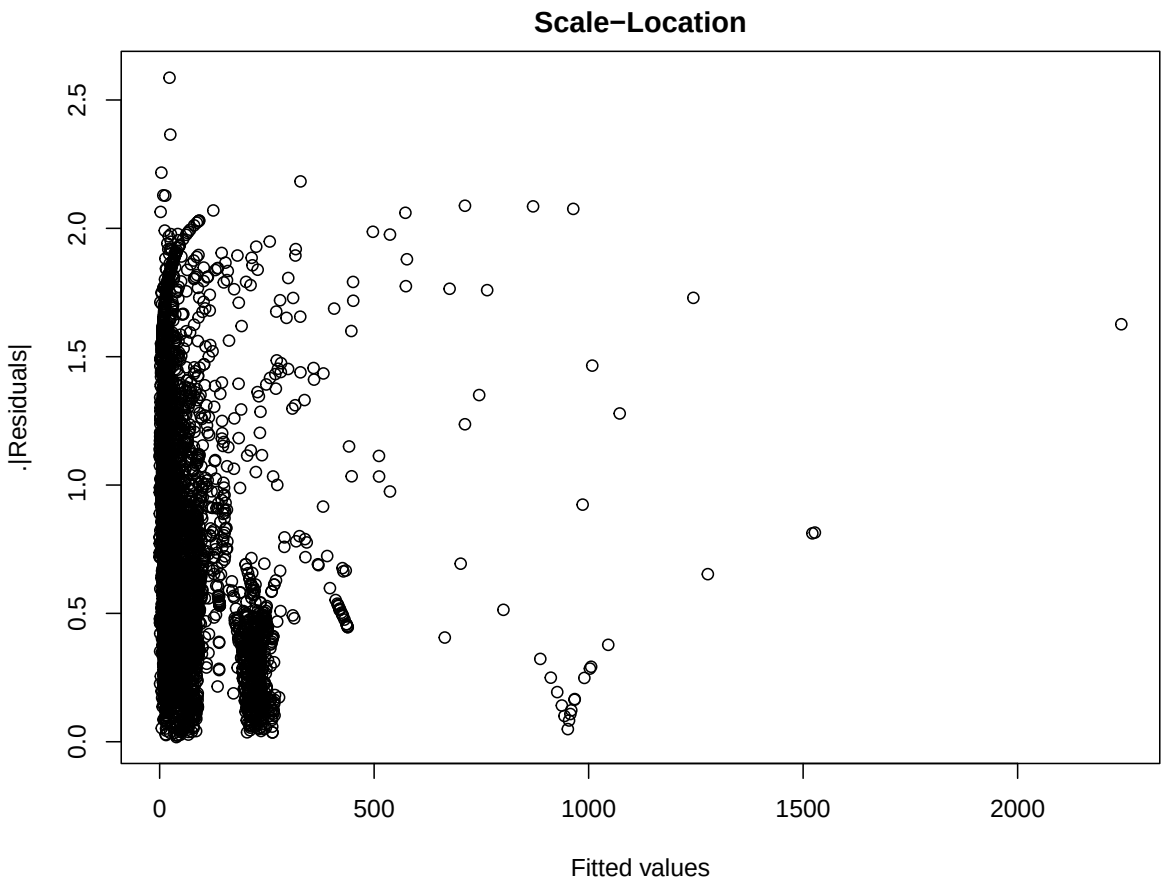


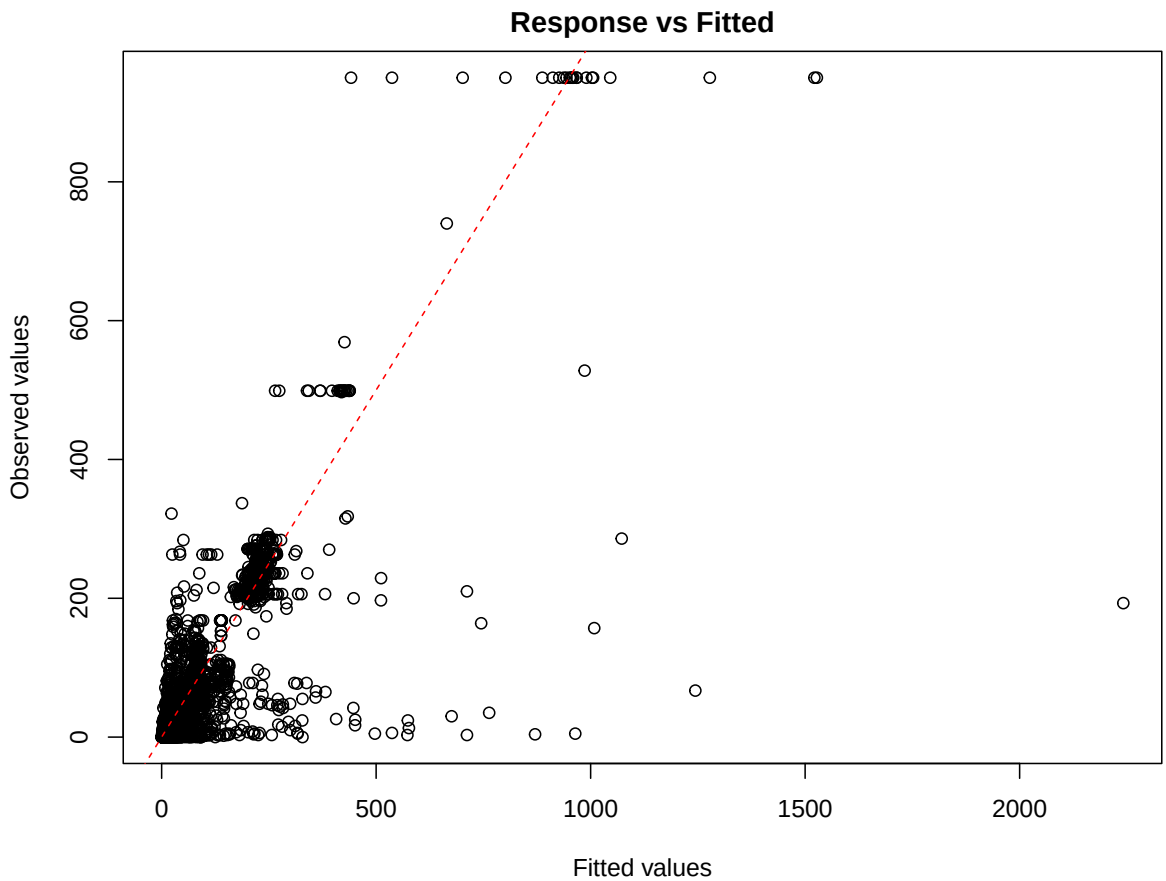

```
### Model 11 : 4c. Temp x Sunlight interaction

Model data dimensions after removing NAs: 5663 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(2.281)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_temp, sunlight_proportion) +
    s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)   3.5073     0.2839   12.36   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                     edf Ref.df   Chi.sq p-value
s(abundance_t_minus_1)             8.621  8.945 10280.85  <2e-16 ***
s(mean_wind_speed)                 6.231  7.395    95.95  <2e-16 ***
s(mean_temp)                       8.044  8.692   430.00  <2e-16 ***
s(sunlight_proportion)             7.759  8.589   151.82  <2e-16 ***
ti(mean_temp,sunlight_proportion) 12.307 13.566   228.56  <2e-16 ***
s(view_id)                         7.791  8.000   352.49  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.587   Deviance explained =   70%
-REML =  25258  Scale est. = 1         n = 5663


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
6
4
2
0
−2
−4
0
500
1000
1500
2000

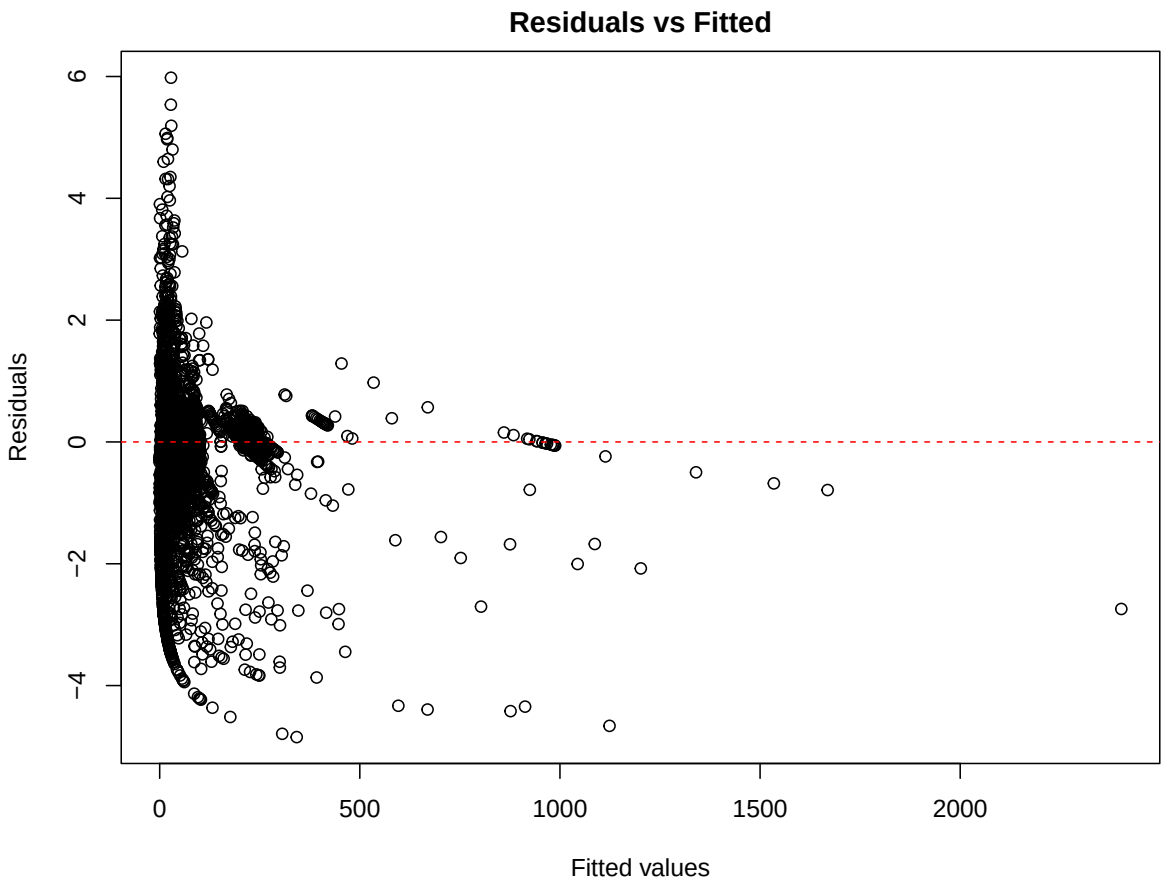

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6

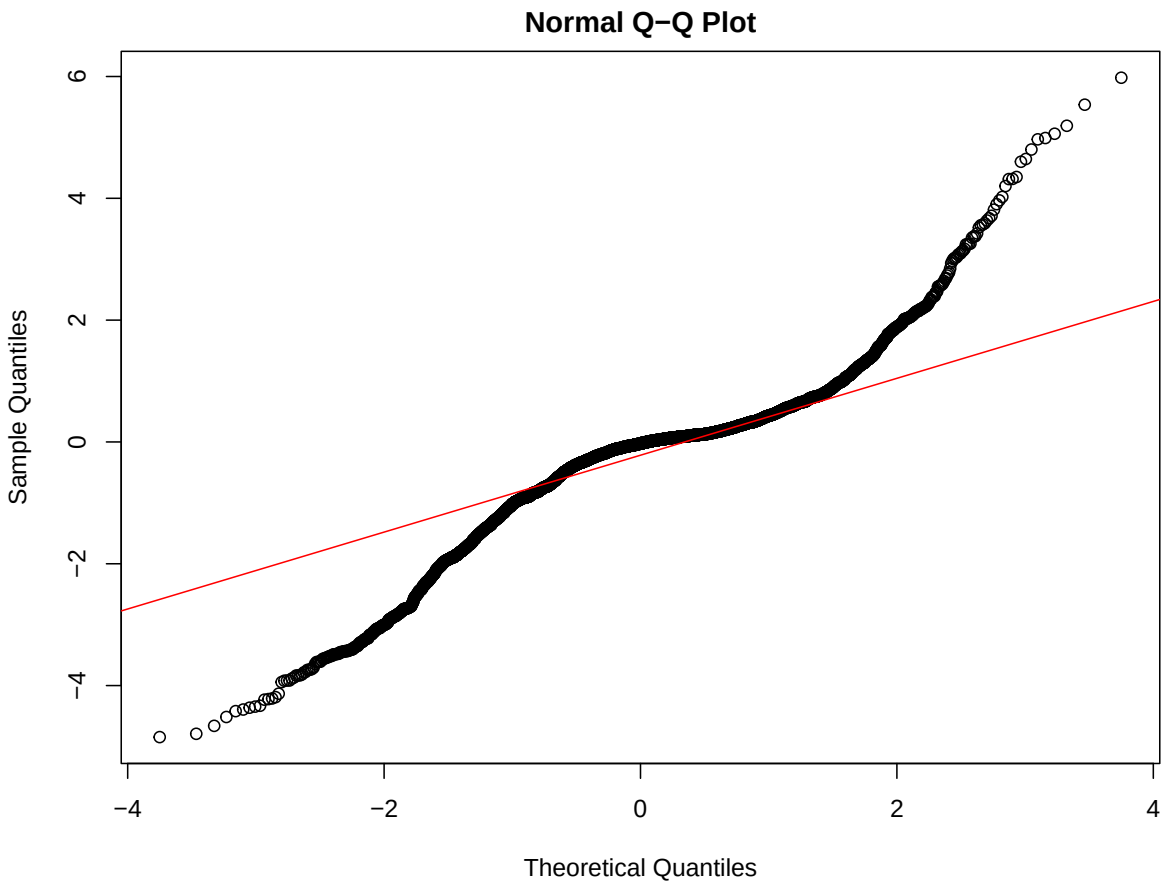

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
Fitted values

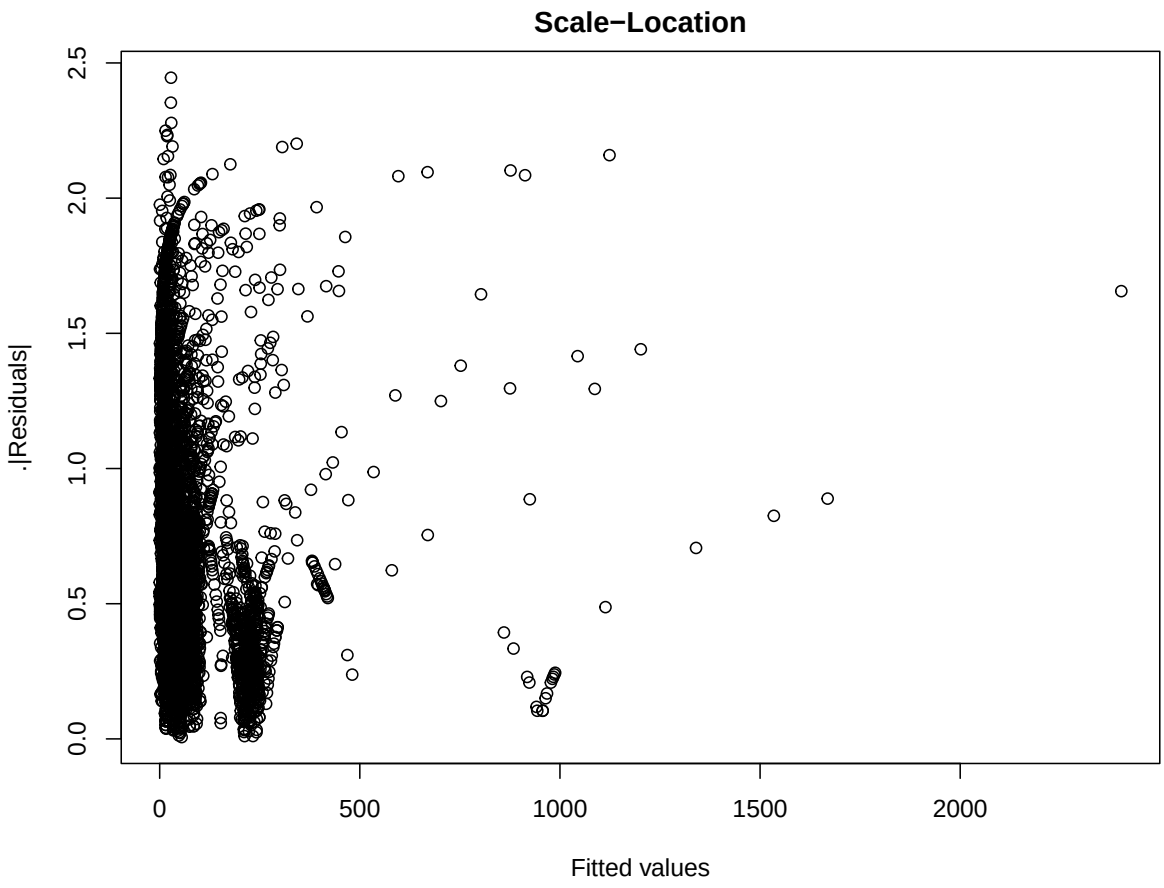


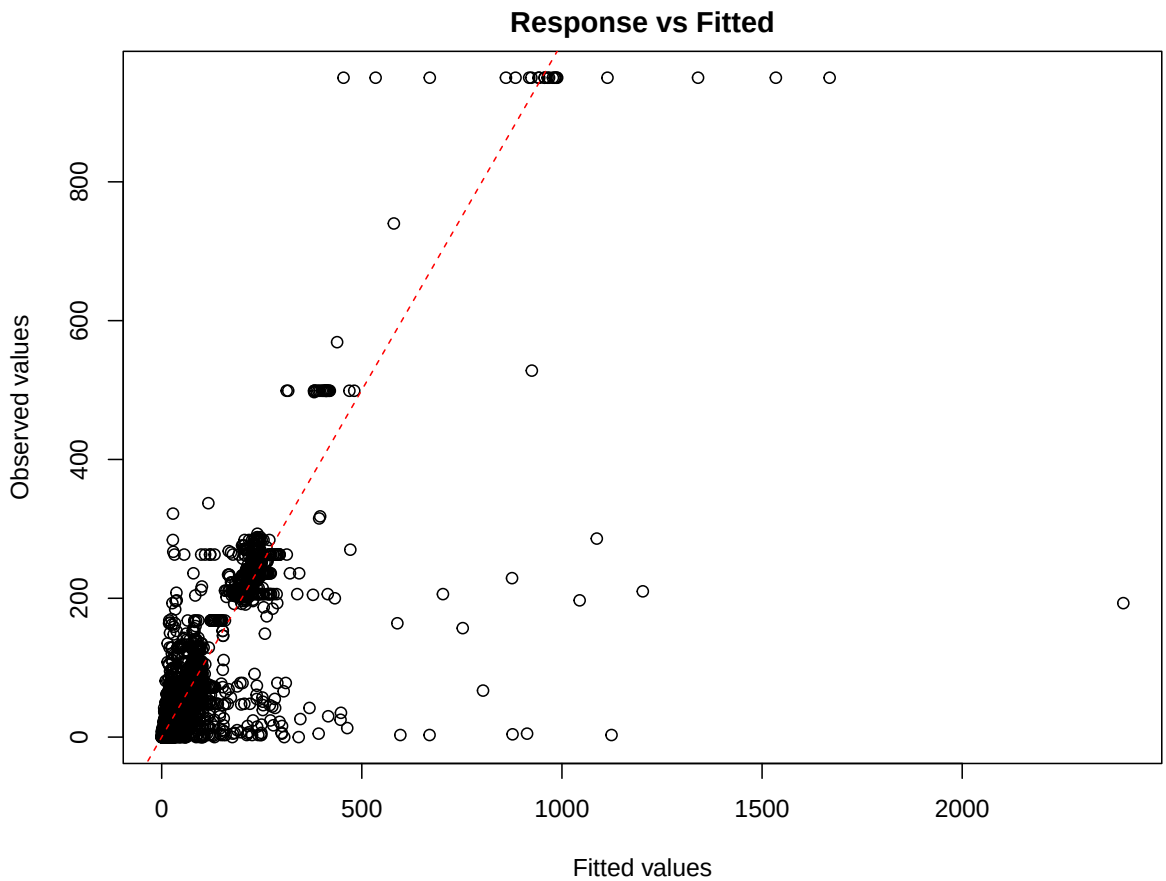

```
### Model 12 : 5a. Wind x Temp + Wind x Sun

Model data dimensions after removing NAs: 5663 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(2.355)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + ti(mean_wind_speed, sunlight_proportion) + s(view_id,
    bs = "re")

Parametric coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3532     0.2834   11.83   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                          edf Ref.df Chi.sq p-value
s(abundance_t_minus_1)                  8.621  8.945 9407.3  <2e-16 ***
s(mean_wind_speed)                      6.437  7.621  115.2  <2e-16 ***
s(mean_temp)                            8.490  8.903  628.7  <2e-16 ***
s(sunlight_proportion)                  6.868  7.938  496.6  <2e-16 ***
ti(mean_wind_speed,mean_temp)          13.855 14.345  232.6  <2e-16 ***
ti(mean_wind_speed,sunlight_proportion)  9.560 10.736  263.4  <2e-16 ***
s(view_id)                              7.791  8.000  396.3  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.573   Deviance explained = 70.9%
-REML =  25207  Scale est. = 1         n = 5663


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
6
4
2
0
−2
−4
0
500
1000
1500
2000

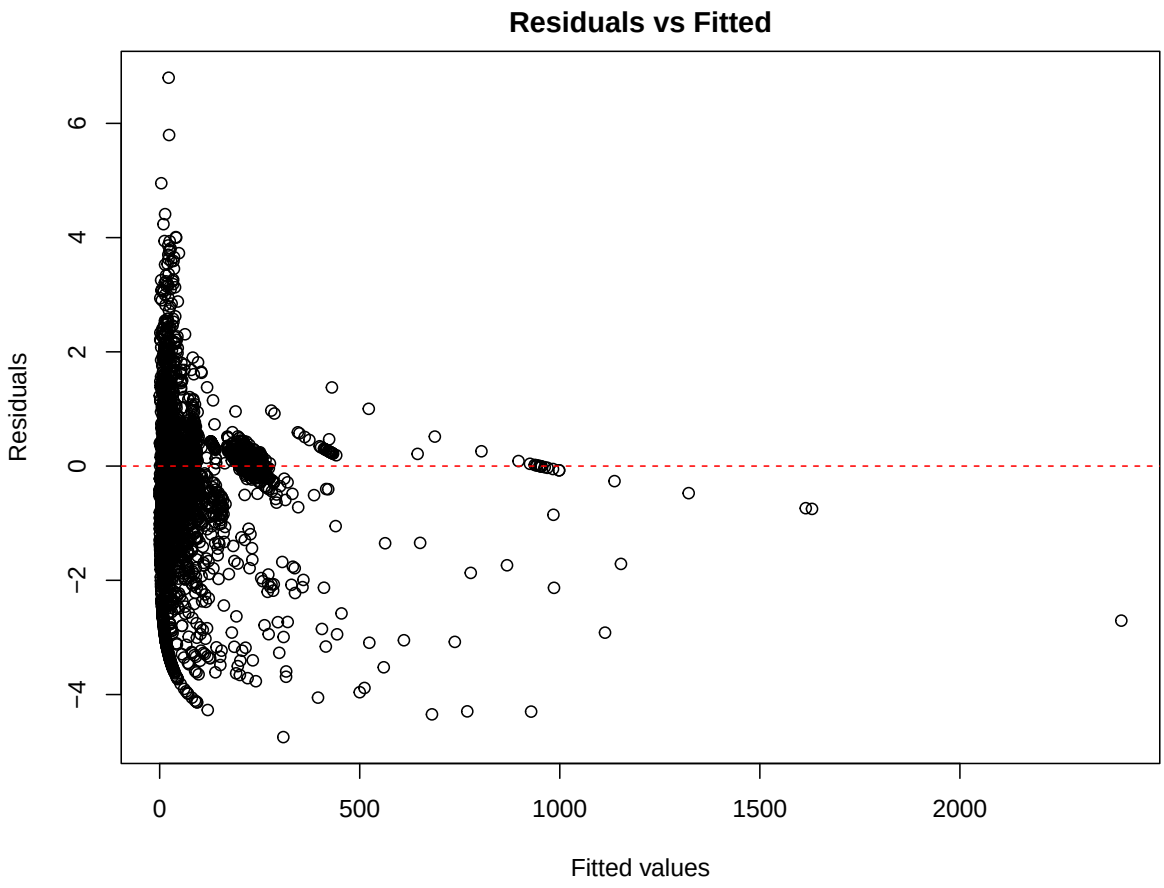

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
6
4
2
0
−2
−4
−4
−2
0
2
4

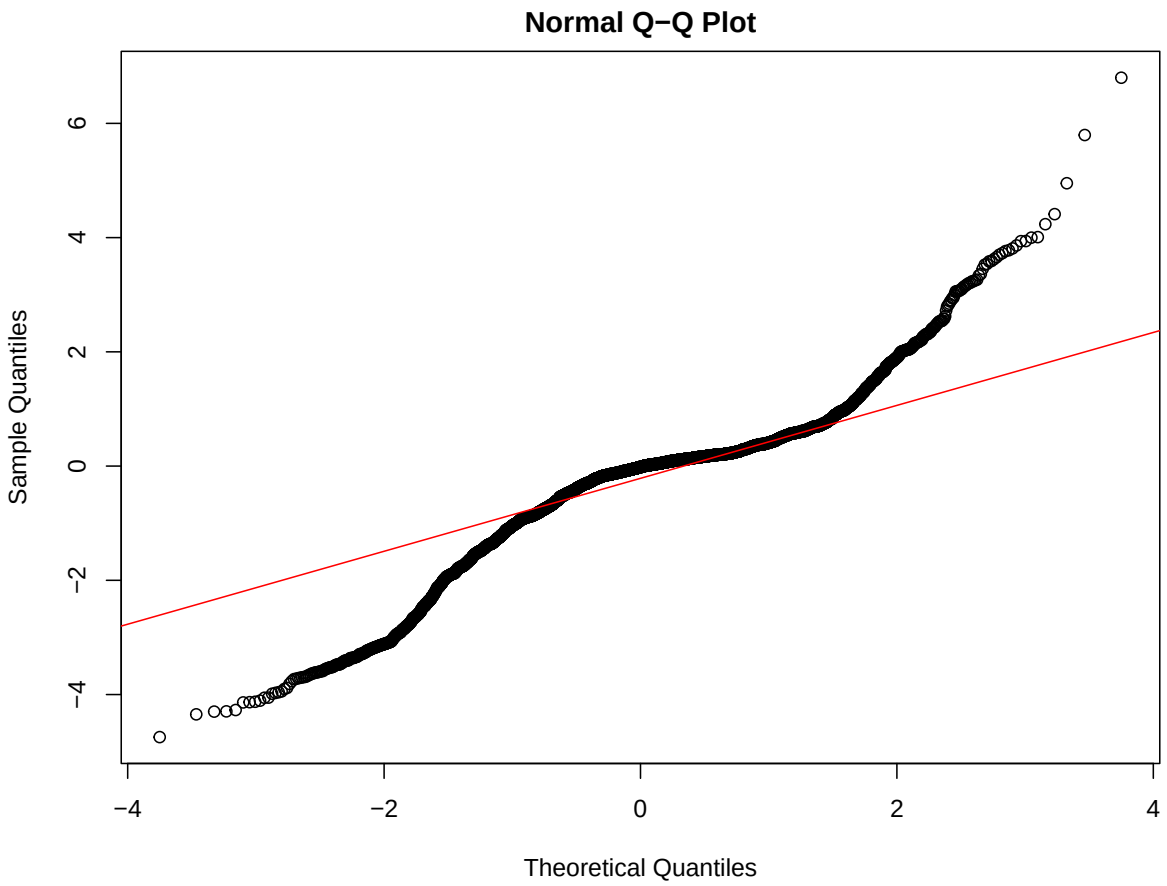

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
Fitted values

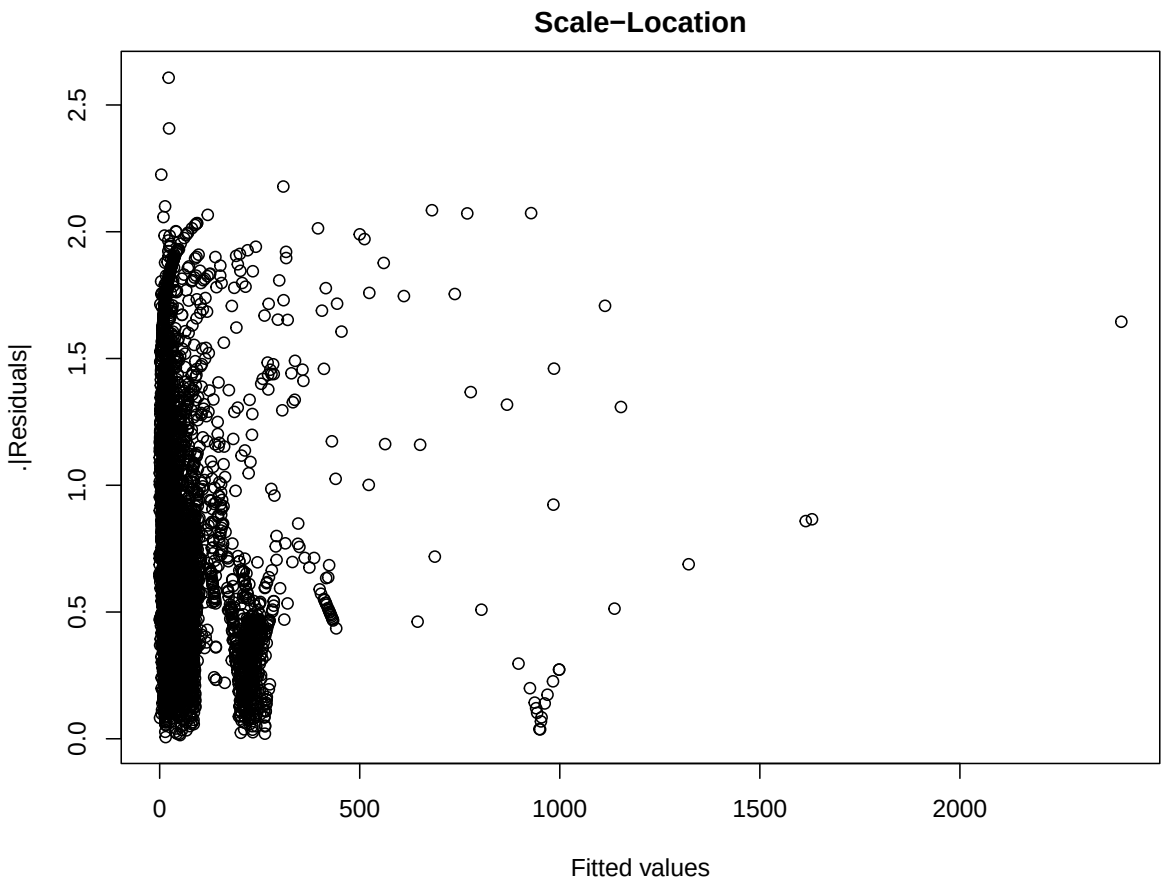


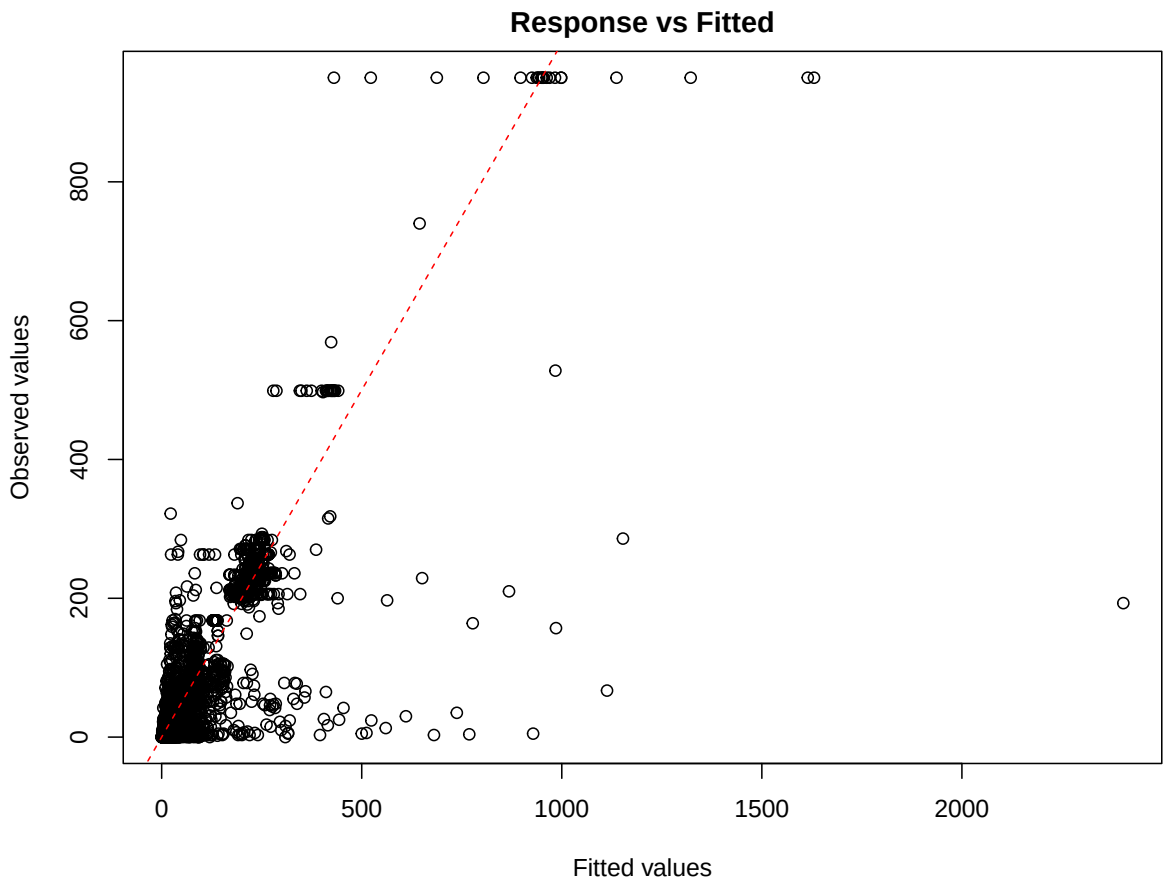

```
### Model 13 : 5b. Wind x Temp + Temp x Sun

Model data dimensions after removing NAs: 5663 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(2.292)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + ti(mean_temp, sunlight_proportion) + s(view_id,
    bs = "re")

Parametric coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.4924     0.2914   11.98   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                     edf Ref.df  Chi.sq  p-value
s(abundance_t_minus_1)             8.624  8.946 9533.42  < 2e-16 ***
s(mean_wind_speed)                 6.003  7.174   62.53  < 2e-16 ***
s(mean_temp)                       8.232  8.784  349.34  < 2e-16 ***
s(sunlight_proportion)             7.708  8.558  148.91  < 2e-16 ***
ti(mean_wind_speed,mean_temp)      4.164  4.702   30.53 1.46e-05 ***
ti(mean_temp,sunlight_proportion) 12.451 13.689  231.37  < 2e-16 ***
s(view_id)                         7.800  8.000  357.19  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.576   Deviance explained = 70.2%
-REML =  25248  Scale est. = 1         n = 5663


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
6
4
2
0
−2
−4
0
500
1000
1500
2000

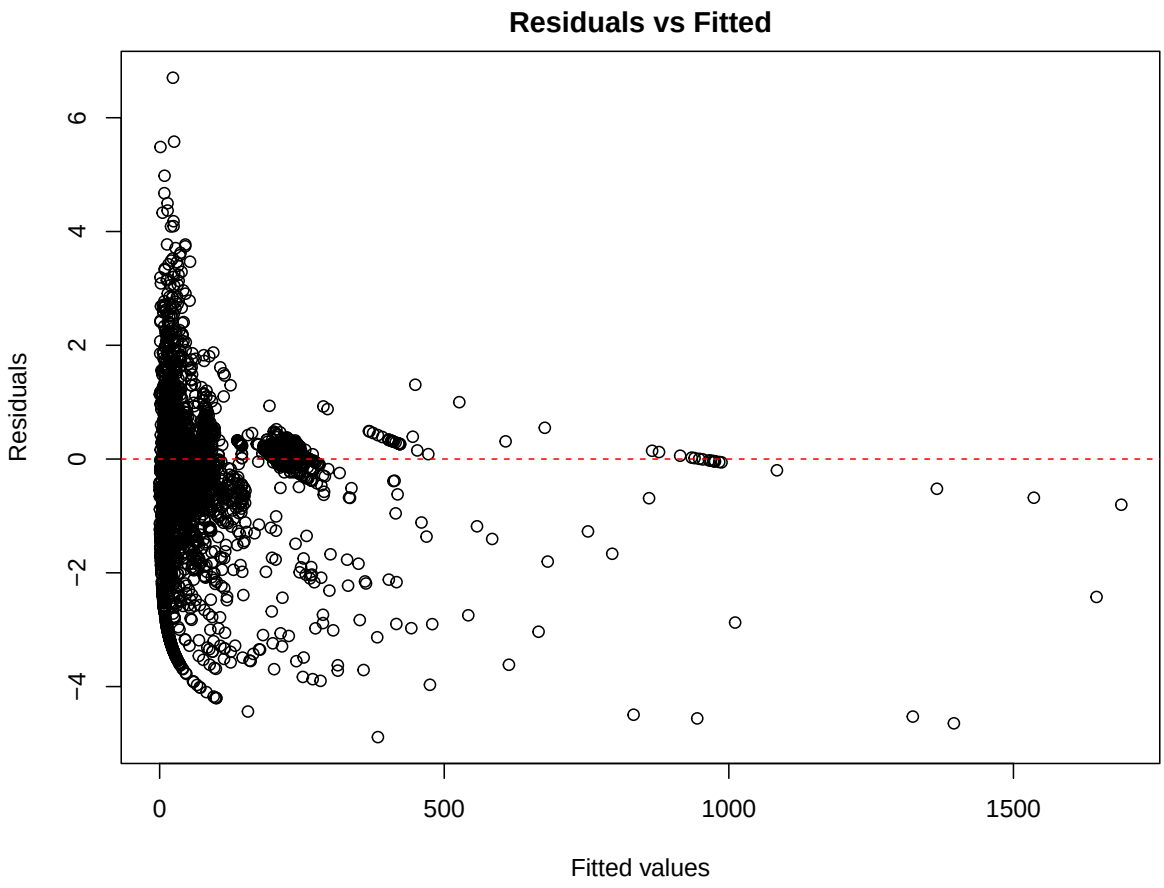

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6

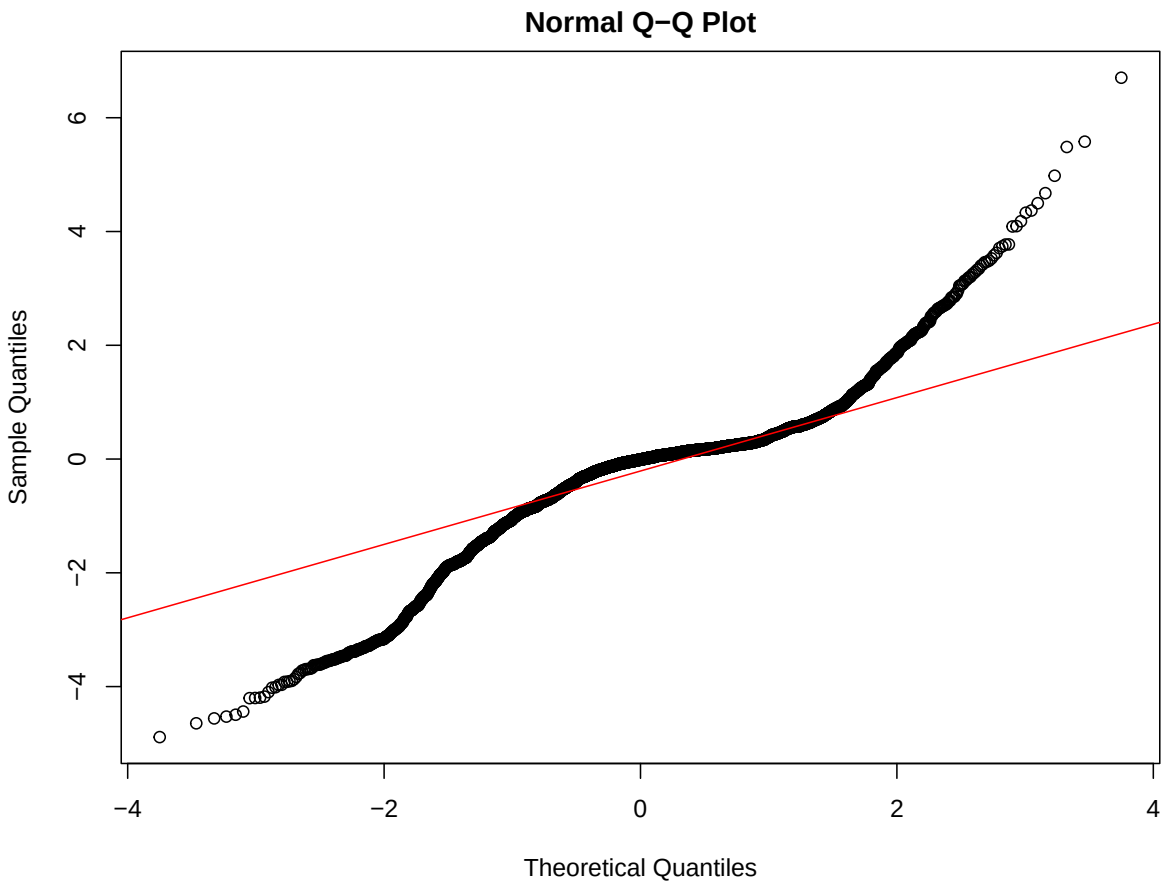

Scale−Location
.|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
Fitted values

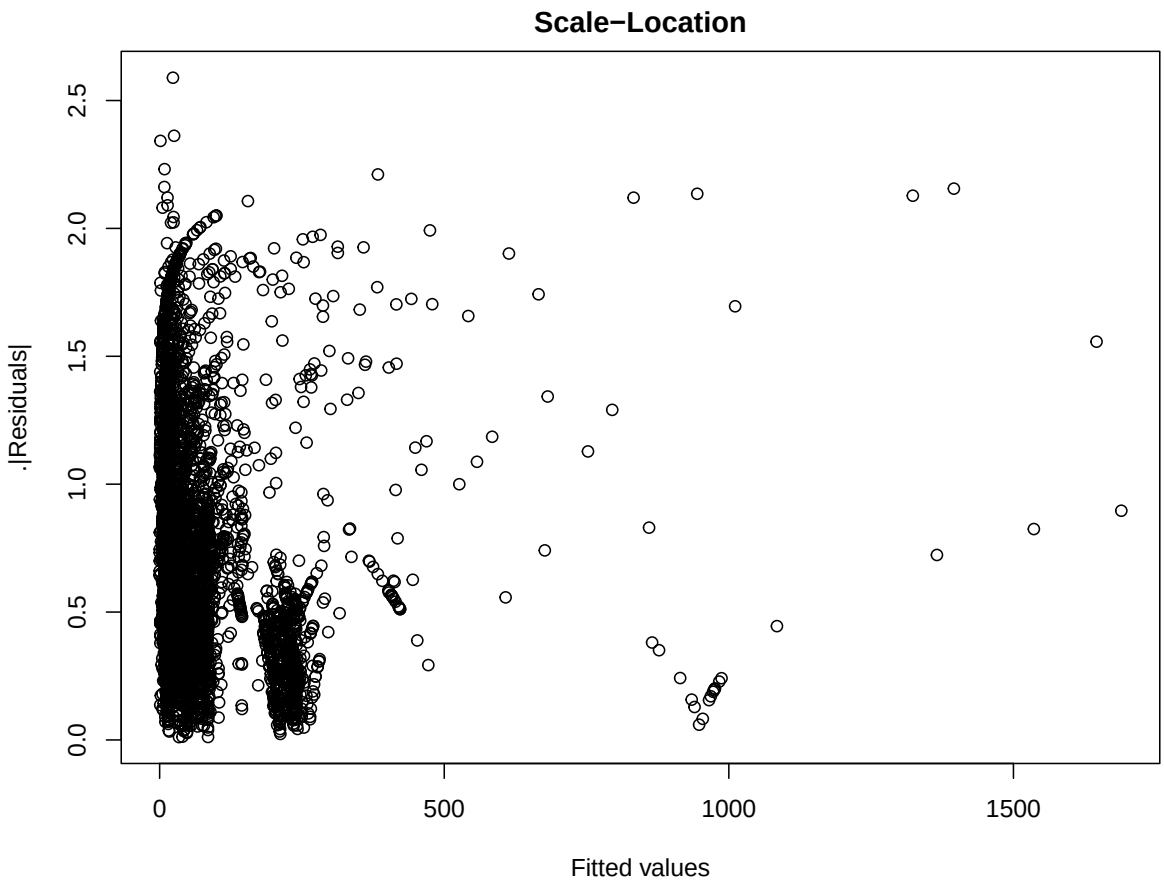


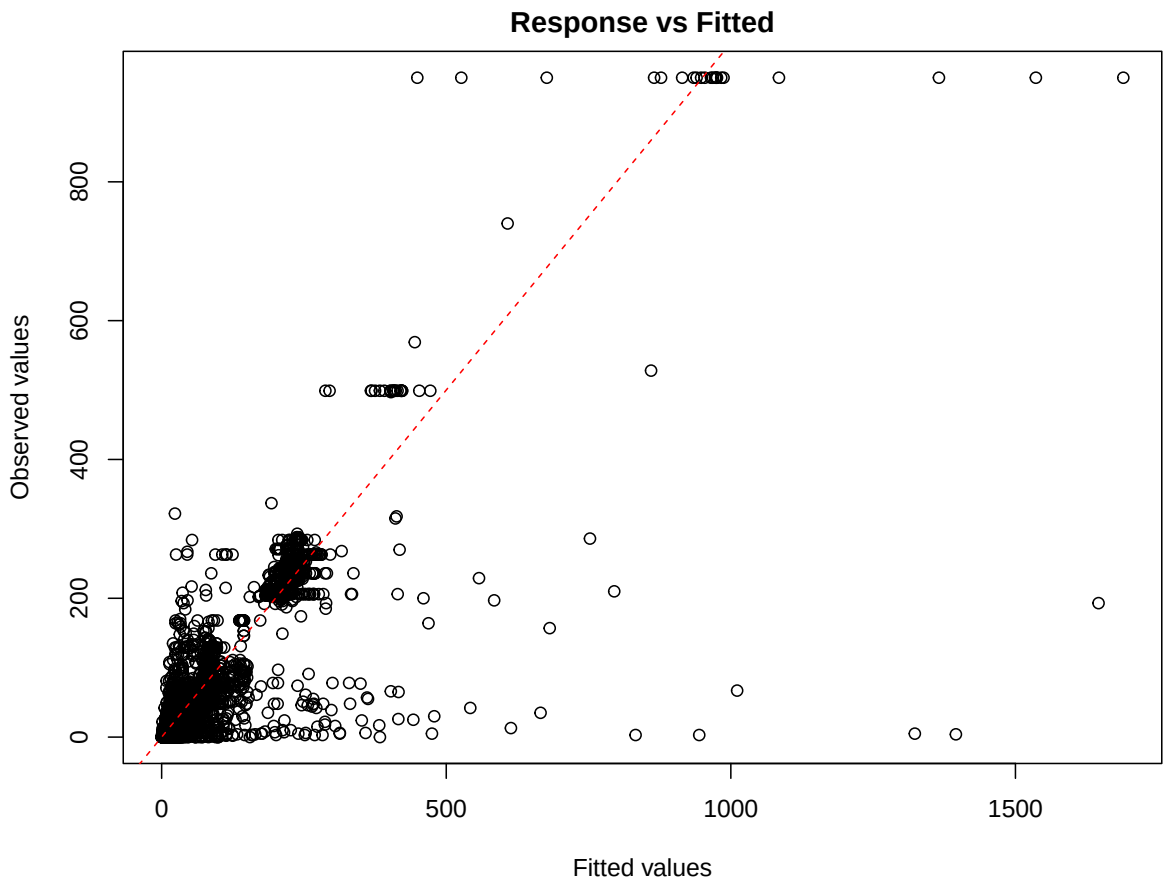

```
### Model 14 : 5c. Wind x Sun + Temp x Sun

Model data dimensions after removing NAs: 5663 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(2.34)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    sunlight_proportion) + ti(mean_temp, sunlight_proportion) +
    s(view_id, bs = "re")

Parametric coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.4977     0.2837   12.33   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                          edf Ref.df   Chi.sq p-value
s(abundance_t_minus_1)                  8.635  8.949 10476.34  <2e-16 ***
s(mean_wind_speed)                      6.603  7.741    85.81  <2e-16 ***
s(mean_temp)                            8.332  8.832   493.56  <2e-16 ***
s(sunlight_proportion)                  7.351  8.321   151.21  <2e-16 ***
ti(mean_wind_speed,sunlight_proportion) 8.605  9.923   144.80  <2e-16 ***
ti(mean_temp,sunlight_proportion)      12.198 13.448   194.77  <2e-16 ***
s(view_id)                              7.793  8.000   369.29  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.606   Deviance explained = 70.7%
-REML =  25205  Scale est. = 1         n = 5663


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
6
4
2
0
−2
−4
0
500
1000
1500

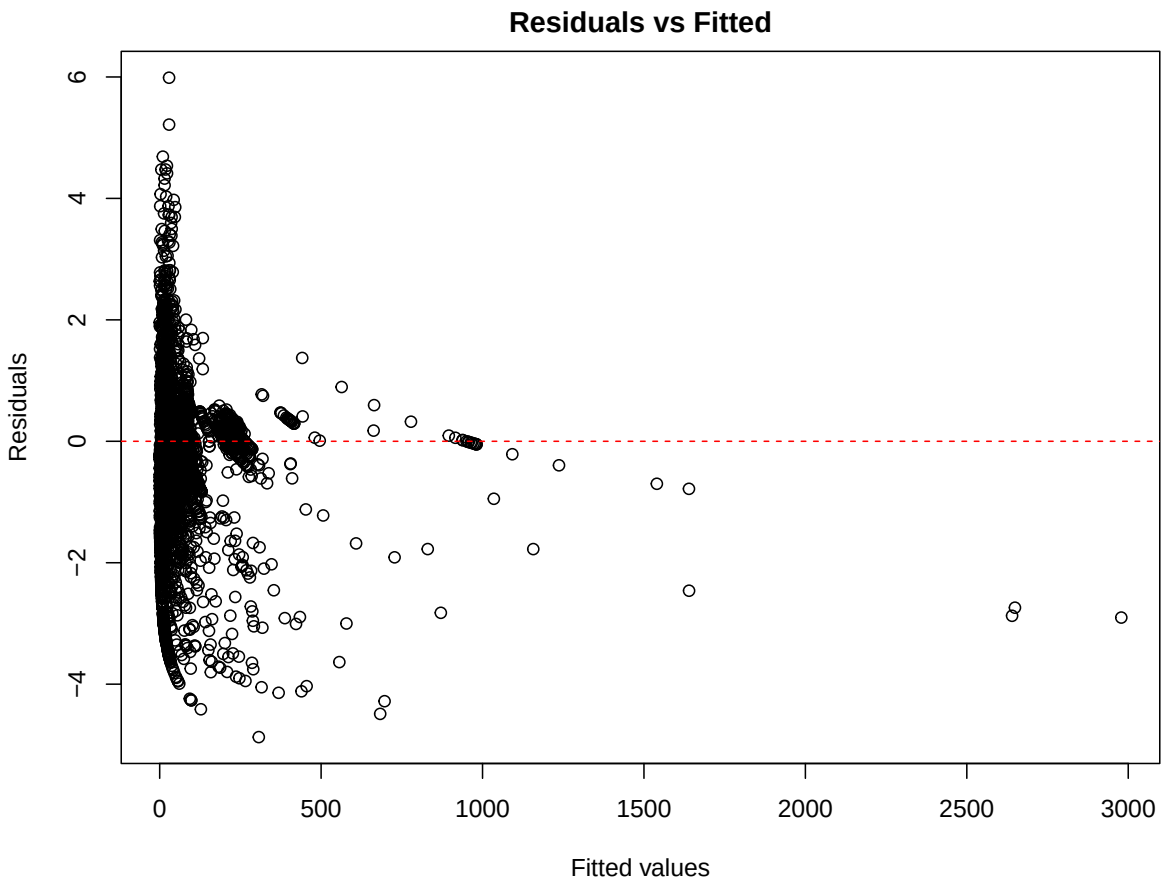

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
6
−4
−2
0
2
4

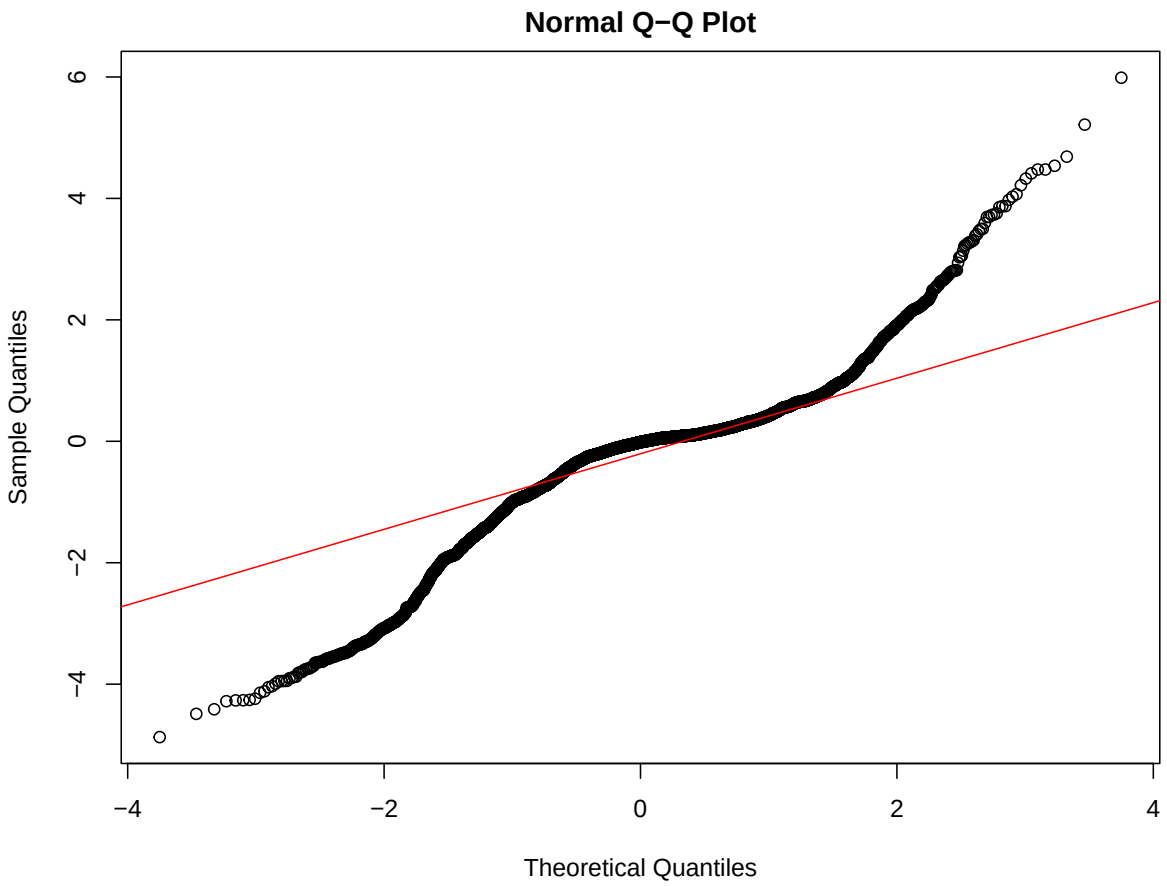

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
Fitted values

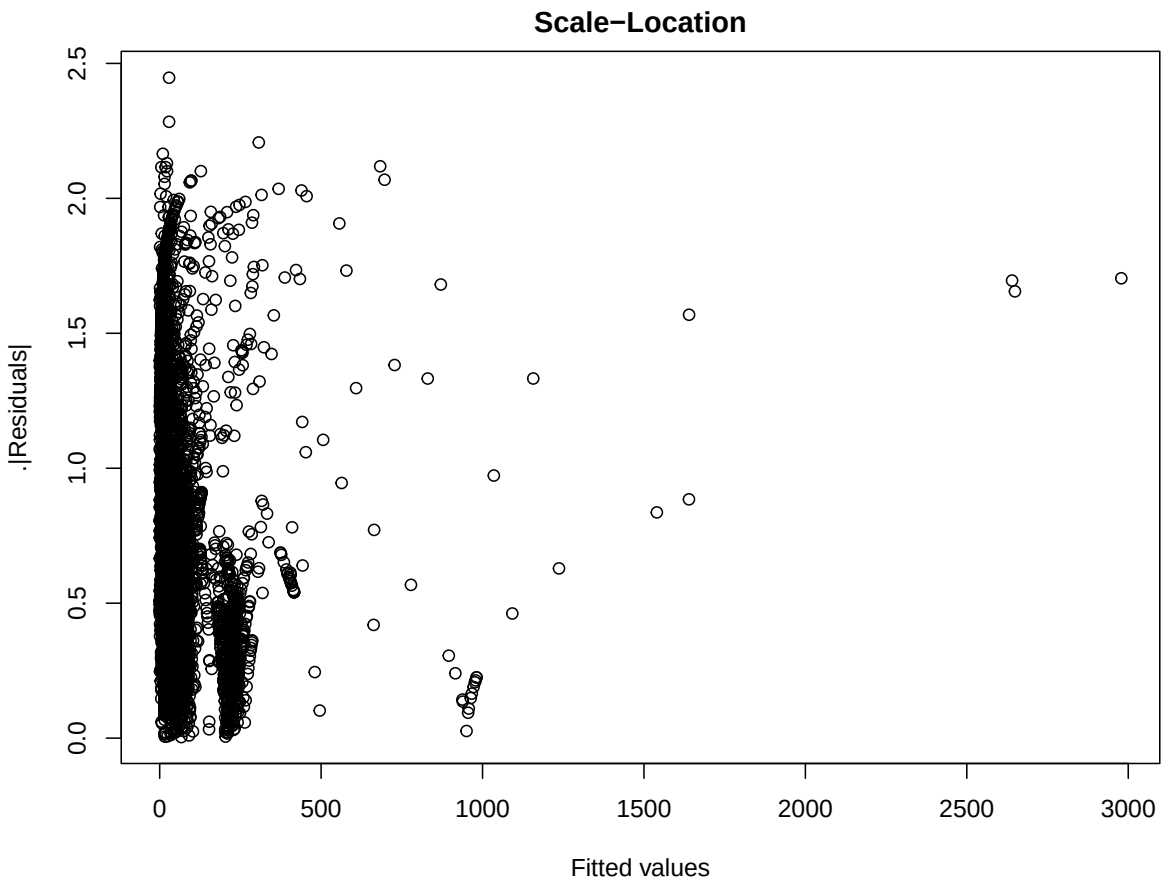


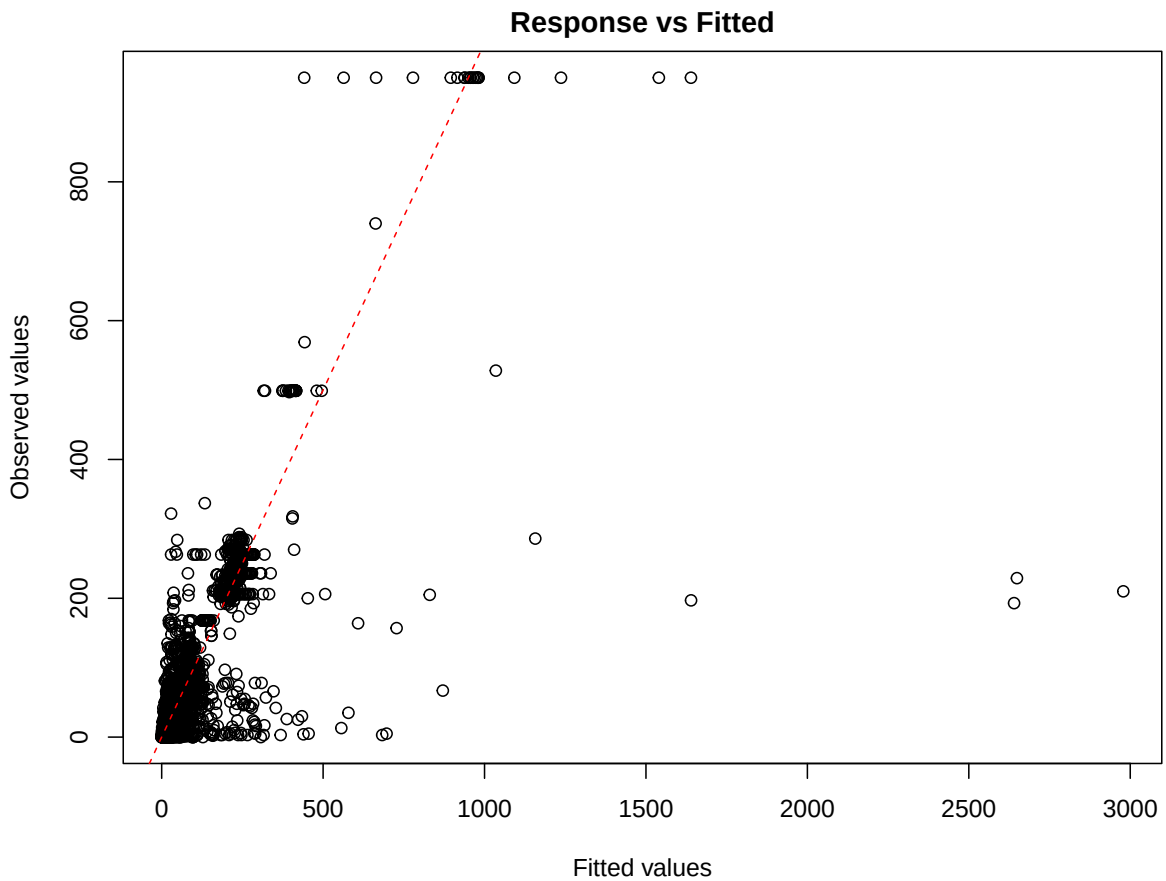

```
### Model 15 : 6. Global model

Model data dimensions after removing NAs: 5663 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(2.453)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + ti(mean_wind_speed, sunlight_proportion) + ti(mean_temp,
    sunlight_proportion) + s(view_id, bs = "re")

Parametric coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3909     0.2838   11.95   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                          edf Ref.df Chi.sq p-value
s(abundance_t_minus_1)                  8.645  8.952 9702.1  <2e-16 ***
s(mean_wind_speed)                      6.583  7.751  104.8  <2e-16 ***
s(mean_temp)                            7.915  8.621  296.4  <2e-16 ***
s(sunlight_proportion)                  6.973  8.031  149.4  <2e-16 ***
ti(mean_wind_speed,mean_temp)          13.859 14.357  232.9  <2e-16 ***
ti(mean_wind_speed,sunlight_proportion) 8.902 10.127  218.2  <2e-16 ***
ti(mean_temp,sunlight_proportion)      12.744 13.945  199.8  <2e-16 ***
s(view_id)                              7.795  8.000  378.8  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =   0.32   Deviance explained =   72%
-REML =  25126  Scale est. = 1         n = 5663


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
6
4
2
0
−2
−4
0
500
1000
1500
2000
2500
3000


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
6
4
2
0
−2
−4
−4
−2
0
2
4


Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000
2500
3000
Fitted values




```r
# Model comparison table
comparison_240m <- data.frame(
  Model = names(aic_values_240m),
  AIC = round(aic_values_240m, 2)
) %>%
  arrange(AIC)

cat("\n### Model Comparison (4-hour lag)\n")
```

```
### Model Comparison (4-hour lag)
```

```r
kable(comparison_240m)
```

	Model	AIC
6. Global model	6. Global model	50016.03
5a. Wind x Temp + Wind x Sun	5a. Wind x Temp + Wind x Sun	50199.82
5c. Wind x Sun + Temp x Sun	5c. Wind x Sun + Temp x Sun	50235.69
5b. Wind x Temp + Temp x Sun	5b. Wind x Temp + Temp x Sun	50329.92
4c. Temp x Sunlight interaction	4c. Temp x Sunlight interaction	50353.42
4b. Wind x Sunlight interaction	4b. Wind x Sunlight interaction	50425.97
4a. Wind x Temp interaction	4a. Wind x Temp interaction	50574.97
3d. All additive	3d. All additive	50586.55
3c. Temp + Sunlight	3c. Temp + Sunlight	50656.47
3a. Wind + Temp	3a. Wind + Temp	51160.97
2b. Temperature only	2b. Temperature only	51212.90
3b. Wind + Sunlight	3b. Wind + Sunlight	51273.00
2c. Sunlight only	2c. Sunlight only	51377.60
2a. Wind only	2a. Wind only	51779.13
1. Baseline	1. Baseline	51872.69

Part 5: Full Day Lag Analysis

```
cat("Preparing full day lag data...\n")
```

```
Preparing full day lag data...
```

```
# Group by view_id and date
full_day_data <- master_df %>%
  filter(!is.na(datetime) & !is.na(view_id)) %>%
  mutate(date = as.Date(datetime)) %>%
  group_by(view_id, date) %>%
  arrange(datetime) %>%
  mutate(
    # Find first observation with butterflies > 0
    has_butterflies = total_butterflies > 0,
    is_first_with_butterflies = has_butterflies & !lag(has_butterflies, default = FALSE)
  ) %>%
  filter(any(is_first_with_butterflies)) %>%  # Keep only days with at least one butterfly ob
  summarise(
    # t-1 observation (first with butterflies)
```

```r
    datetime_t_minus_1 = datetime[which(is_first_with_butterflies)[1]],
    abundance_t_minus_1 = total_butterflies[which(is_first_with_butterflies)[1]],

    # t observation (last of day)
    datetime = last(datetime),
    total_butterflies = last(total_butterflies),

    # Keep wind meter name for joining
    wind_meter_name = first(wind_meter_name),

    .groups = "drop"
  ) %>%
  # Calculate time delta and filter
  mutate(
    time_delta_hours = as.numeric(difftime(datetime, datetime_t_minus_1, units = "hours"))
  ) %>%
  filter(time_delta_hours > 4)

cat("Full day pairs found:", nrow(full_day_data), "\n")
```

```
Full day pairs found: 125
```

```r
cat("Time delta range:", round(min(full_day_data$time_delta_hours), 1),
    "to", round(max(full_day_data$time_delta_hours), 1), "hours\n")
```

```
Time delta range: 7 to 23.9 hours
```

```r
# Calculate environmental predictors for full day intervals
full_day_with_predictors <- full_day_data %>%
  rowwise() %>%
  mutate(
    # Wind metrics
    wind_data = list({
      wind_subset <- wind %>%
        filter(
          wind_meter_name == .data$wind_meter_name,
          time >= datetime_t_minus_1,
          time <= datetime
        )
      if (nrow(wind_subset) > 0) {
```

```
      list(
        mean_wind_speed = mean(wind_subset$speed, na.rm = TRUE),
        max_wind_speed = max(wind_subset$speed, na.rm = TRUE),
        sd_wind_speed = sd(wind_subset$speed, na.rm = TRUE),
        gust_factor = mean(wind_subset$gust, na.rm = TRUE) / mean(wind_subset$speed, na.rm
        minutes_above_2mps = sum(wind_subset$speed > 2, na.rm = TRUE)
      )
    } else {
      list(
        mean_wind_speed = NA_real_,
        max_wind_speed = NA_real_,
        sd_wind_speed = NA_real_,
        gust_factor = NA_real_,
        minutes_above_2mps = NA_real_
      )
    }
  }),
  mean_wind_speed = wind_data$mean_wind_speed,
  max_wind_speed = wind_data$max_wind_speed,
  sd_wind_speed = wind_data$sd_wind_speed,
  gust_factor = wind_data$gust_factor,
  minutes_above_2mps = wind_data$minutes_above_2mps,

  # Temperature metrics
  temp_data = list({
    temp_subset <- master_df %>%
      filter(
        view_id == .data$view_id,
        datetime >= datetime_t_minus_1,
        datetime <= .data$datetime
      )
    if (nrow(temp_subset) > 0) {
      list(
        mean_temp = mean(temp_subset$temperature, na.rm = TRUE),
        max_temp = max(temp_subset$temperature, na.rm = TRUE),
        min_temp = min(temp_subset$temperature, na.rm = TRUE),
        sd_temp = sd(temp_subset$temperature, na.rm = TRUE)
      )
    } else {
      list(
        mean_temp = NA_real_,
        max_temp = NA_real_,
```

```
        min_temp = NA_real_,
        sd_temp = NA_real_
      )
    }
  }),
  mean_temp = temp_data$mean_temp,
  max_temp = temp_data$max_temp,
  min_temp = temp_data$min_temp,
  sd_temp = temp_data$sd_temp,

  # Sunlight metrics
  sun_data = list({
    sun_subset <- master_df %>%
      filter(
        view_id == .data$view_id,
        datetime >= datetime_t_minus_1,
        datetime <= .data$datetime
      )
    if (nrow(sun_subset) > 0) {
      total_sun <- sum(sun_subset$butterflies_direct_sun, na.rm = TRUE)
      total_count <- sum(sun_subset$total_butterflies, na.rm = TRUE)
      if (total_count > 0) {
        overall_prop <- total_sun / total_count
        # Calculate per-observation proportions
        obs_props <- ifelse(sun_subset$total_butterflies > 0,
                            sun_subset$butterflies_direct_sun / sun_subset$total_butterflies
                            NA_real_)
        list(
          sunlight_proportion = overall_prop,
          max_prop = max(obs_props, na.rm = TRUE),
          min_prop = min(obs_props, na.rm = TRUE),
          sd_prop = sd(obs_props, na.rm = TRUE)
        )
      } else {
        list(
          sunlight_proportion = NA_real_,
          max_prop = NA_real_,
          min_prop = NA_real_,
          sd_prop = NA_real_
        )
      }
    } else {
```

```r
        list(
          sunlight_proportion = NA_real_,
          max_prop = NA_real_,
          min_prop = NA_real_,
          sd_prop = NA_real_
        )
      }
    }),
    sunlight_proportion = sun_data$sunlight_proportion,
    max_prop_sunlight = sun_data$max_prop,
    min_prop_sunlight = sun_data$min_prop,
    sd_prop_sunlight = sun_data$sd_prop
  ) %>%
  ungroup() %>%
  select(-wind_data, -temp_data, -sun_data)

# Filter for complete cases
data_full_day <- full_day_with_predictors %>%
  filter(!is.na(mean_wind_speed) & !is.na(mean_temp))

cat("Final full day dataset rows:", nrow(data_full_day), "\n")
```

```
Final full day dataset rows: 125
```

Exploratory Data Analysis (Full Day)

```r
cat("Data structure:\n")
```

```
Data structure:
```

```r
glimpse(data_full_day)
```

```
Rows: 125
Columns: 21
$ view_id             <fct> 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
$ date                <date> 2023-11-17, 2023-11-18, 2023-11-19, 2023-11-20, 2~
$ datetime_t_minus_1  <dttm> 2023-11-17 11:40:01, 2023-11-18 00:00:01, 2023-11~
$ abundance_t_minus_1 <dbl> 121, 54, 19, 2, 436, 499, 206, 1, 2, 168, 236, 26,~
```

```
$ datetime            <dttm> 2023-11-17 23:55:01, 2023-11-18 23:55:01, 2023-11~
$ total_butterflies   <dbl> 54, 19, 2, 0, 499, 206, 0, 2, 168, 236, 26, 15, 37~
$ wind_meter_name     <chr> "JazzPlay", "JazzPlay", "JazzPlay", "JazzPlay", "S~
$ time_delta_hours    <dbl> 12.25000, 23.91667, 23.83333, 23.83333, 9.00000, 2~
$ mean_wind_speed     <dbl> 2.2716738, 1.4147838, 0.1985714, 1.1108467, 0.1808~
$ max_wind_speed      <dbl> 4.2, 5.0, 2.0, 4.3, 1.8, 2.3, 2.8, 3.4, 3.3, 3.0, ~
$ sd_wind_speed       <dbl> 0.7711548, 0.8540580, 0.3589021, 0.7926403, 0.3206~
$ gust_factor         <dbl> 1.401285, 1.440162, 2.217672, 1.541010, 2.034462, ~
$ minutes_above_2mps  <int> 280, 683, 0, 137, 0, 5, 25, 85, 156, 173, 195, 138~
$ mean_temp           <dbl> 13.25099, 13.18943, 13.02735, 12.97943, 13.02589, ~
$ max_temp            <dbl> 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37~
$ min_temp            <dbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,~
$ sd_temp             <dbl> 3.951748, 3.971907, 4.000563, 4.047503, 4.219018, ~
$ sunlight_proportion <dbl> 0.02921473, 0.03045164, 0.03255667, 0.03318173, 0.~
$ max_prop_sunlight   <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ min_prop_sunlight   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ sd_prop_sunlight    <dbl> 0.2328945, 0.2368612, 0.2461221, 0.2520011, 0.2523~
```

```r
cat("\n\nData summary:\n")
```

```
Data summary:
```

```r
summary(data_full_day)
```

```
    view_id        date             datetime_t_minus_1
 2      :55   Min.   :2023-11-17   Min.   :2023-11-17 11:40:01.00
 4      :52   1st Qu.:2023-12-22   1st Qu.:2023-12-22 00:20:01.00
 1      : 4   Median :2024-01-05   Median :2024-01-05 00:00:01.00
 9      : 4   Mean   :2024-01-03   Mean   :2024-01-03 06:08:56.68
 5      : 3   3rd Qu.:2024-01-19   3rd Qu.:2024-01-19 00:20:01.00
 11     : 3   Max.   :2024-02-04   Max.   :2024-02-04 00:00:01.00
 (Other): 4
 abundance_t_minus_1    datetime                      total_butterflies
 Min.   :  1.00      Min.   :2023-11-17 23:55:01.0   Min.   :  0.0
 1st Qu.:  6.00      1st Qu.:2023-12-22 23:50:01.0   1st Qu.:  3.0
 Median : 37.00      Median :2024-01-05 23:36:04.0   Median : 26.0
 Mean   : 78.16      Mean   :2024-01-04 03:15:34.0   Mean   : 77.6
 3rd Qu.:109.00      3rd Qu.:2024-01-19 23:50:01.0   3rd Qu.: 91.0
 Max.   :499.00      Max.   :2024-02-04 10:50:01.0   Max.   :950.0
```

```
 wind_meter_name    time_delta_hours mean_wind_speed  max_wind_speed
 Length:125         Min.   : 7.00    Min.   :0.1012   Min.   :1.600
 Class :character   1st Qu.:23.50    1st Qu.:0.3617   1st Qu.:2.900
 Mode  :character   Median :23.50    Median :0.6096   Median :3.400
                    Mean   :21.11    Mean   :0.7048   Mean   :3.836
                    3rd Qu.:23.50    3rd Qu.:0.8764   3rd Qu.:4.300
                    Max.   :23.92    Max.   :2.3941   Max.   :9.700

 sd_wind_speed     gust_factor    minutes_above_2mps   mean_temp
 Min.   :0.2398   Min.   :1.346   Min.   :   0.0     Min.   : 9.686
 1st Qu.:0.5133   1st Qu.:1.476   1st Qu.:  35.0     1st Qu.:12.820
 Median :0.6241   Median :1.518   Median : 177.0     Median :13.019
 Mean   :0.6488   Mean   :1.542   Mean   : 406.6     Mean   :13.124
 3rd Qu.:0.7849   3rd Qu.:1.594   3rd Qu.: 558.0     3rd Qu.:13.658
 Max.   :1.4544   Max.   :2.218   Max.   :3876.0     Max.   :14.397

    max_temp        min_temp        sd_temp       sunlight_proportion
 Min.   :11.00   Min.   :3.000   Min.   :0.5333   Min.   :0.002424
 1st Qu.:37.00   1st Qu.:3.000   1st Qu.:4.3022   1st Qu.:0.033951
 Median :37.00   Median :3.000   Median :4.5068   Median :0.034954
 Mean   :35.66   Mean   :3.768   Mean   :4.4176   Mean   :0.055235
 3rd Qu.:37.00   3rd Qu.:5.000   3rd Qu.:4.6698   3rd Qu.:0.074669
 Max.   :37.00   Max.   :9.000   Max.   :5.4154   Max.   :0.154394

 max_prop_sunlight min_prop_sunlight sd_prop_sunlight
 Min.   :0.2000    Min.   :0         Min.   :0.03306
 1st Qu.:1.0000    1st Qu.:0         1st Qu.:0.24936
 Median :1.0000    Median :0         Median :0.25270
 Mean   :0.9931    Mean   :0         Mean   :0.25218
 3rd Qu.:1.0000    3rd Qu.:0         3rd Qu.:0.25648
 Max.   :1.0000    Max.   :0         Max.   :0.29015
```

```r
# Check for data quality issues
cat("\n\nData quality checks:\n")
```

```
Data quality checks:
```

```r
cat("Total rows:", nrow(data_full_day), "\n")
```

```
Total rows: 125
```

```r
cat("Complete cases:", sum(complete.cases(data_full_day)), "\n")
```

```
Complete cases: 125
```

```r
cat("Rows with abundance_t_minus_1 > 0:", sum(data_full_day$abundance_t_minus_1 > 0, na.rm =
```

```
Rows with abundance_t_minus_1 > 0: 125
```

```r
cat("Rows with total_butterflies > 0:", sum(data_full_day$total_butterflies > 0, na.rm = TRU
```

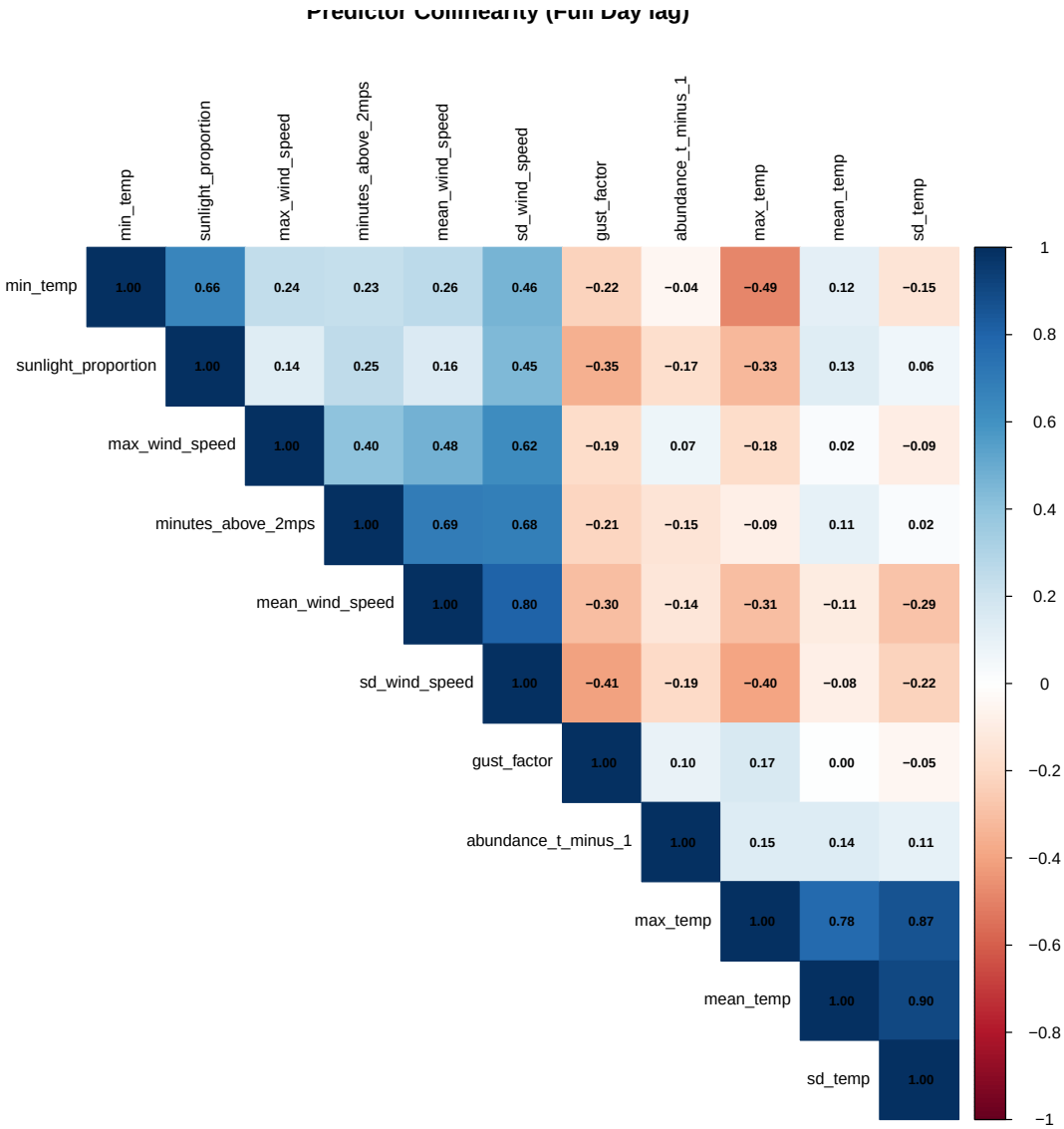
```
Rows with total_butterflies > 0: 108
```

```r
cat("Unique view_ids:", length(unique(data_full_day$view_id)), "\n")
```

```
Unique view_ids: 9
```

```r
# Select numeric predictors for correlation analysis
predictors_full_day <- data_full_day %>%
  select(abundance_t_minus_1, mean_wind_speed, max_wind_speed, sd_wind_speed,
         gust_factor, minutes_above_2mps, mean_temp, max_temp, min_temp, sd_temp,
         sunlight_proportion) %>%
  na.omit()

if (nrow(predictors_full_day) > 1) {
  cor_matrix_full_day <- cor(predictors_full_day, use = "complete.obs")
  if (!any(is.na(cor_matrix_full_day) | is.infinite(cor_matrix_full_day))) {
    corrplot(cor_matrix_full_day, method = "color", type = "upper",
             order = "hclust", tl.cex = 0.8, tl.col = "black",
             addCoef.col = "black", number.cex = 0.7,
             title = "Predictor Collinearity (Full Day lag)")
  } else {
    corrplot(cor_matrix_full_day, method = "color", type = "upper",
             order = "original", tl.cex = 0.8, tl.col = "black",
             addCoef.col = "black", number.cex = 0.7,
```

```
            title = "Predictor Collinearity (Full Day lag)")
  }
} else {
  cat("Not enough data for correlation analysis\n")
}
```



Model Fitting (Full Day)

```
# Define comprehensive model formulas
model_formulas_full_day <- list(
  # 1. Baseline Model
  "1. Baseline" = total_butterflies ~ s(abundance_t_minus_1) + s(view_id, bs = "re"),

  # 2. Single Environmental Factor Models
  "2a. Wind only" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(view_
  "2b. Temperature only" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(vie
  "2c. Sunlight only" = total_butterflies ~ s(abundance_t_minus_1) + s(sunlight_proportion) +

  # 3. Additive Models (No Interactions)
  "3a. Wind + Temp" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(mea
  "3b. Wind + Sunlight" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s
  "3c. Temp + Sunlight" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(sunli
  "3d. All additive" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(me

  # 4. Single Interaction Models
  "4a. Wind x Temp interaction" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_sp
  "4b. Wind x Sunlight interaction" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wi
  "4c. Temp x Sunlight interaction" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wi

  # 5. Multiple Interaction Models
  "5a. Wind x Temp + Wind x Sun" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_s
  "5b. Wind x Temp + Temp x Sun" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_s
  "5c. Wind x Sun + Temp x Sun" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_sp

  # 6. Global Model with All Interactions
  "6. Global model" = total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) + s(mea
)

# Storage for results
models_full_day <- list()
aic_values_full_day <- numeric()

# Fit each model
for (i in seq_along(model_formulas_full_day)) {
  model_name <- names(model_formulas_full_day)[i]
  formula <- model_formulas_full_day[[i]]

  cat("\n### Model", i, ":", model_name, "\n\n")
```

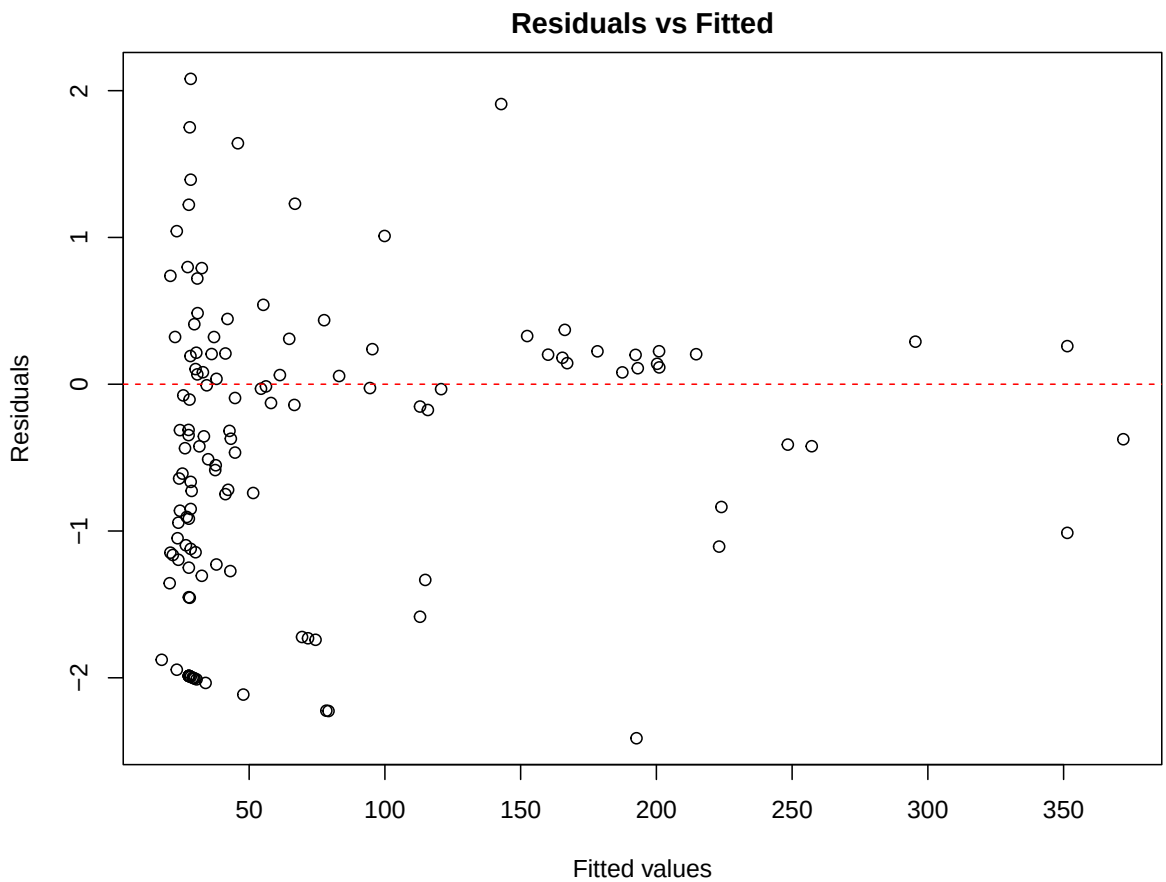
```r
tryCatch({
  # Check data before fitting
  model_vars <- all.vars(formula)
  model_data <- data_full_day %>%
    select(all_of(model_vars), view_id) %>%
    na.omit()

  cat("Model data dimensions after removing NAs:", nrow(model_data), "x", ncol(model_data)

  # Fit GAM
  gam_model <- gam(
    formula,
    family = nb(),
    data = model_data
  )

  cat("GAM fitted successfully\n")

  # Store model
  models_full_day[[model_name]] <- gam_model
  aic_values_full_day[model_name] <- AIC(gam_model)

  # Print summary
  cat("Model Summary:\n")
  print(summary(gam_model))

  # Diagnostic plots - individual plots for better visibility in PDF
  cat("\n\nDiagnostic plots:\n")

  # Set up for individual plots
  par(mfrow = c(1, 1), mar = c(4, 4, 2, 1))

  # Plot 1: Residuals vs Fitted
  plot(fitted(gam_model), residuals(gam_model),
       xlab = "Fitted values", ylab = "Residuals",
       main = "Residuals vs Fitted")
  abline(h = 0, col = "red", lty = 2)

  # Plot 2: Q-Q plot
  qqnorm(residuals(gam_model))
  qqline(residuals(gam_model), col = "red")
```

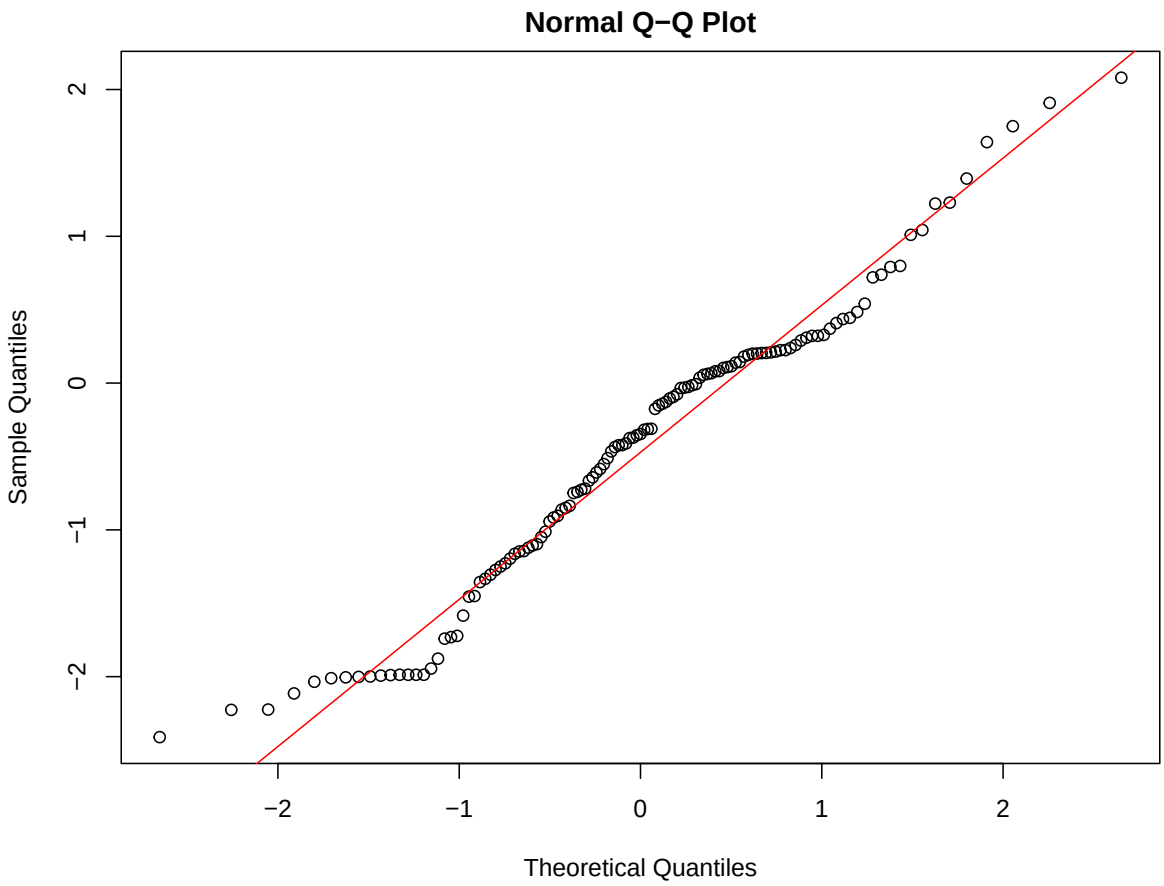
```r
    # Plot 3: Scale-Location (if appropriate)
    plot(fitted(gam_model), sqrt(abs(residuals(gam_model))),
         xlab = "Fitted values", ylab = "√|Residuals|",
         main = "Scale-Location")

    # Plot 4: Response vs Fitted
    plot(fitted(gam_model), model_data$total_butterflies,
         xlab = "Fitted values", ylab = "Observed values",
         main = "Response vs Fitted")
    abline(a = 0, b = 1, col = "red", lty = 2)

  }, error = function(e) {
    cat("Error fitting model:", e$message, "\n")
    aic_values_full_day[model_name] <- NA
  })
}
```

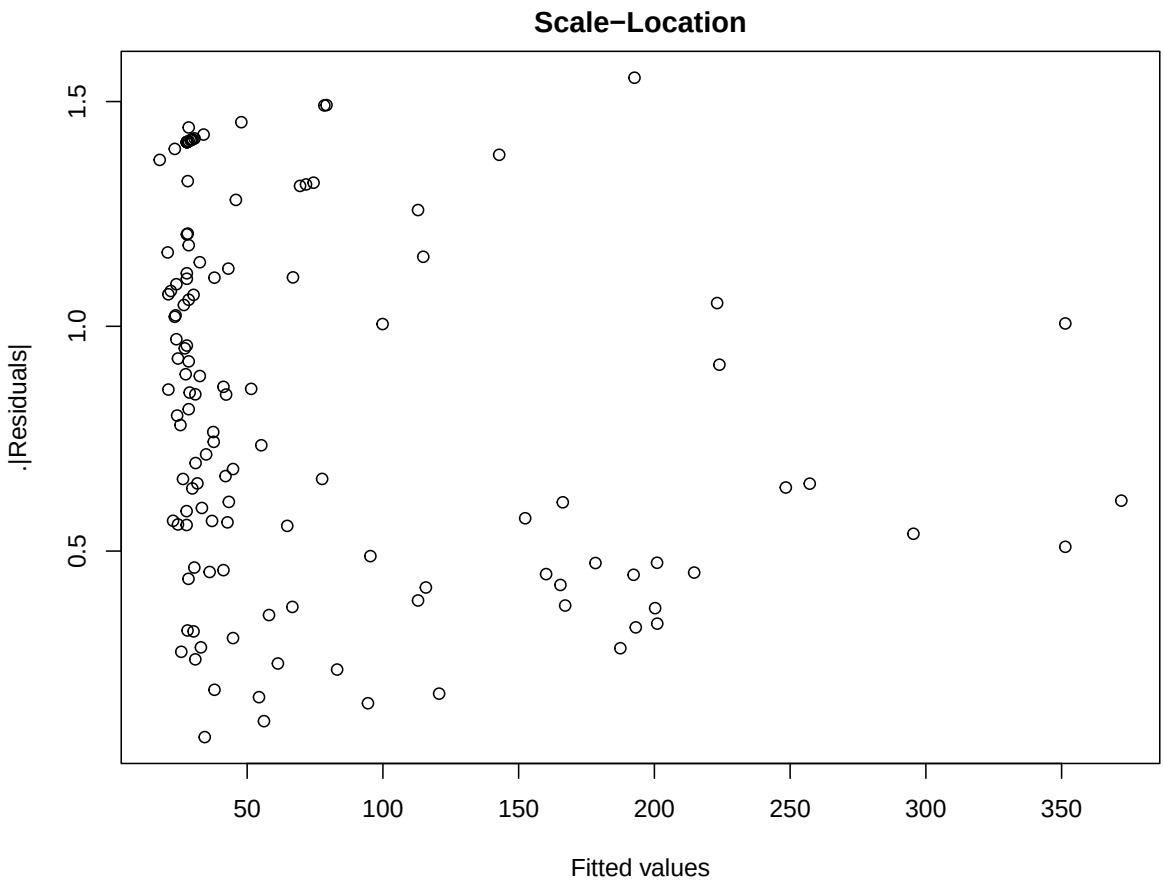
```
### Model 1 : 1. Baseline

Model data dimensions after removing NAs: 125 x 3
GAM fitted successfully
Model Summary:

Family: Negative Binomial(0.486)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.9108     0.2798   13.98   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                         edf Ref.df Chi.sq  p-value
s(abundance_t_minus_1) 2.092  2.581 37.425 3.04e-07 ***
s(view_id)             2.788  8.000  3.661    0.231
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

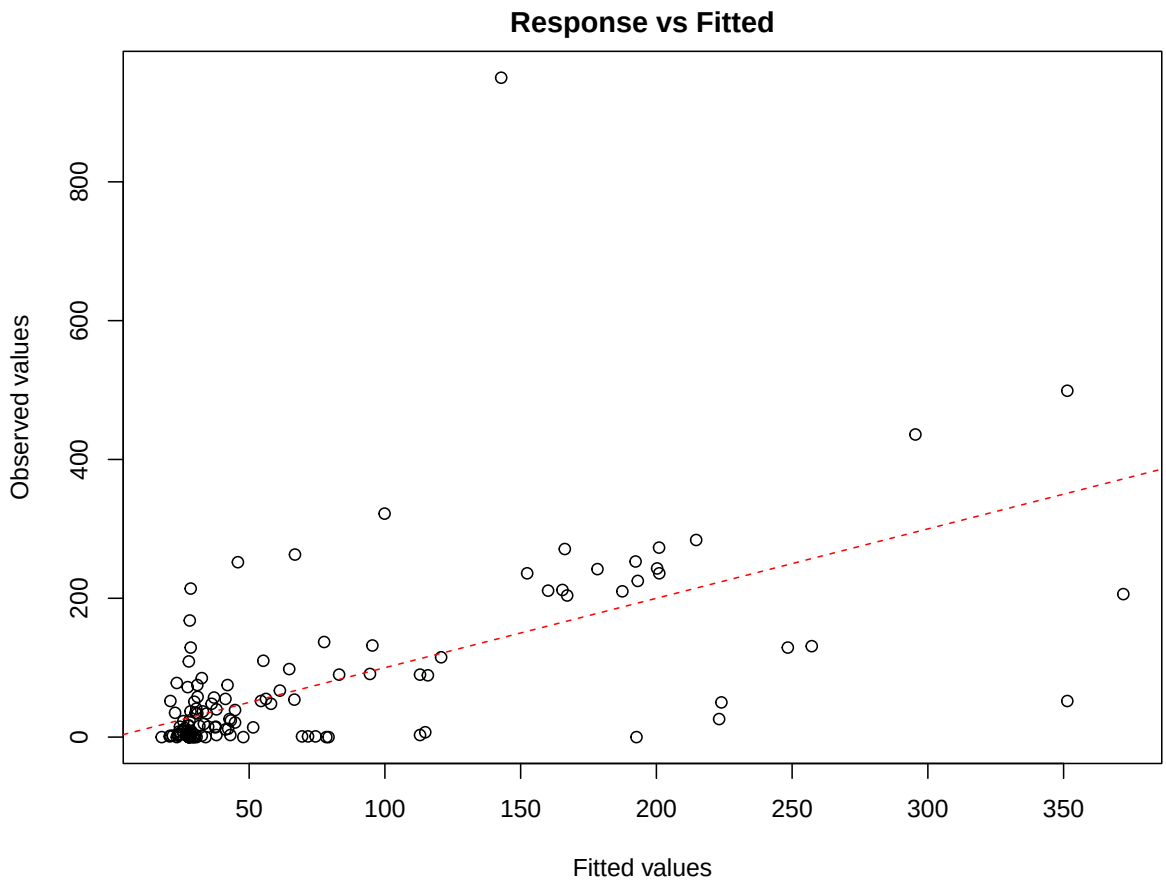
```
R-sq.(adj) =  0.302   Deviance explained = 25.4%
-REML = 608.78  Scale est. = 1         n = 125
```

Diagnostic plots:



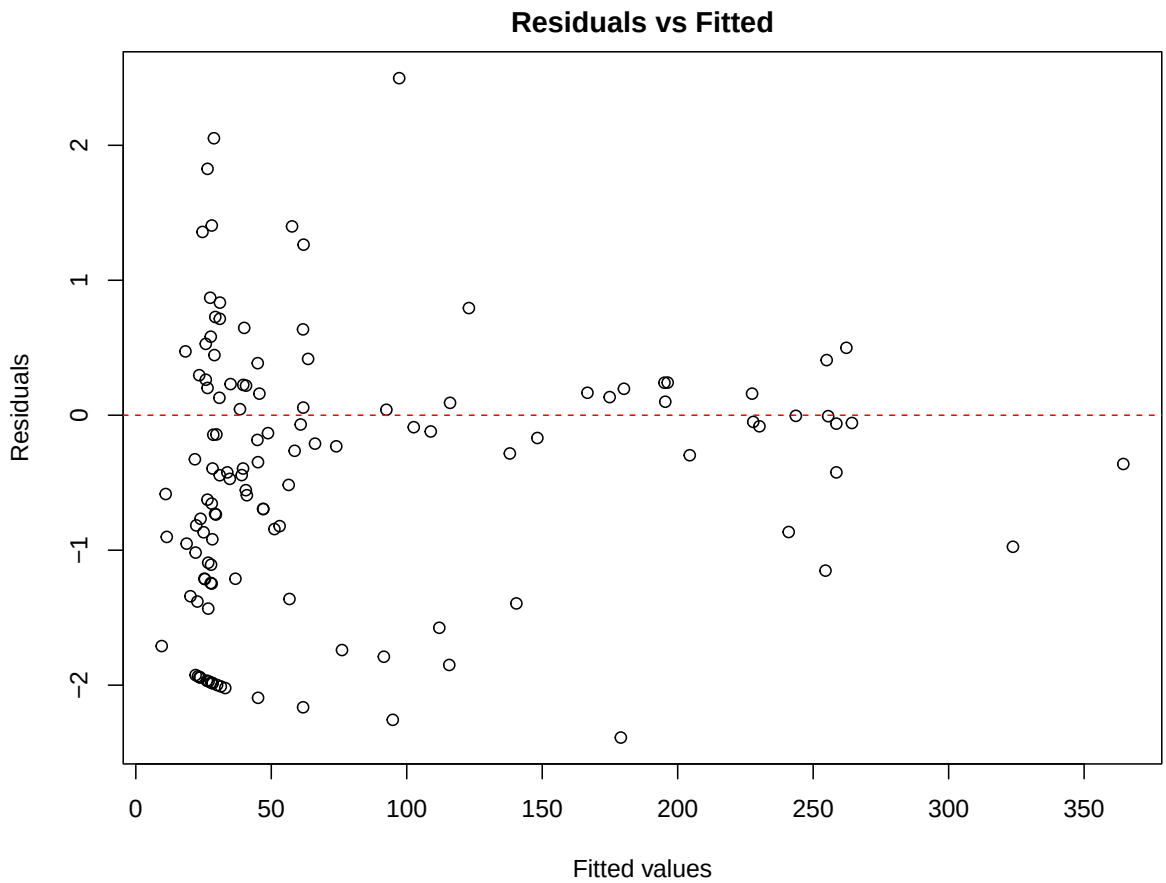

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−2
−1
0
1
2

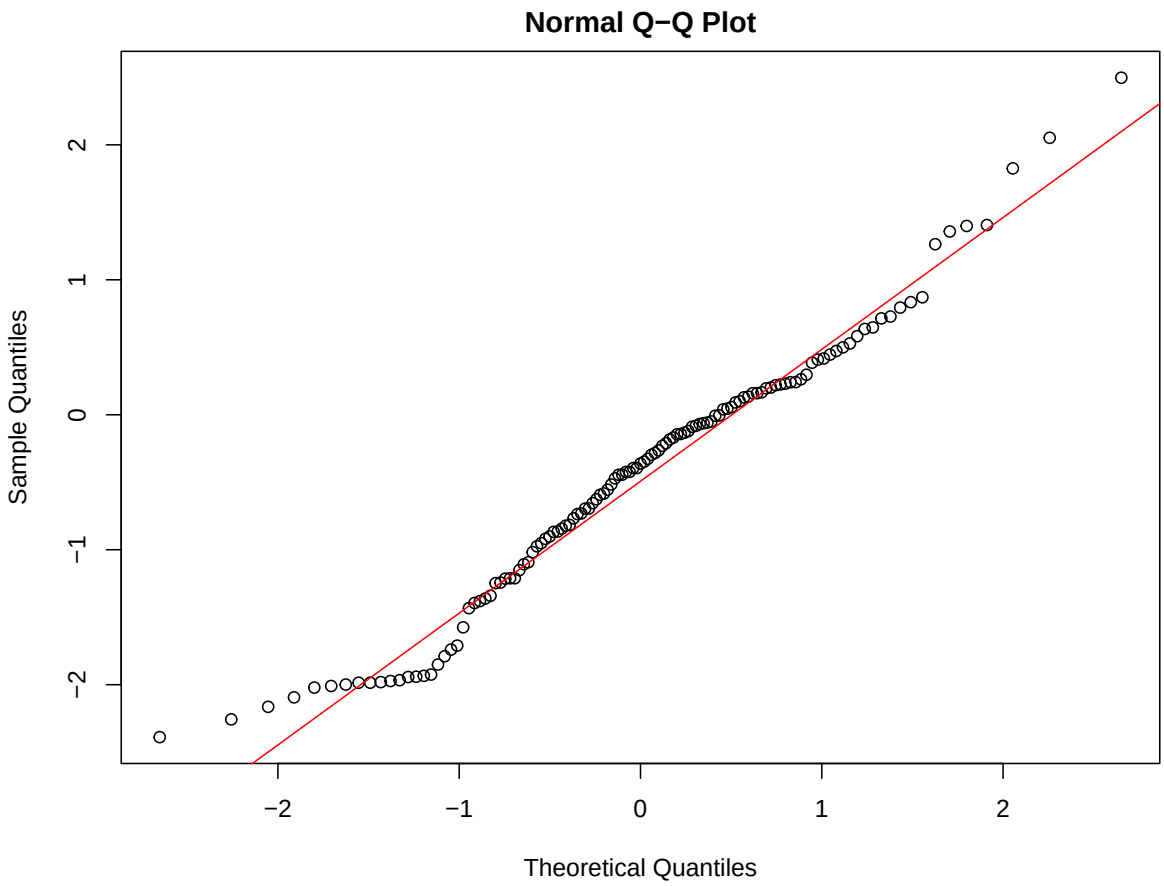

Scale−Location
√|Residuals|
0.5
1.0
1.5
50
100
150
200
250
300
350
Fitted values

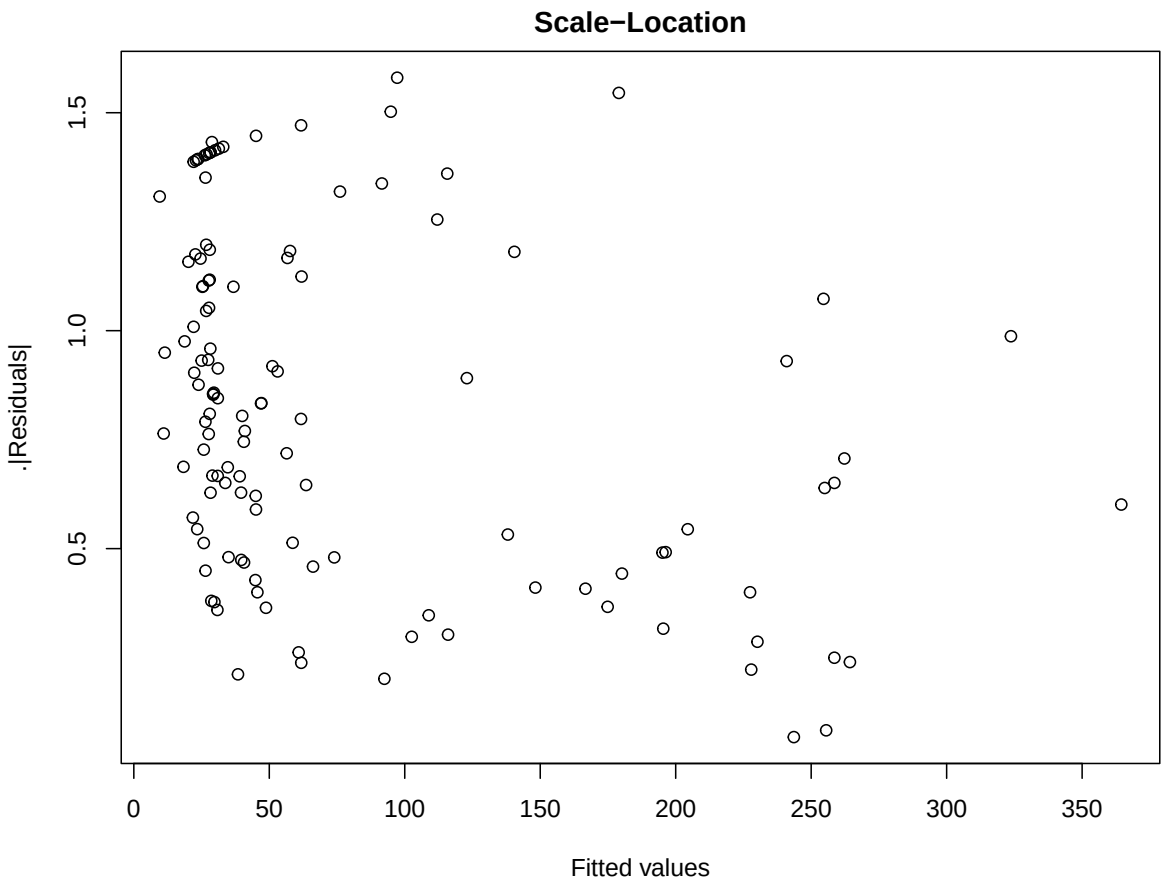


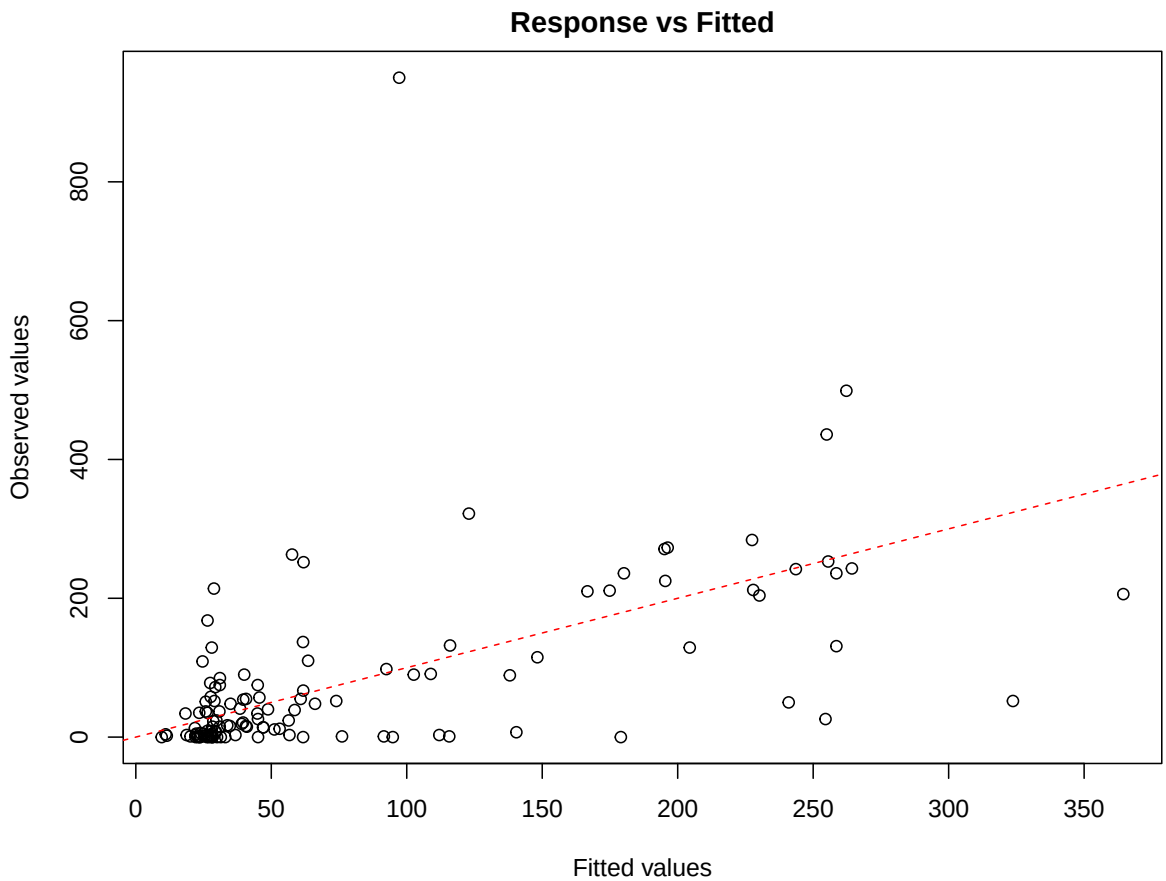

```
### Model 2 : 2a. Wind only

Model data dimensions after removing NAs: 125 x 4
GAM fitted successfully
Model Summary:

Family: Negative Binomial(0.482)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.9545     0.1297   30.48   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                            edf Ref.df Chi.sq p-value
s(abundance_t_minus_1) 2.331063  2.872 41.543  <2e-16 ***
s(mean_wind_speed)     2.248172  2.786  5.219   0.215
s(view_id)             0.001833  8.000  0.002   0.407
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.249   Deviance explained = 24.7%
-REML = 609.07  Scale est. = 1         n = 125


Diagnostic plots:
```



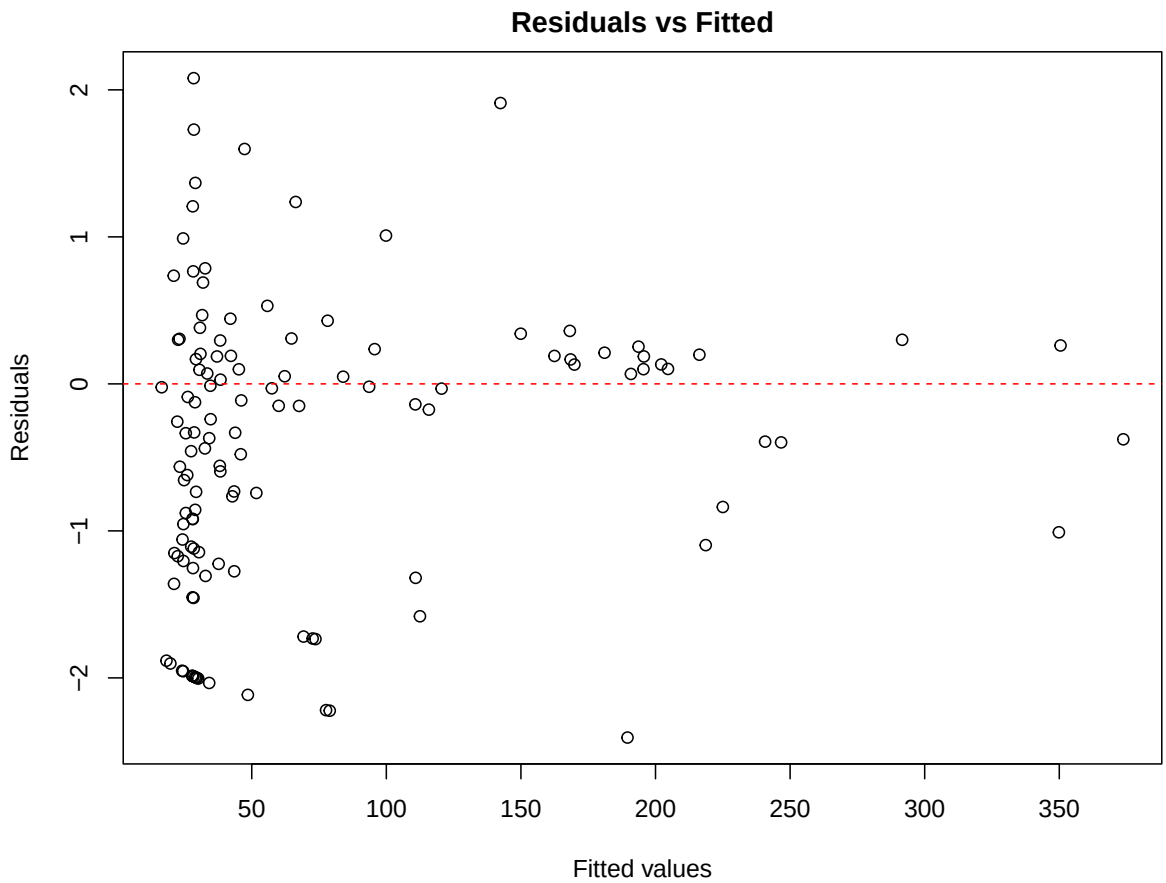

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−2
−1
0
1
2

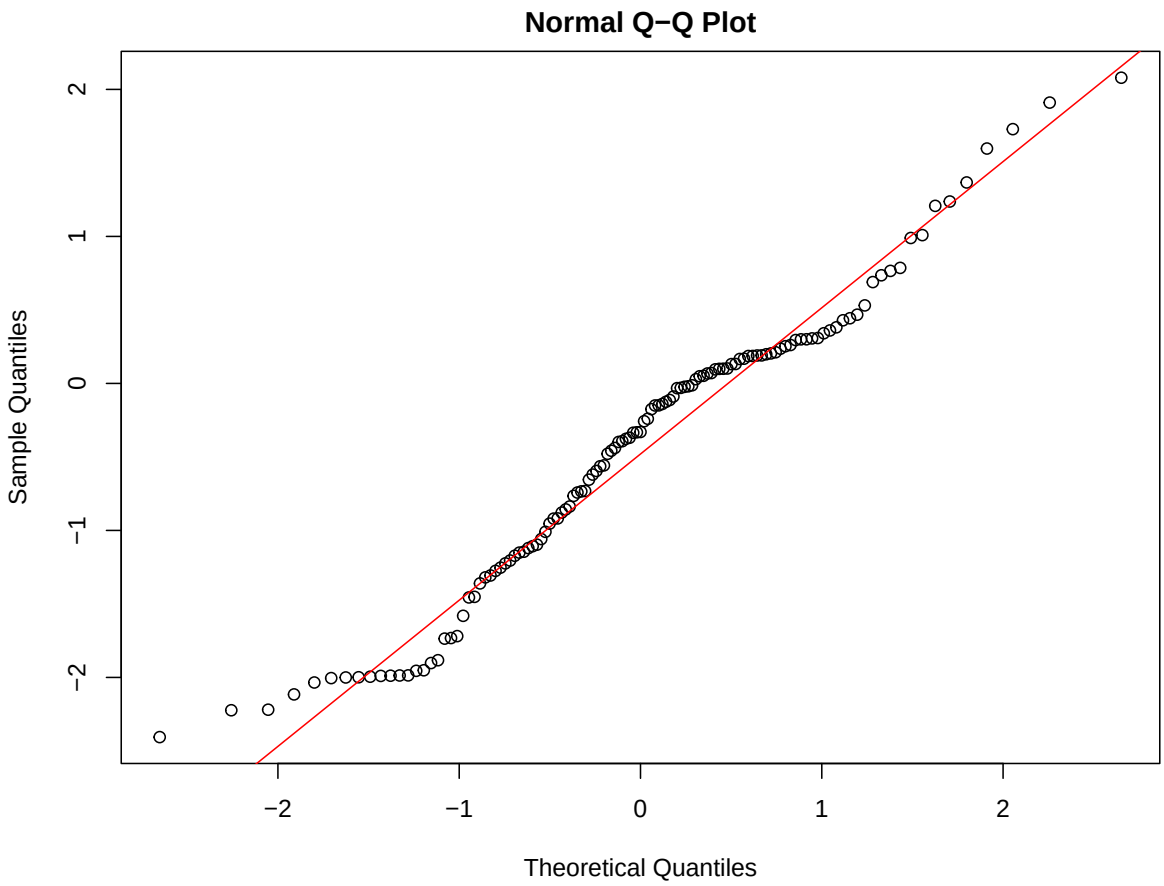

Scale−Location
√|Residuals|
0.5
1.0
1.5
0
50
100
150
200
250
300
350
Fitted values

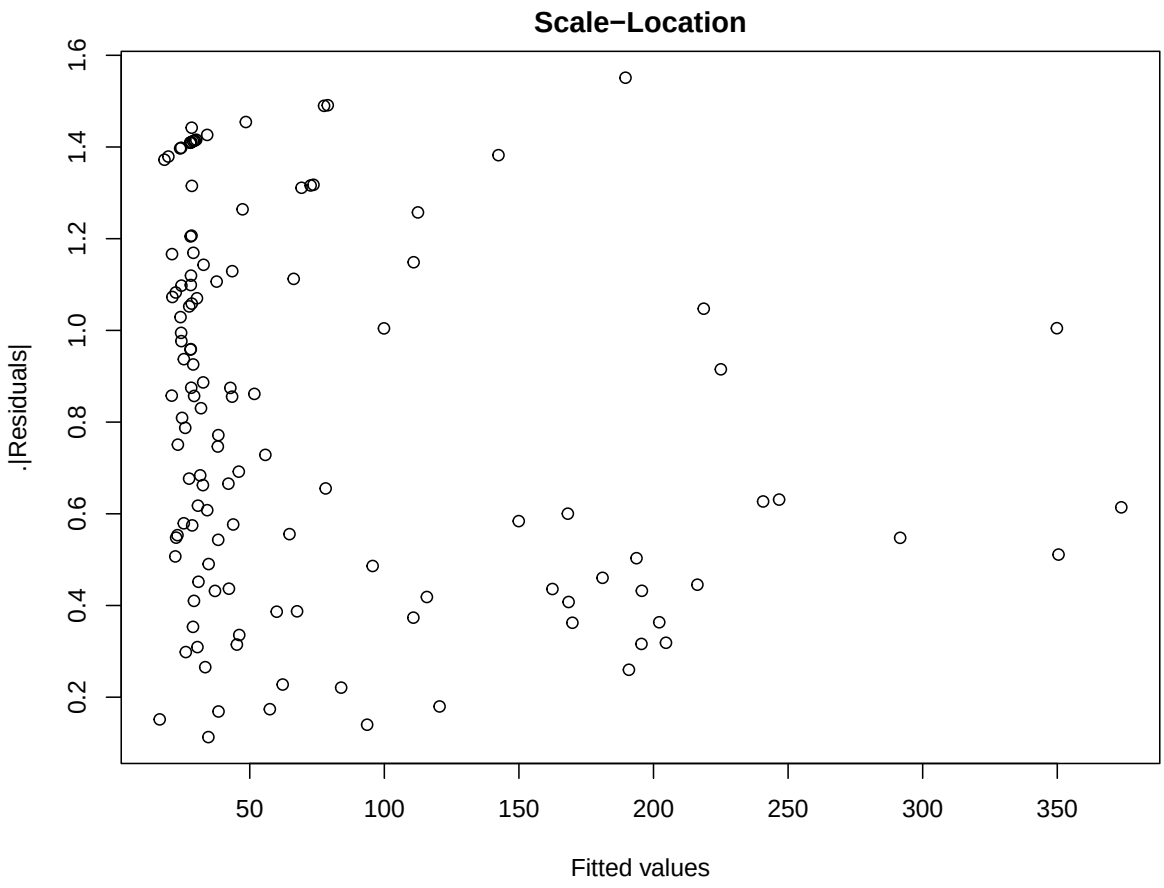


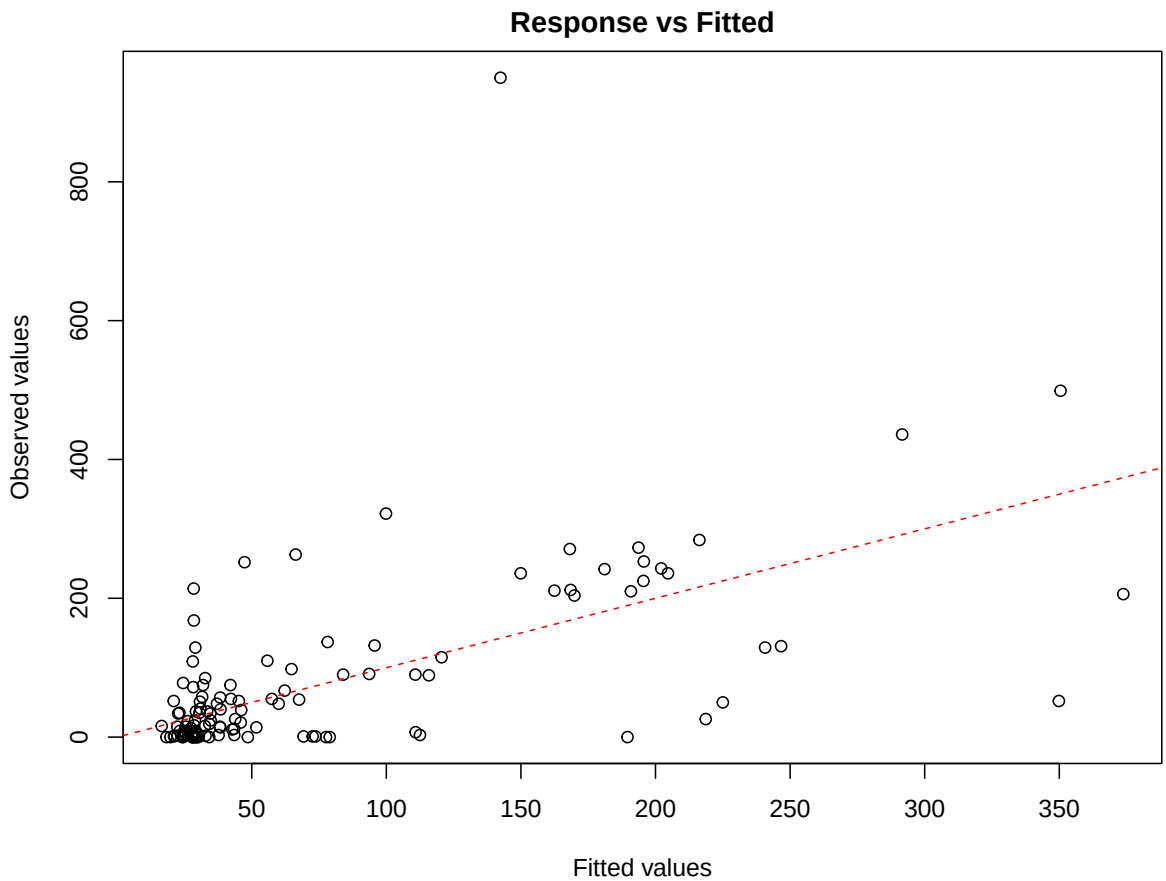

```
### Model 3 : 2b. Temperature only

Model data dimensions after removing NAs: 125 x 4
GAM fitted successfully
Model Summary:

Family: Negative Binomial(0.485)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(view_id,
    bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    3.900      0.279   13.97   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df Chi.sq  p-value
s(abundance_t_minus_1) 2.067  2.550 35.323 7.41e-07 ***
s(mean_temp)           1.364  1.621  0.881    0.667
s(view_id)             2.741  8.000  3.496    0.244
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.299   Deviance explained = 25.9%
-REML = 609.56  Scale est. = 1         n = 125


Diagnostic plots:
```



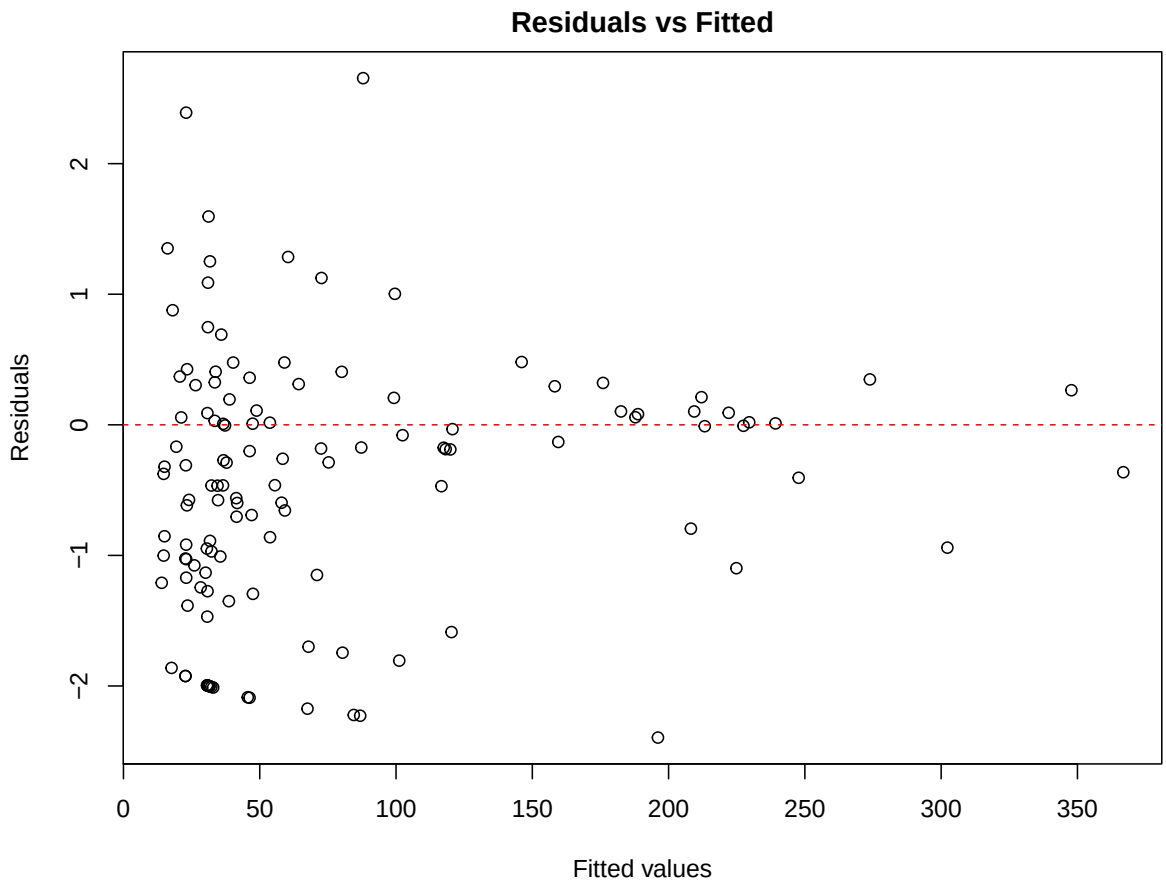

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−2
−1
0
1
2

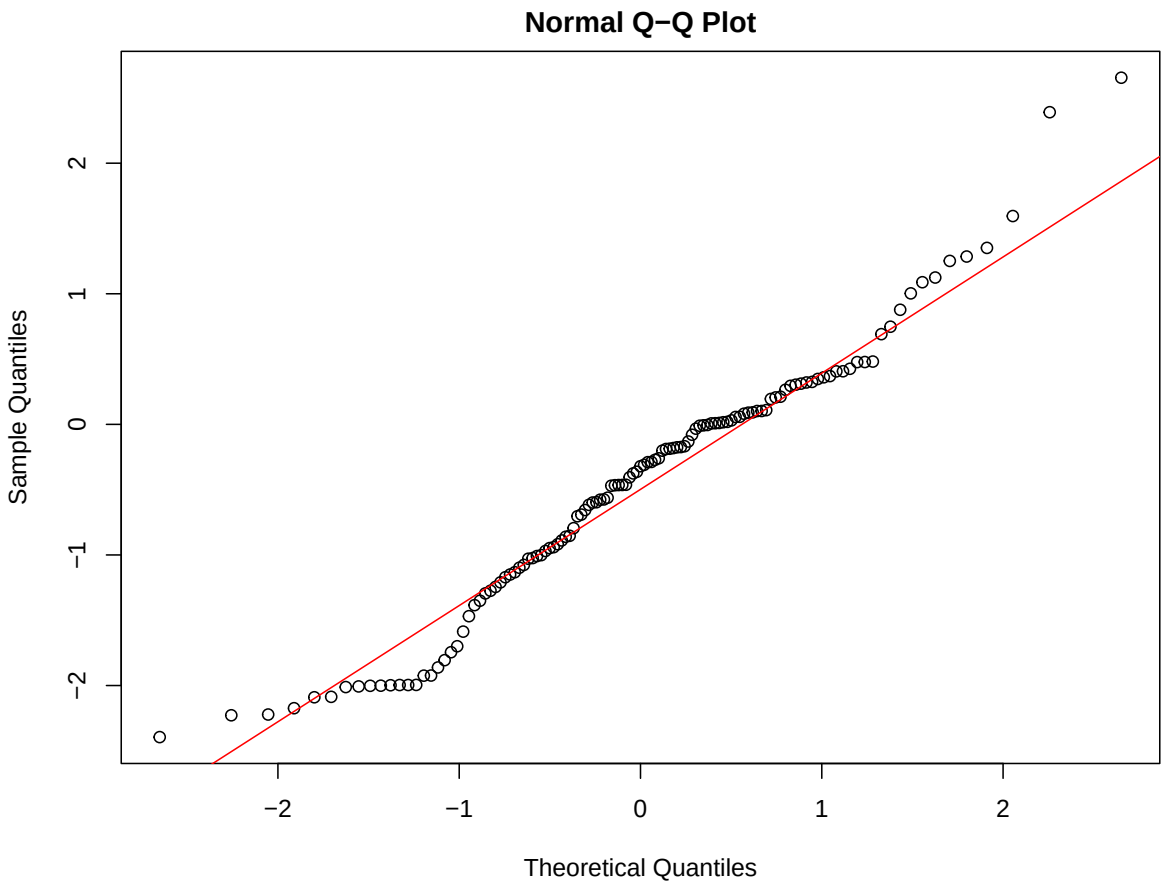

Scale−Location
√|Residuals|
Fitted values
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
50
100
150
200
250
300
350

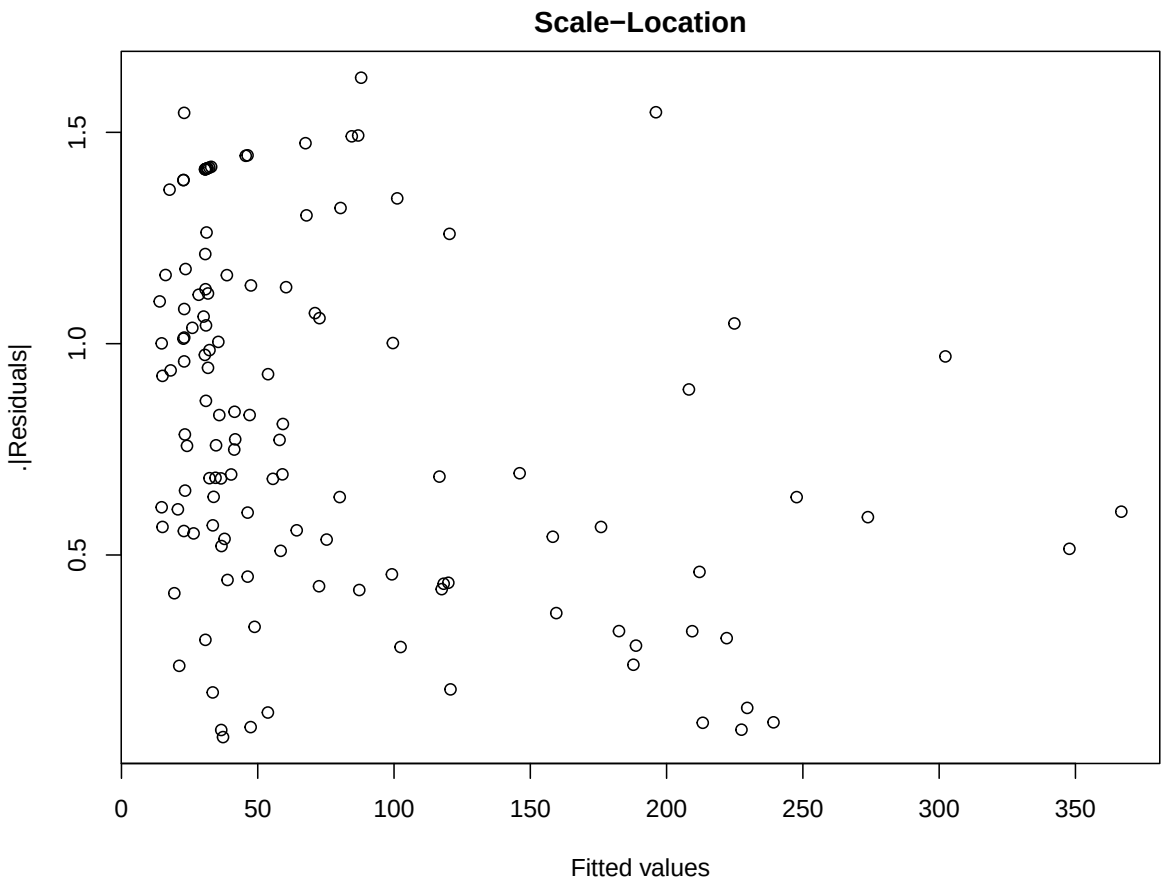


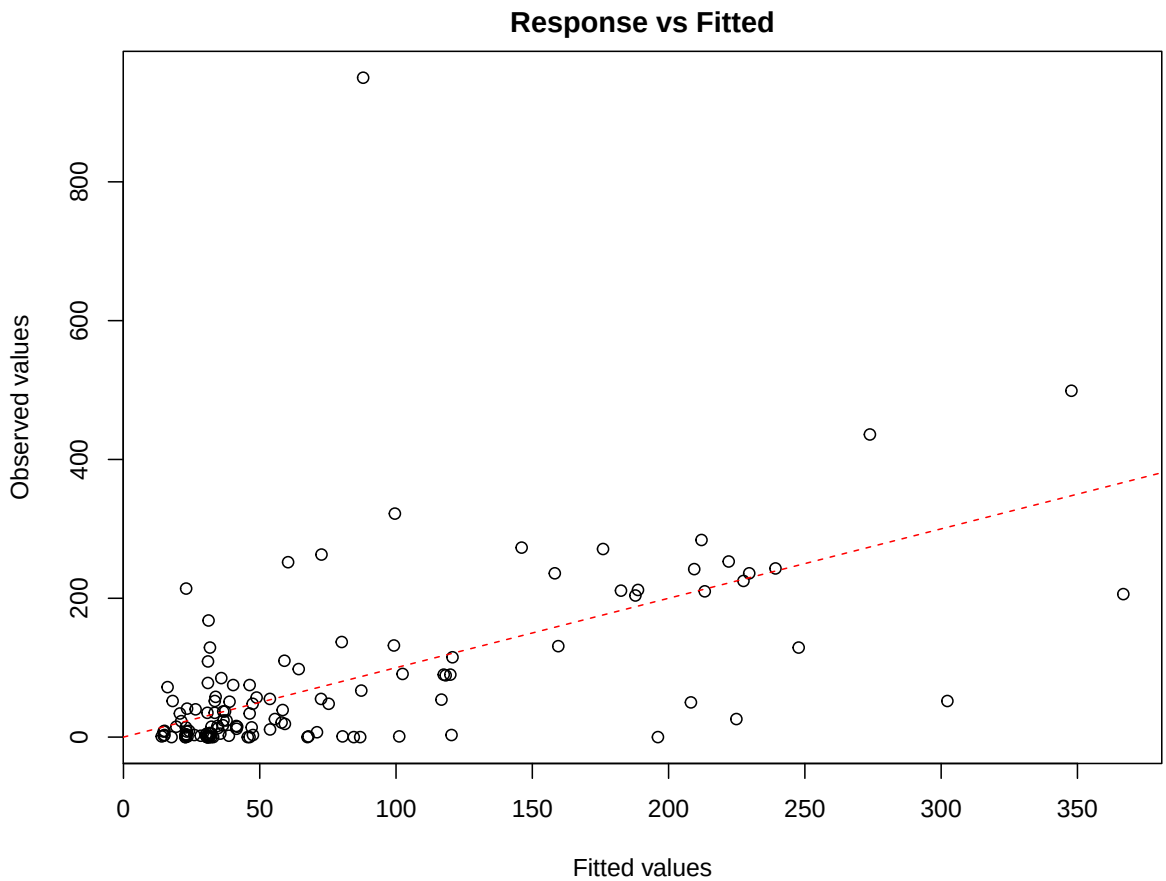

```
### Model 4 : 2c. Sunlight only

Model data dimensions after removing NAs: 125 x 4
GAM fitted successfully
Model Summary:

Family: Negative Binomial(0.476)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(sunlight_proportion) +
    s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.9802     0.1304   30.52   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                             edf Ref.df Chi.sq  p-value
s(abundance_t_minus_1) 2.0882806  2.591 32.711 1.44e-06 ***
s(sunlight_proportion) 1.0006397  1.001  3.438   0.0638 .
s(view_id)             0.0001309  8.000  0.000   0.4270
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.283   Deviance explained =   23%
-REML = 608.65  Scale est. = 1         n = 125


Diagnostic plots:
```



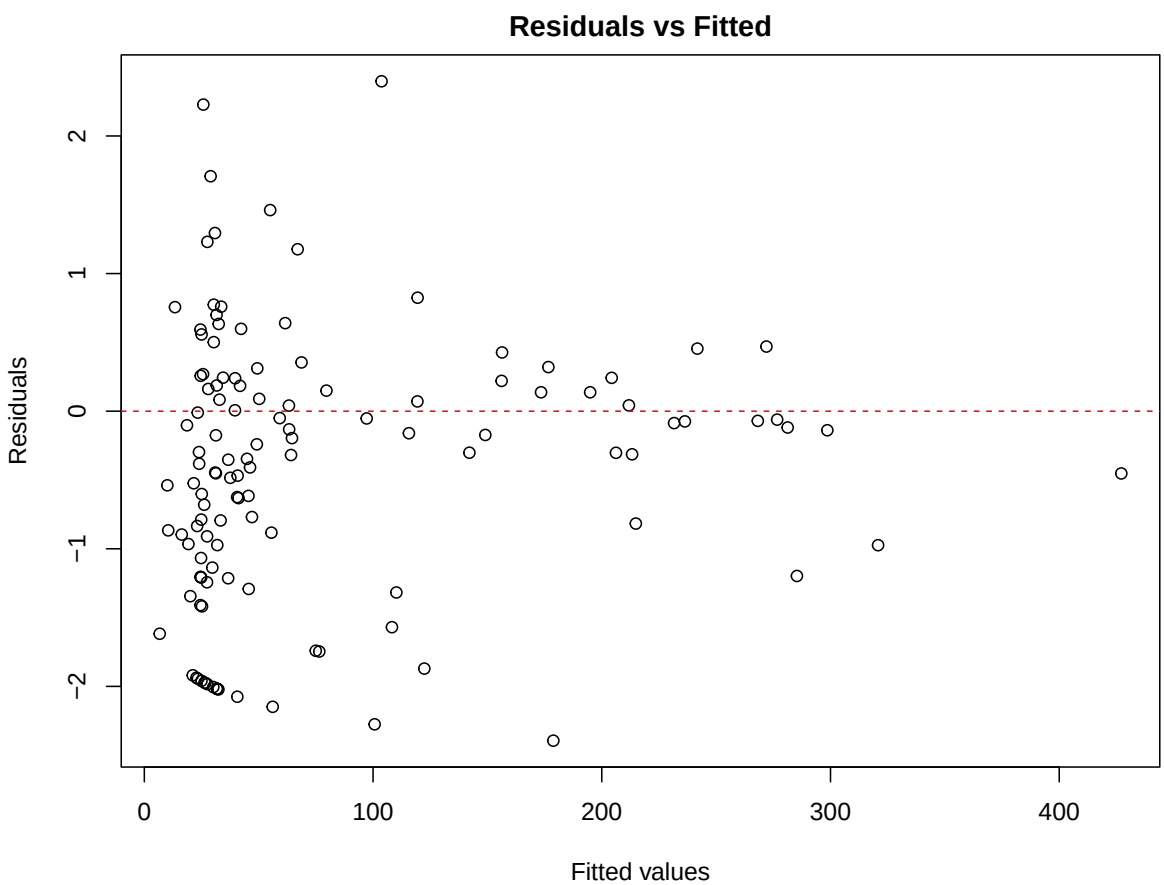

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−2
−1
0
1
2

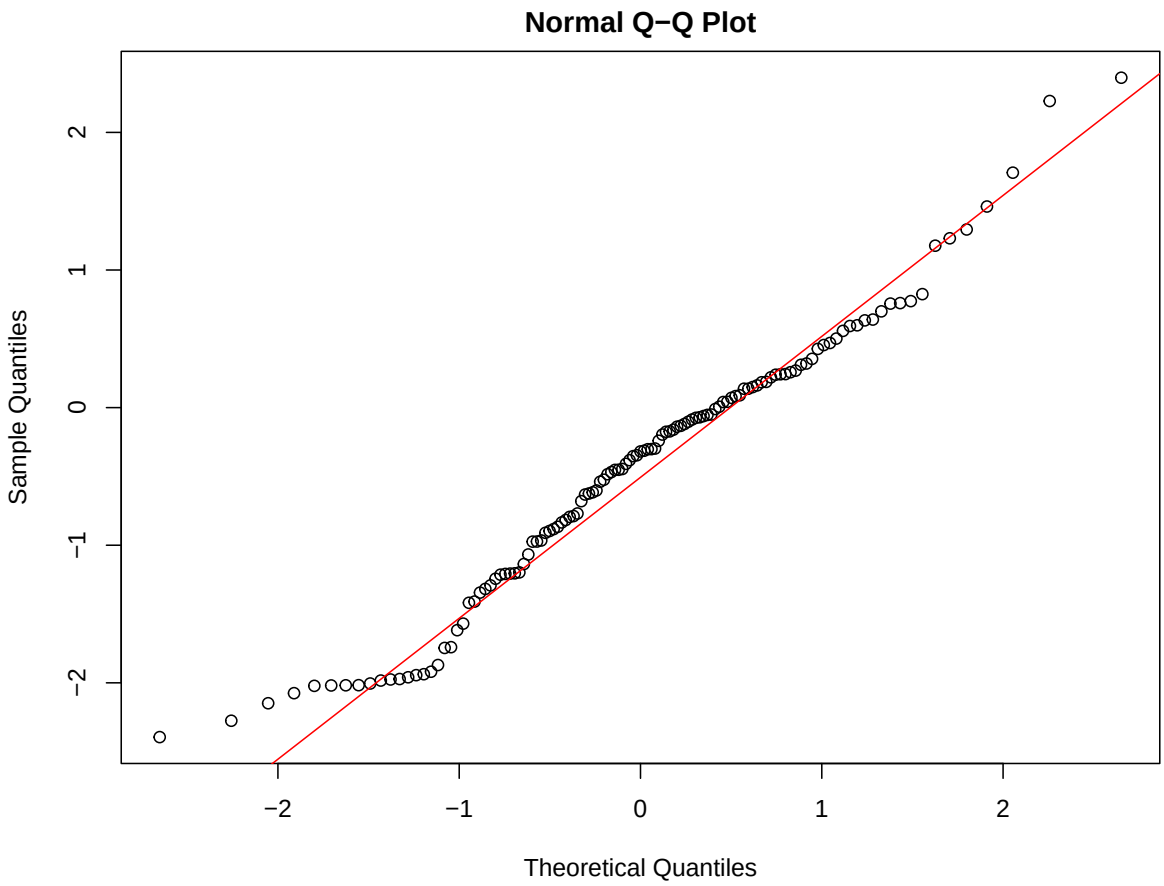

Scale−Location
√|Residuals|
0.5
1.0
1.5
0
50
100
150
200
250
300
350
Fitted values

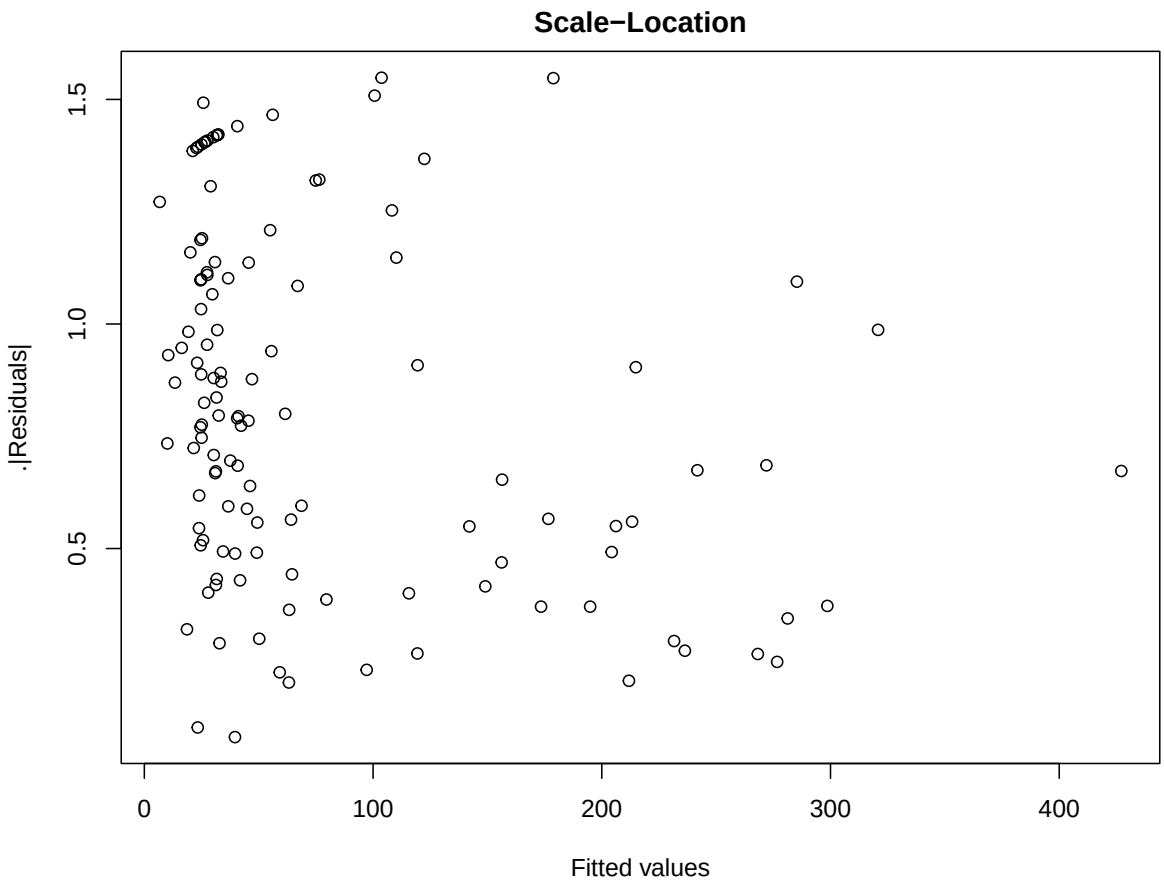


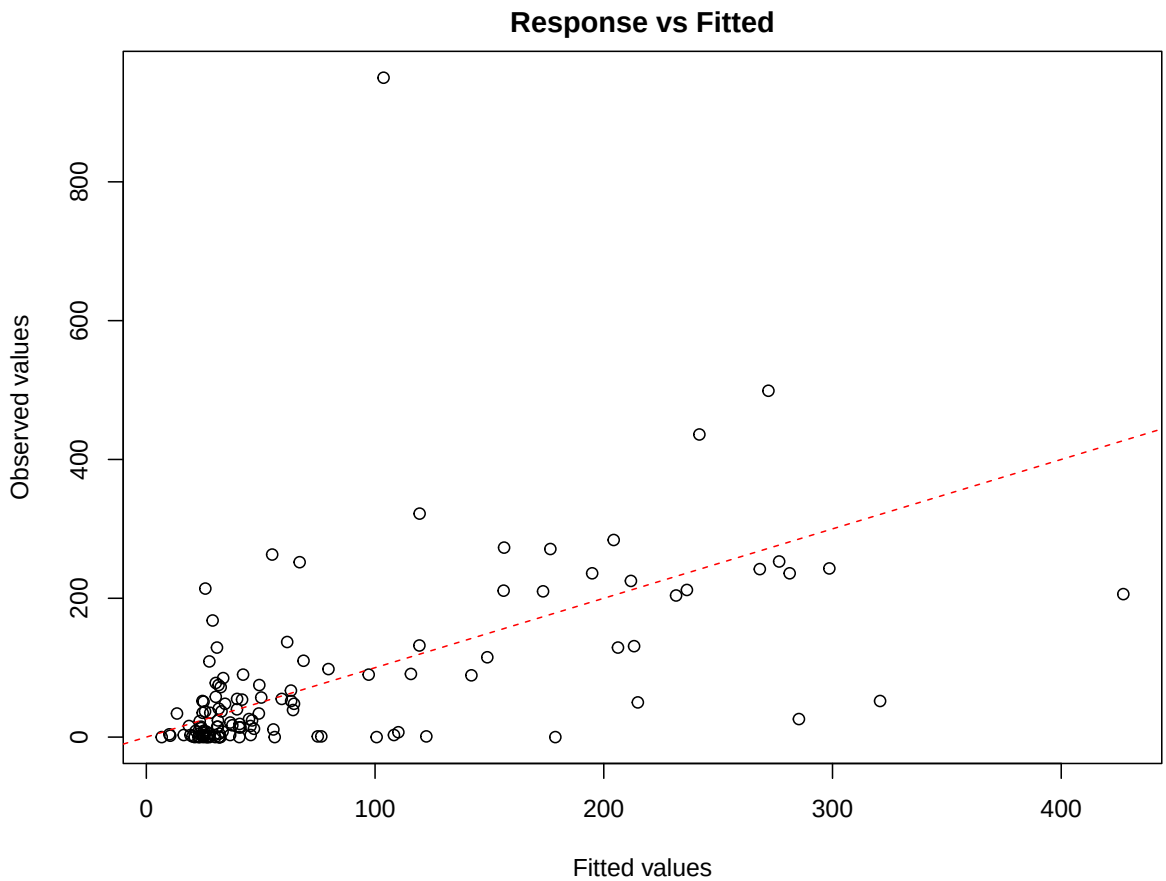

```
### Model 5 : 3a. Wind + Temp

Model data dimensions after removing NAs: 125 x 5
GAM fitted successfully
Model Summary:

Family: Negative Binomial(0.485)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.9319     0.1293    30.4   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                            edf Ref.df Chi.sq p-value
s(abundance_t_minus_1) 2.227629  2.747 39.996  <2e-16 ***
s(mean_wind_speed)     2.456670  3.035  4.852   0.179
s(mean_temp)           1.698262  2.060  1.846   0.423
s(view_id)             0.001081  8.000  0.001   0.499
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =   0.23   Deviance explained = 26.1%
-REML = 609.68  Scale est. = 1         n = 125


Diagnostic plots:
```



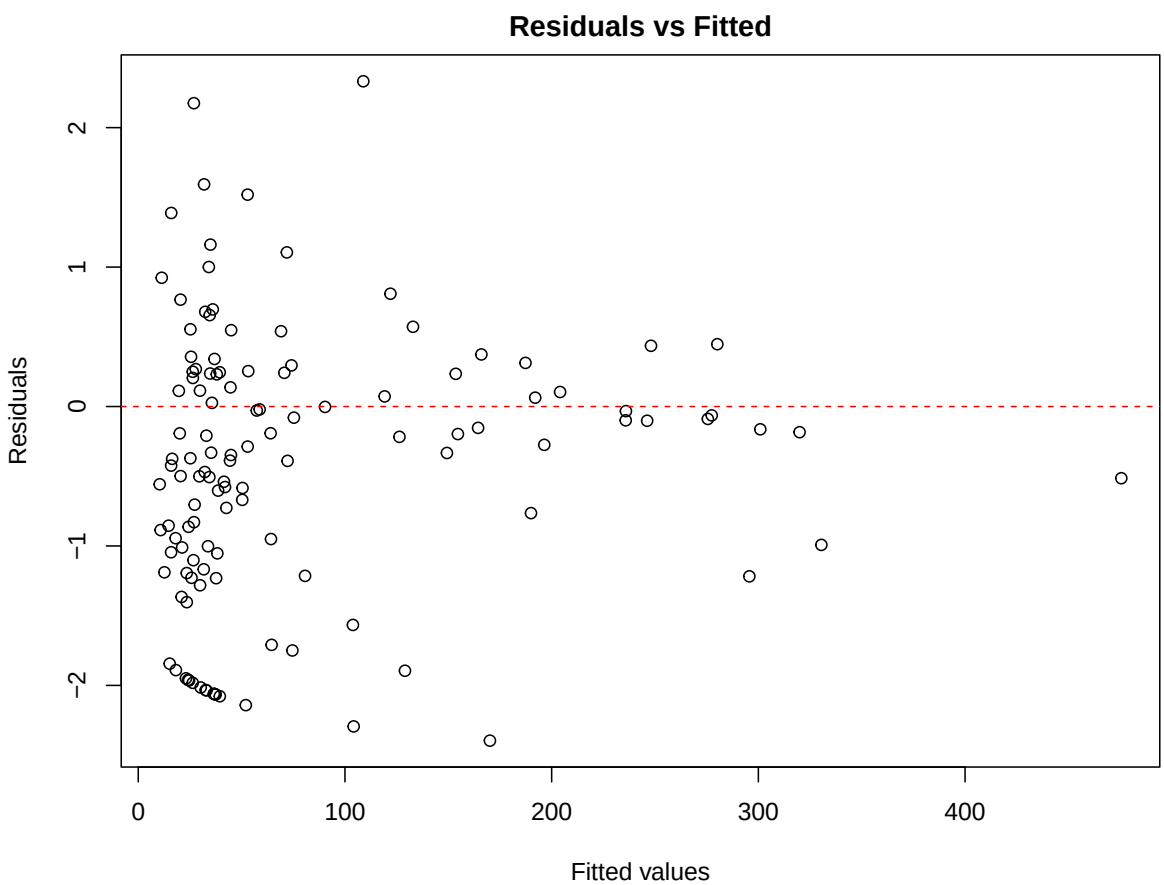

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−2
−1
0
1
2

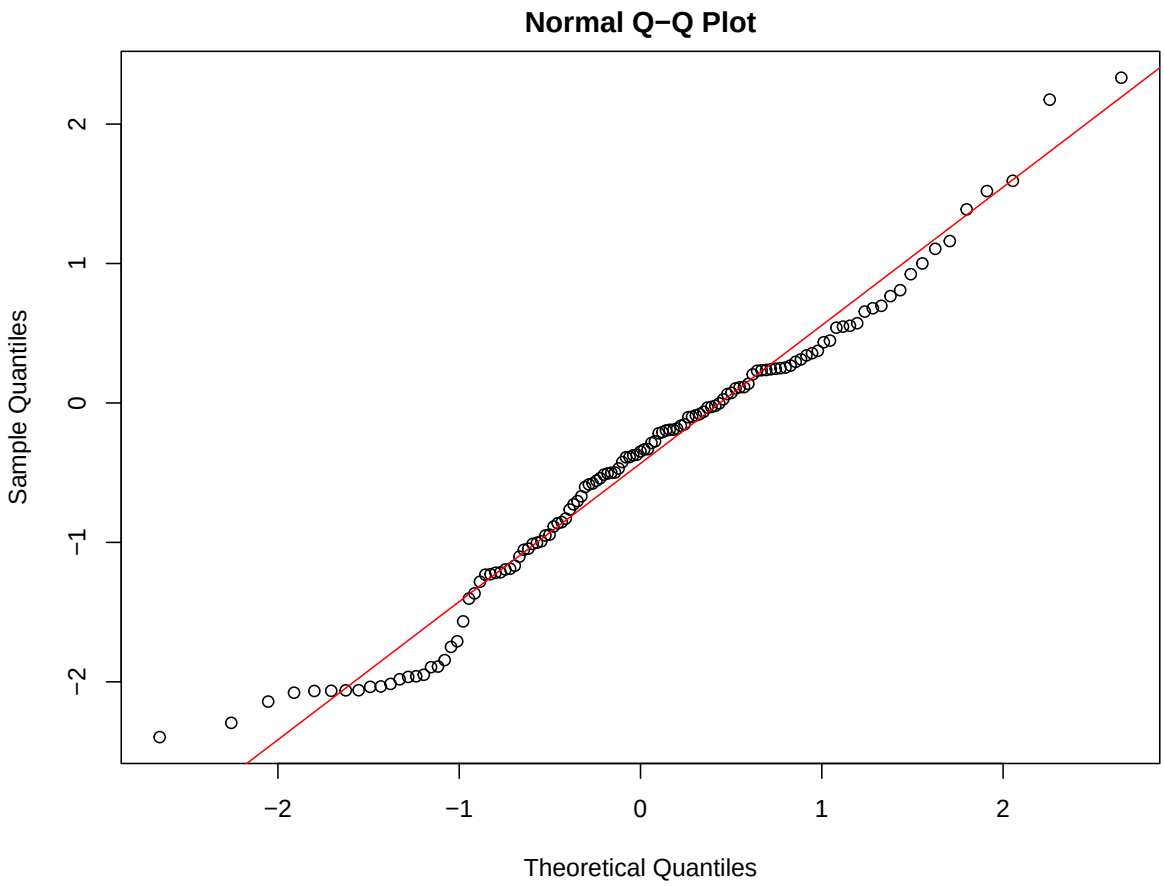

Scale−Location
√|Residuals|
0.5
1.0
1.5
0
100
200
300
400
Fitted values

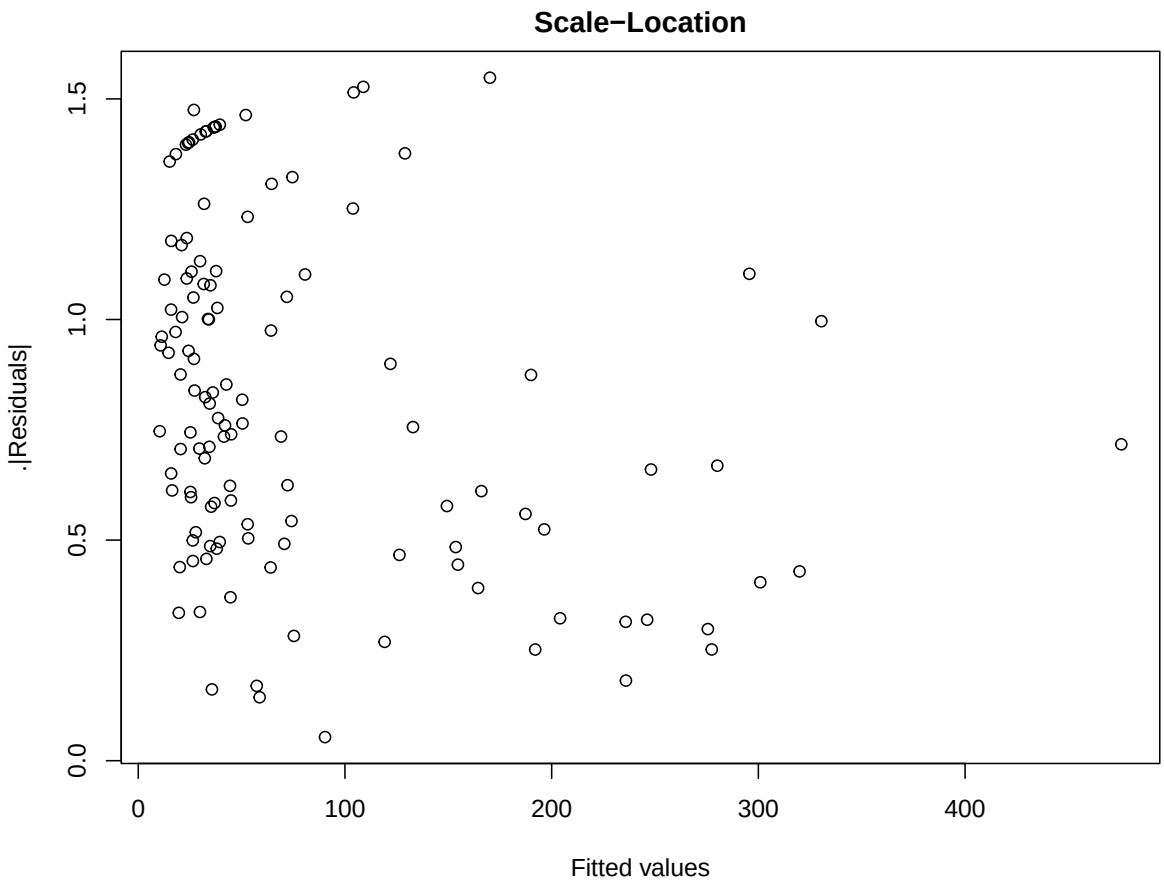


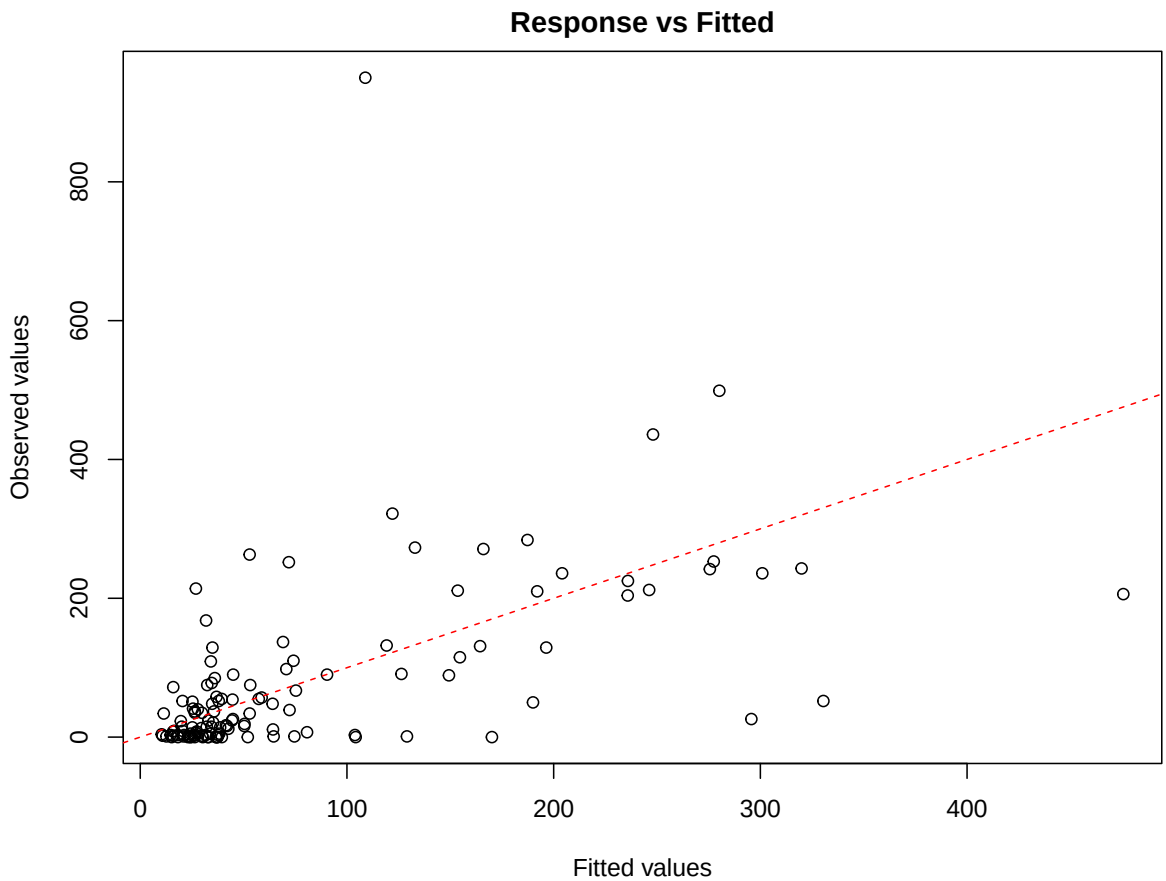

```
### Model 6 : 3b. Wind + Sunlight

Model data dimensions after removing NAs: 125 x 5
GAM fitted successfully
Model Summary:

Family: Negative Binomial(0.491)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(sunlight_proportion) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.9259     0.1285   30.54   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                             edf Ref.df Chi.sq  p-value
s(abundance_t_minus_1) 2.1132047  2.617 36.906 1.12e-06 ***
s(mean_wind_speed)     2.4587876  3.057  5.411   0.1436
s(sunlight_proportion) 1.0002324  1.000  4.103   0.0428 *
s(view_id)             0.0002106  8.000  0.000   0.6399
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.229   Deviance explained = 26.5%
-REML = 608.33  Scale est. = 1         n = 125


Diagnostic plots:
```



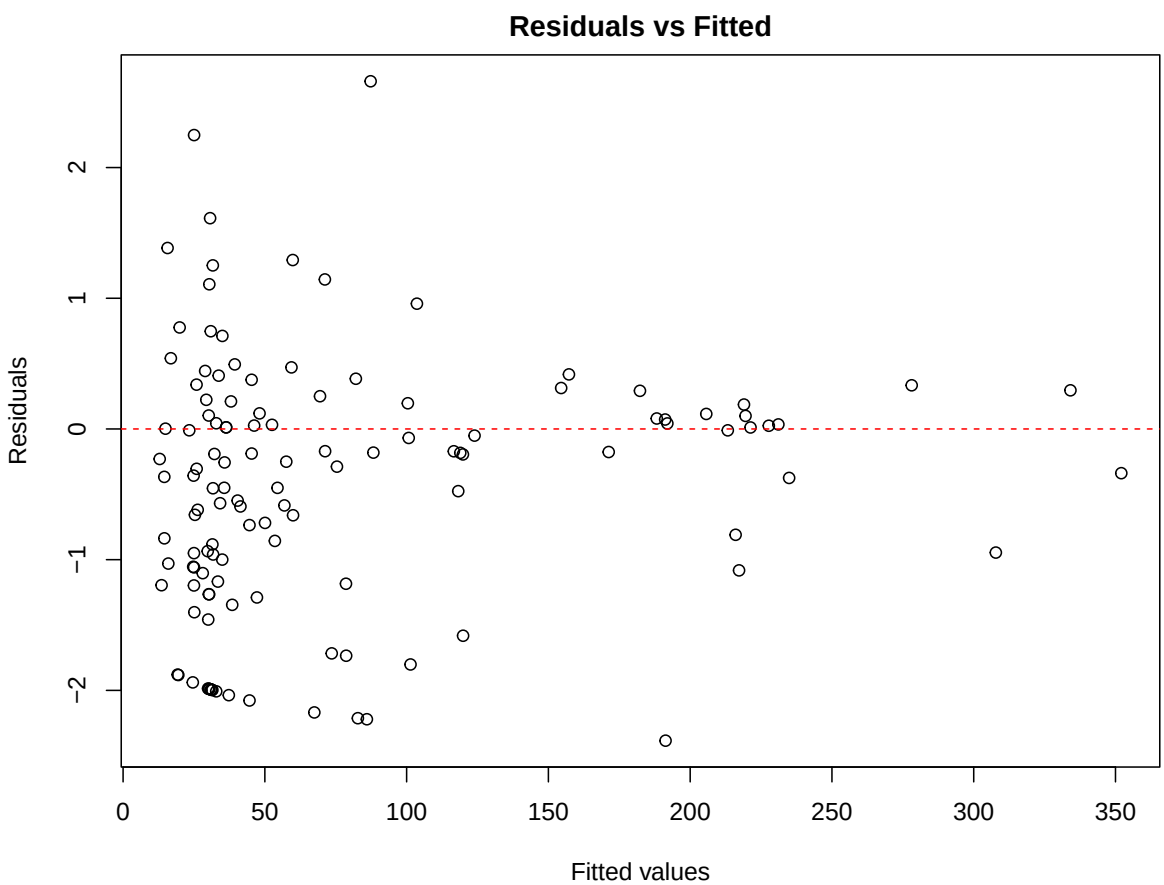

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−2
−1
0
1
2

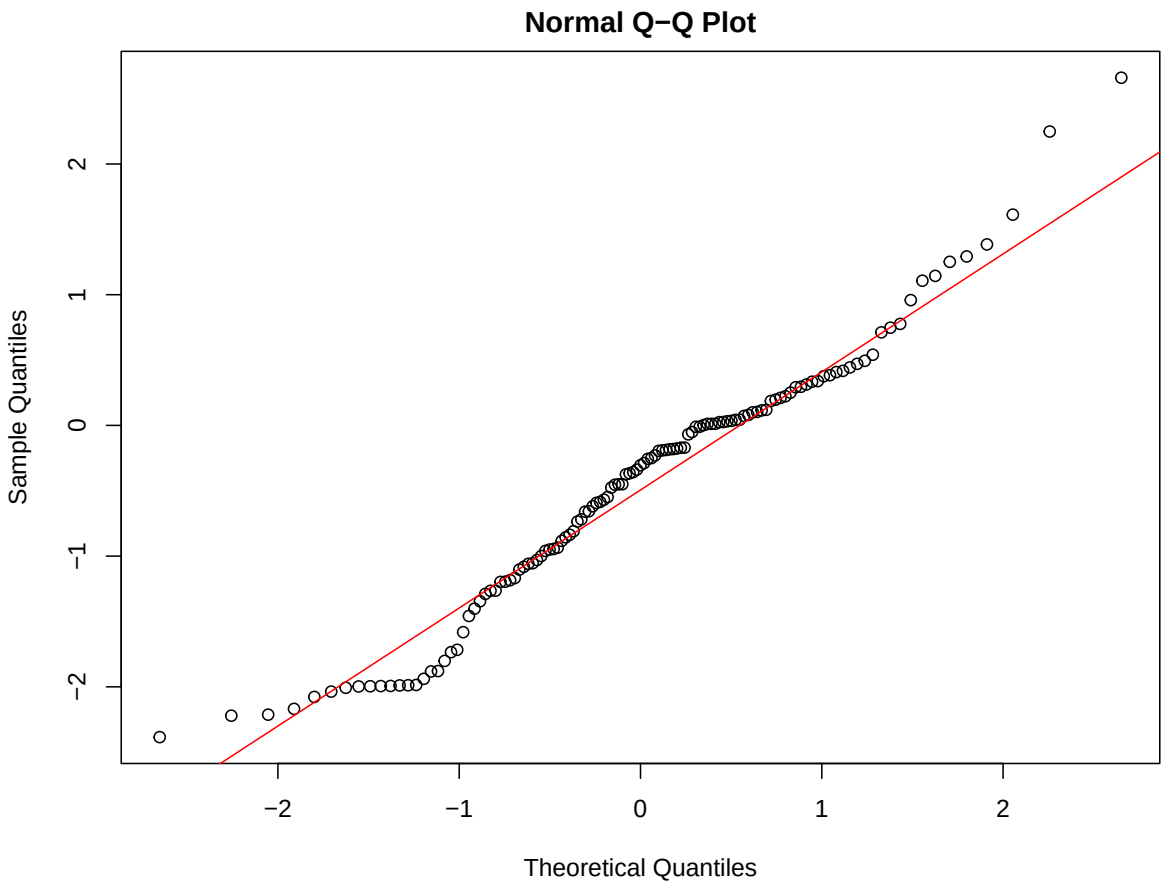

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
0
100
200
300
400
Fitted values

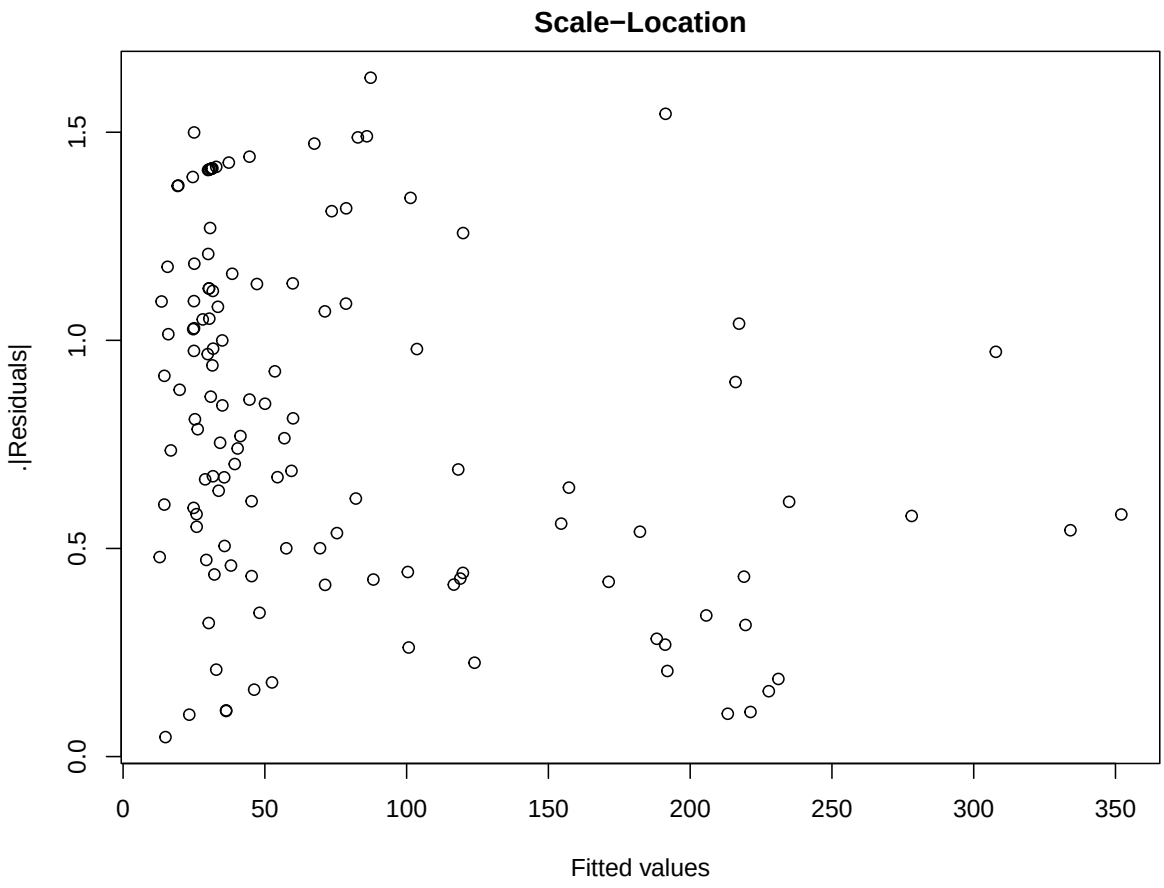


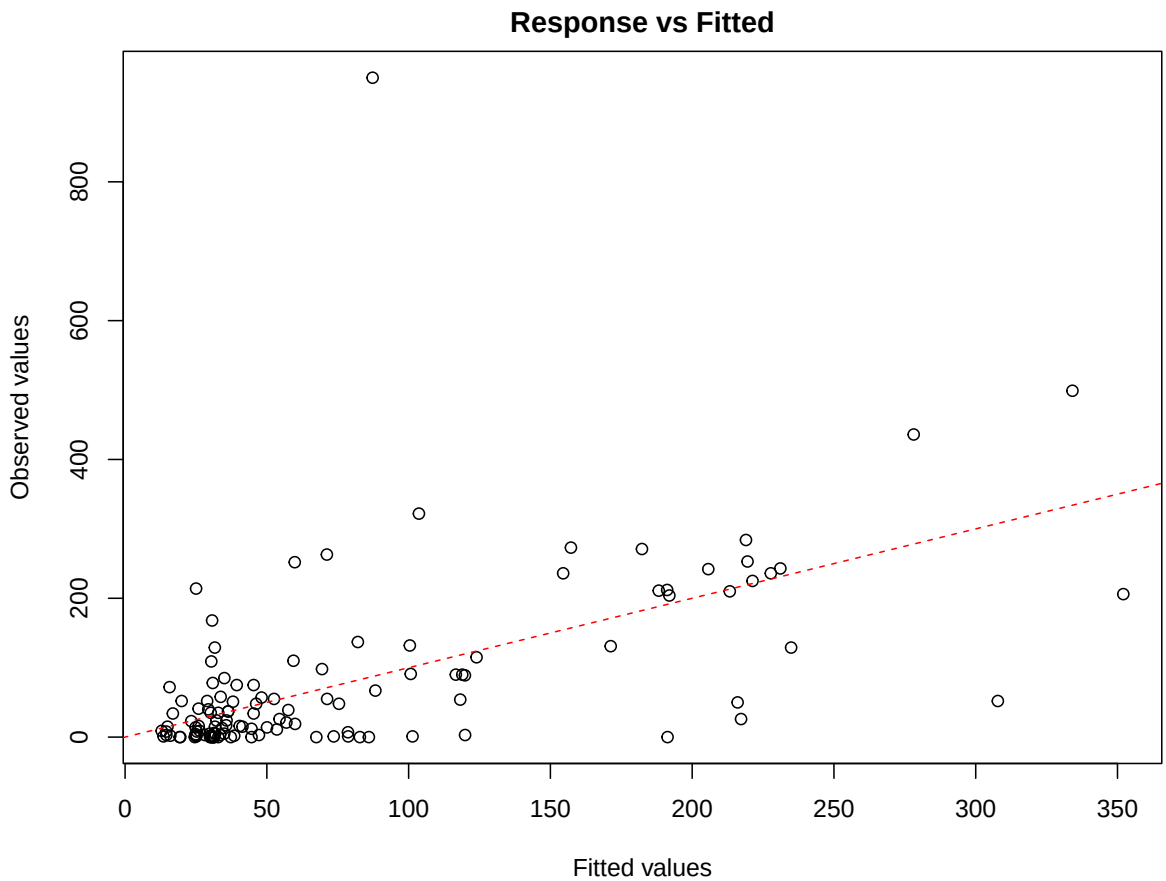

```
### Model 7 : 3c. Temp + Sunlight

Model data dimensions after removing NAs: 125 x 5
GAM fitted successfully
Model Summary:

Family: Negative Binomial(0.474)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_temp) + s(sunlight_proportion) +
    s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.9778     0.1307   30.43   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                             edf Ref.df Chi.sq  p-value
s(abundance_t_minus_1) 2.102e+00  2.607 31.073 1.45e-06 ***
s(mean_temp)           1.001e+00  1.002  0.258   0.6126
s(sunlight_proportion) 1.000e+00  1.000  3.378   0.0661 .
s(view_id)             4.027e-05  8.000  0.000   0.4400
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.281   Deviance explained = 23.2%
-REML = 609.58  Scale est. = 1         n = 125


Diagnostic plots:
```



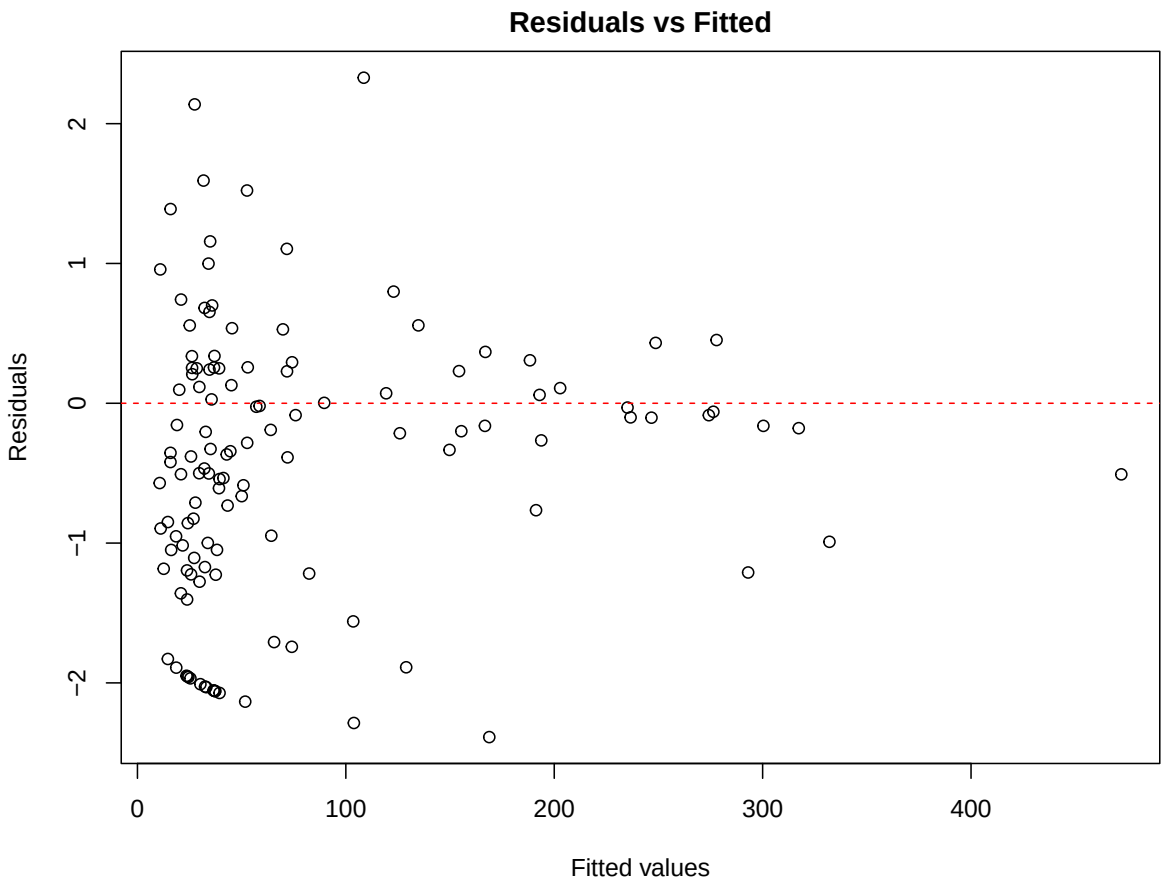

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−2
−1
0
1
2

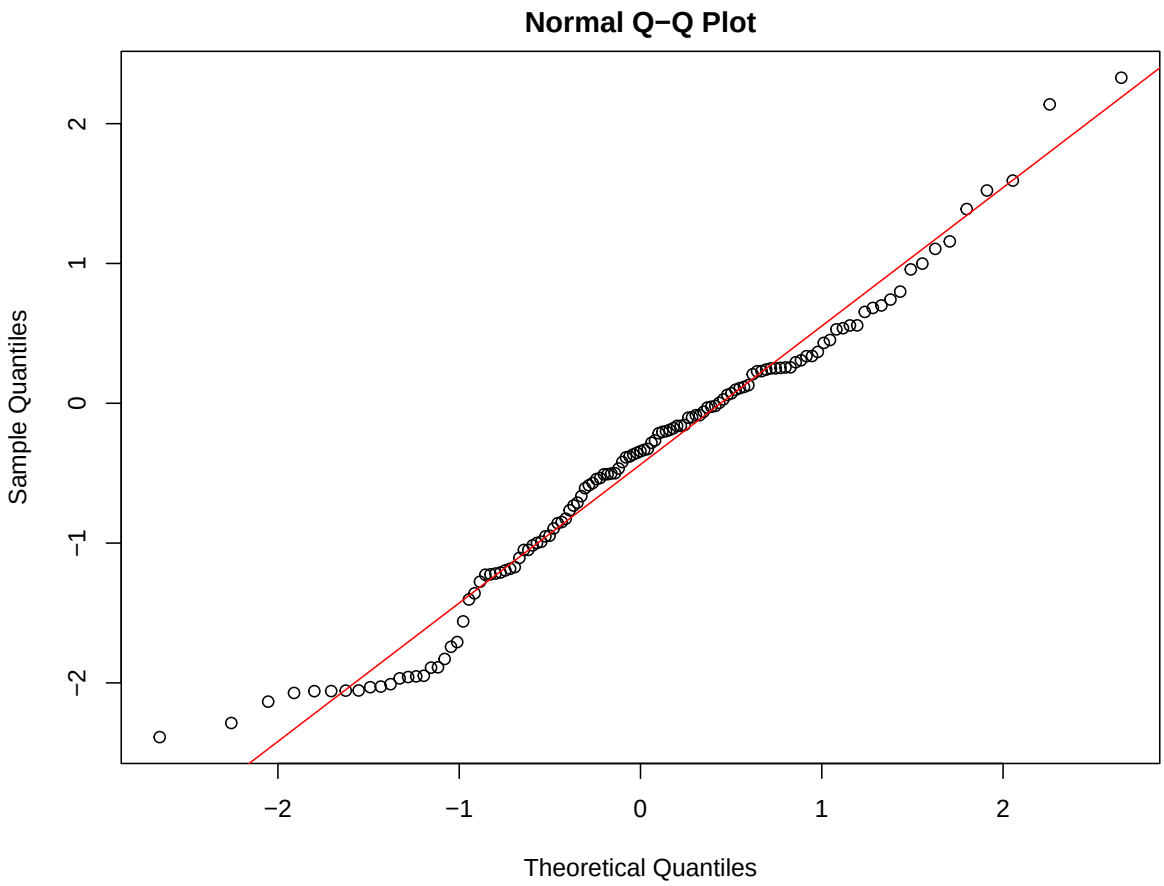

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
0
50
100
150
200
250
300
350
Fitted values

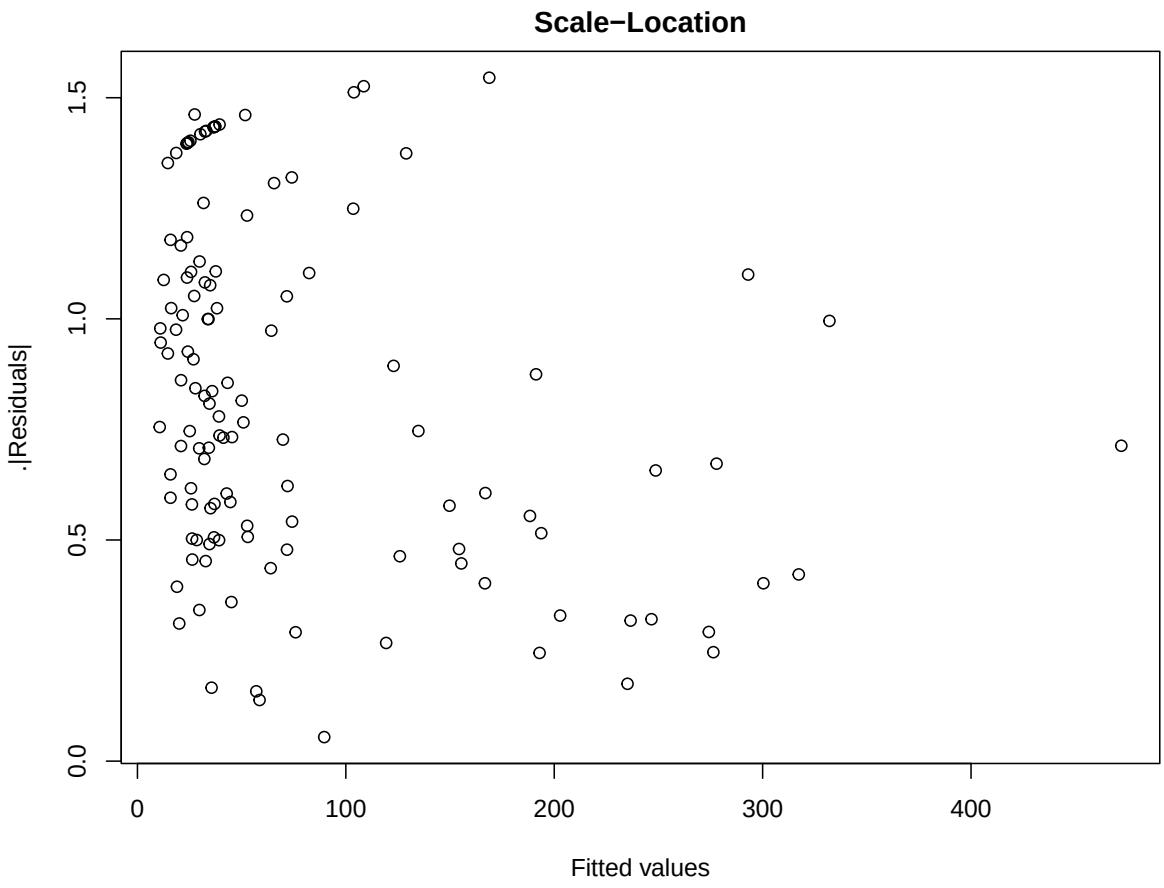


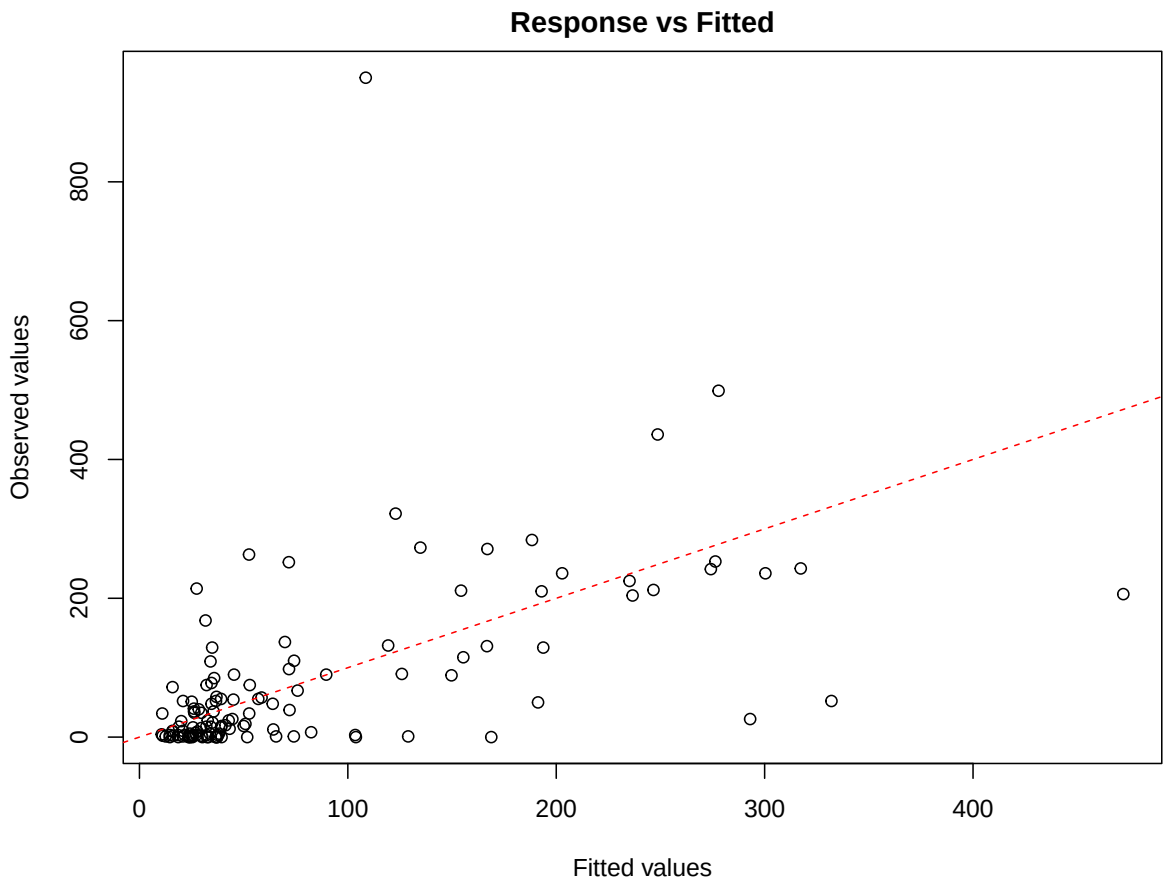

```
### Model 8 : 3d. All additive

Model data dimensions after removing NAs: 125 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(0.487)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    3.926      0.129   30.43   <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                            edf Ref.df Chi.sq  p-value
s(abundance_t_minus_1) 2.1111294  2.613 35.511 1.19e-06 ***
s(mean_wind_speed)     2.4580459  3.050  5.307   0.1500
s(mean_temp)           1.0003649  1.001  0.013   0.9106
s(sunlight_proportion) 1.0001642  1.000  3.971   0.0463 *
s(view_id)             0.0001372  8.000  0.000   0.6385
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.224   Deviance explained = 26.5%
-REML = 609.33  Scale est. = 1         n = 125


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
2
1
0
−1
−2
0
100
200
300
400

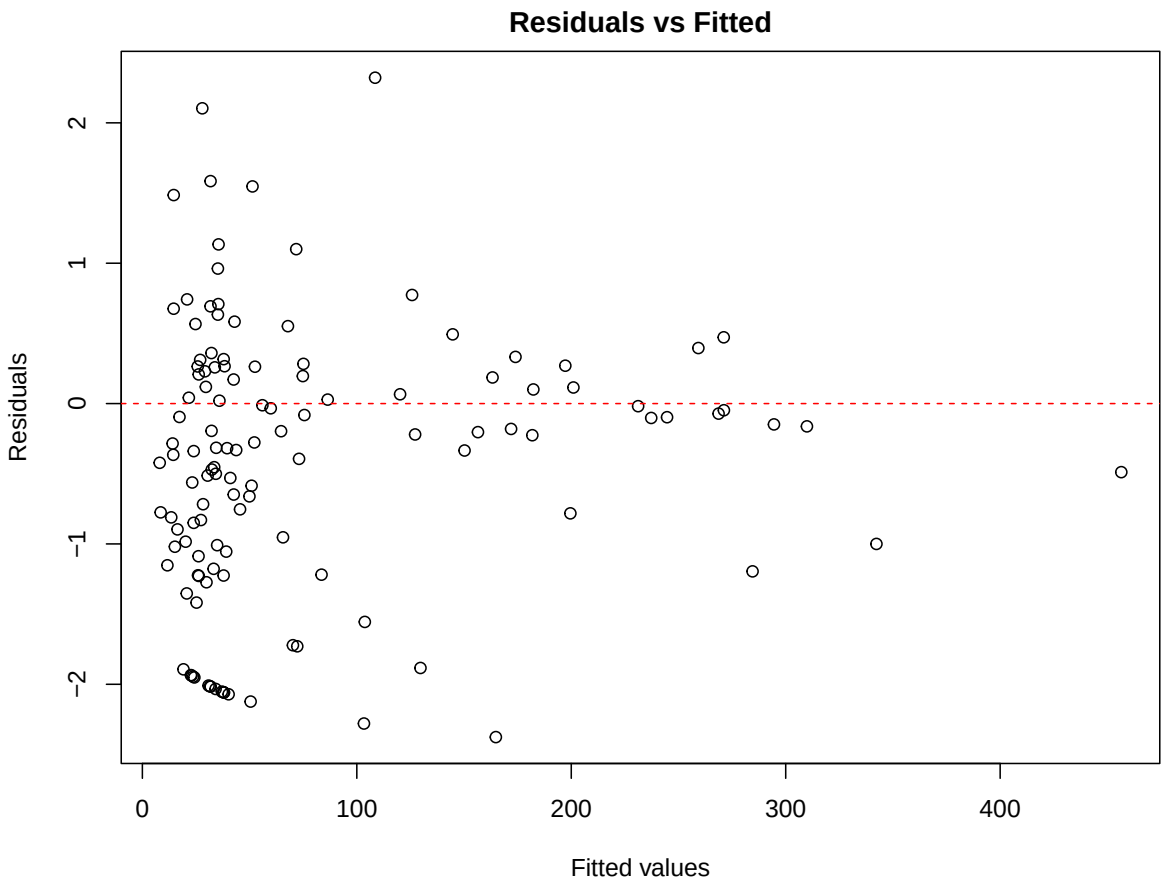

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−2
−1
0
1
2

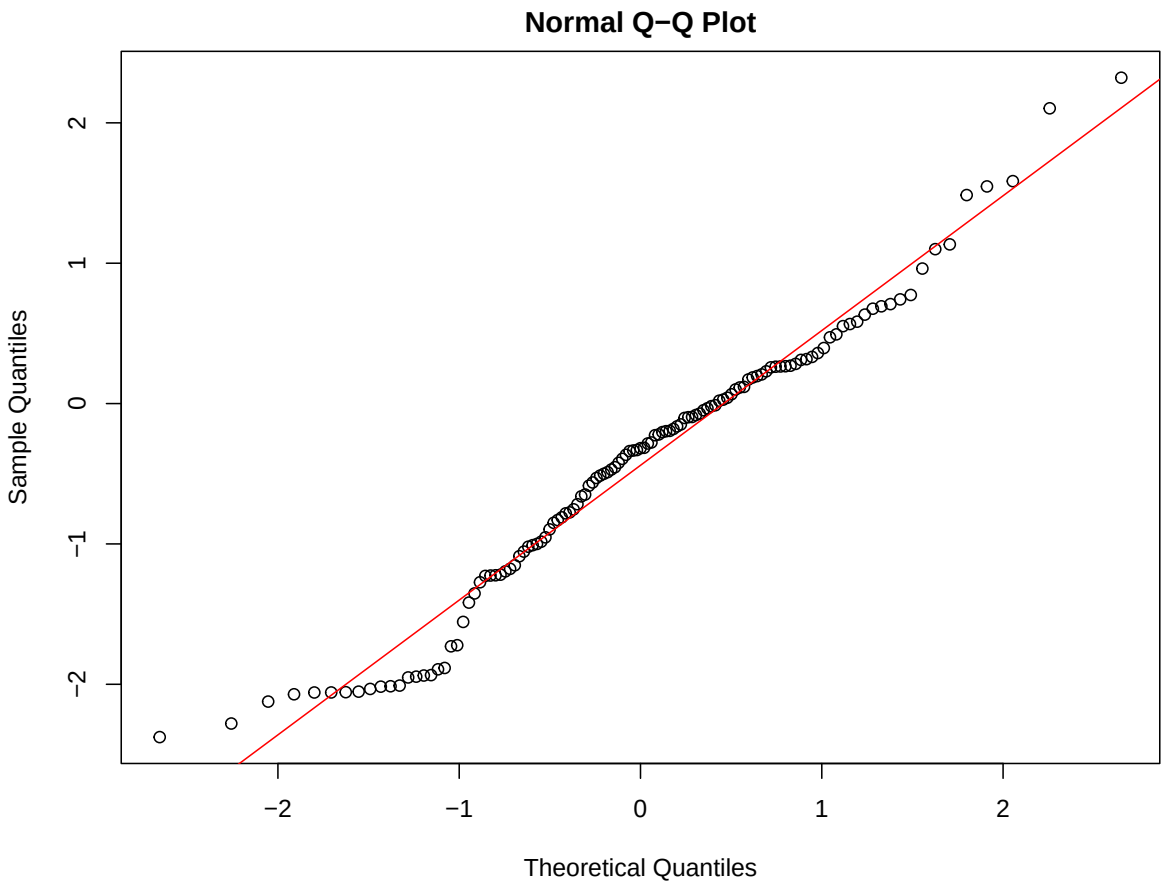

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
0
100
200
300
400
Fitted values

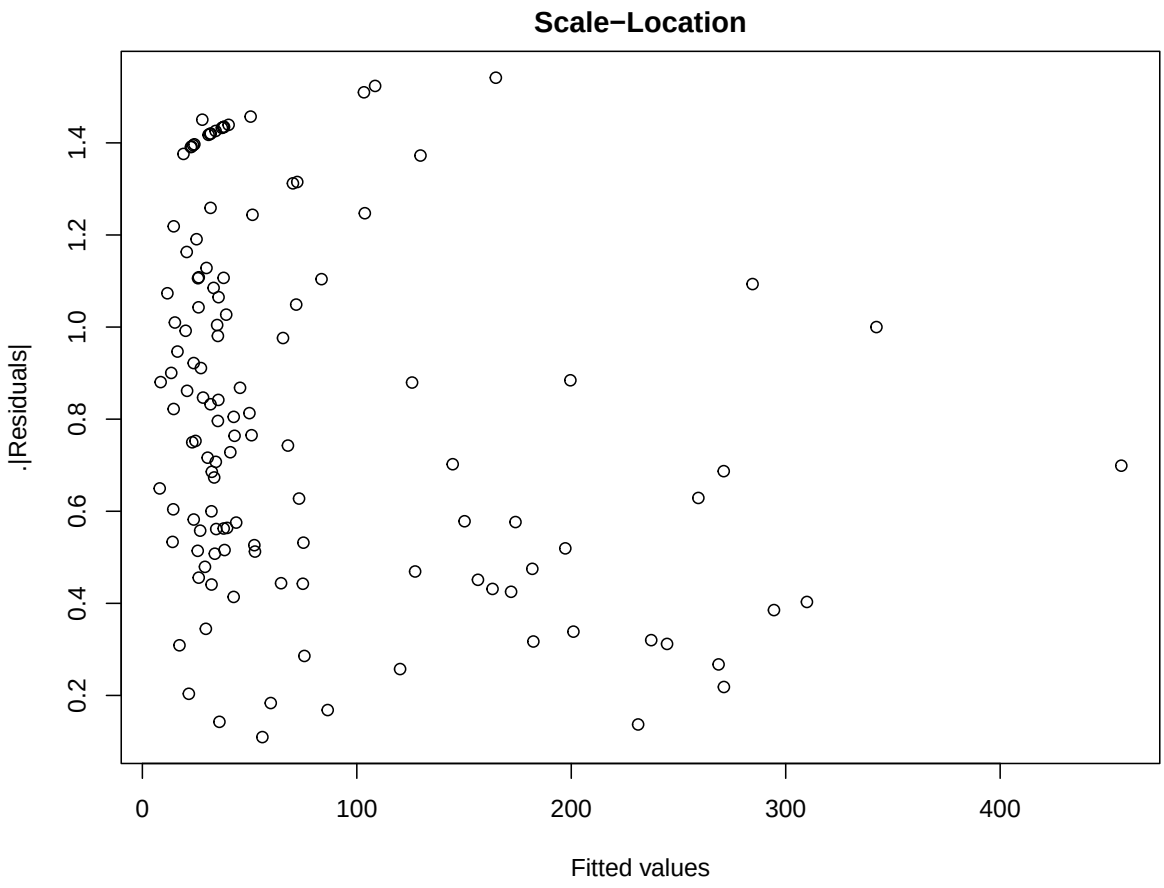


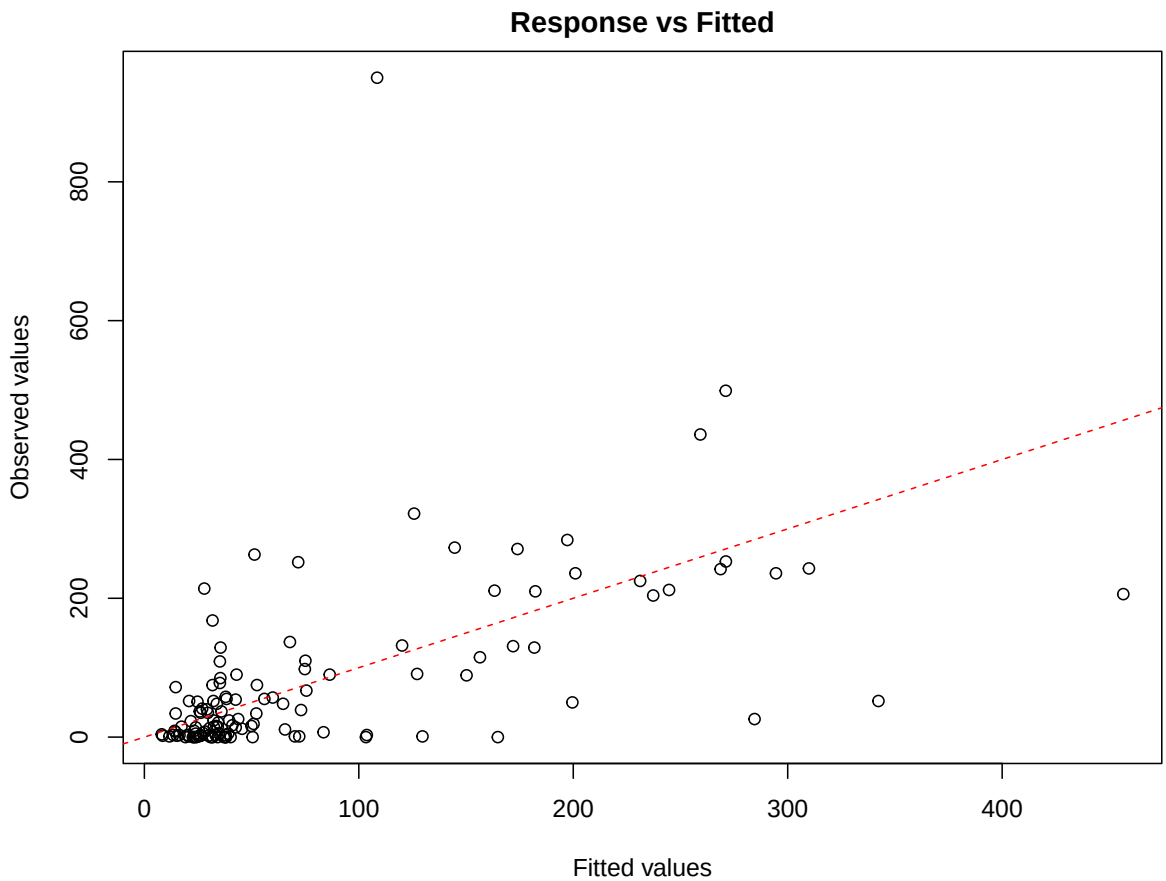

```
### Model 9 : 4a. Wind x Temp interaction

Model data dimensions after removing NAs: 125 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(0.484)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)    3.917      0.130   30.13   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                   edf Ref.df Chi.sq p-value
s(abundance_t_minus_1)        2.1003379  2.600 33.415 1.4e-06 ***
s(mean_wind_speed)            2.3903105  2.966  5.884  0.1507
s(mean_temp)                  1.0003563  1.001  0.155  0.6941
s(sunlight_proportion)        1.0002711  1.001  4.783  0.0288 *
ti(mean_wind_speed,mean_temp) 1.0010587  1.002  0.448  0.5047
s(view_id)                    0.0001848  8.000  0.000  0.6421
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.226   Deviance explained = 26.5%
-REML = 607.25  Scale est. = 1         n = 125


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
2
1
0
−1
−2
0
100
200
300
400

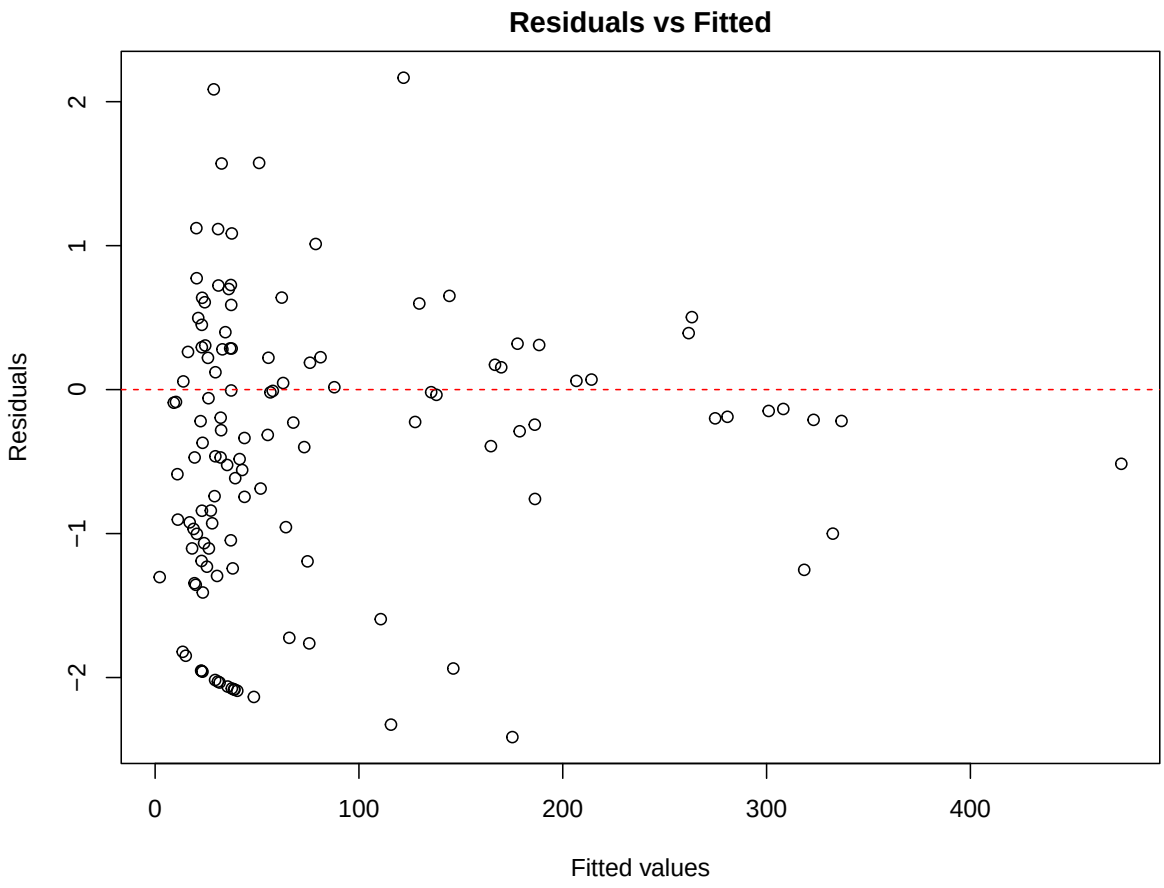

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−2
−1
0
1
2

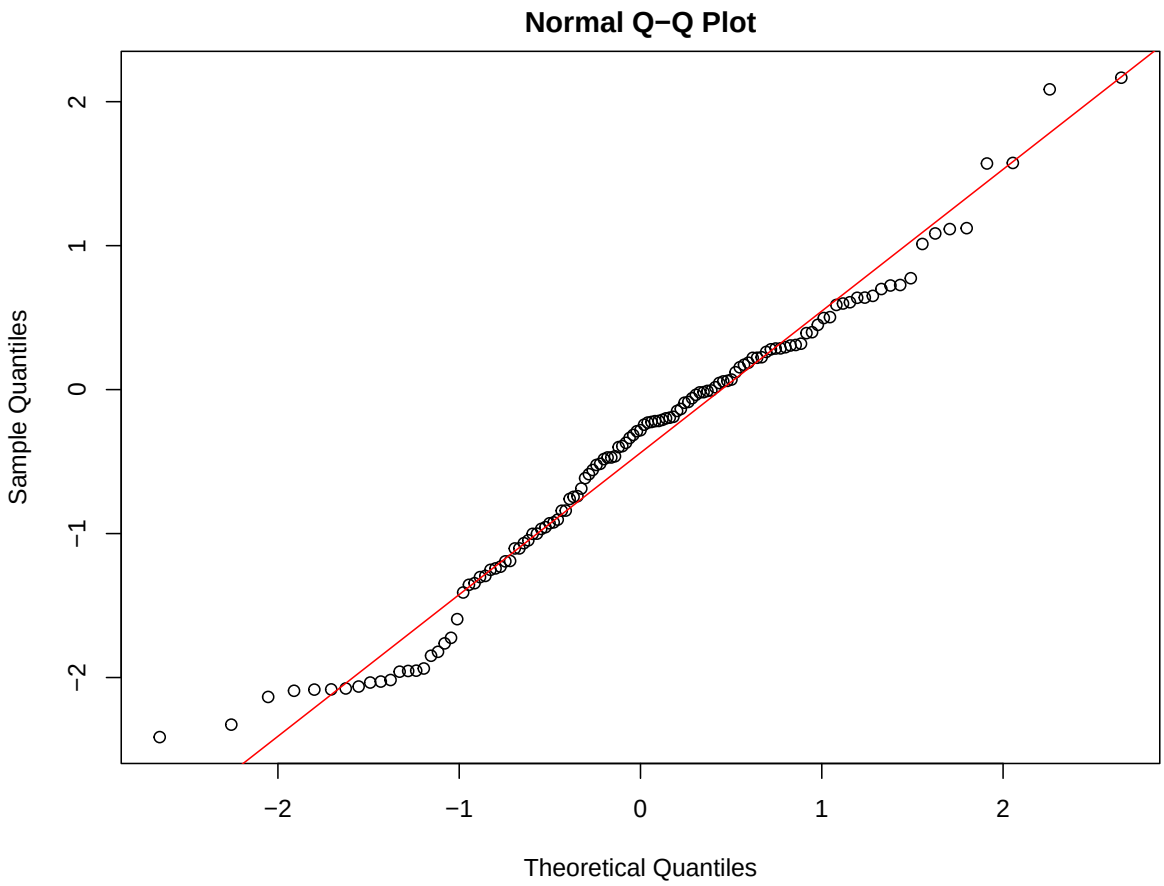

Scale−Location
√|Residuals|
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0
100
200
300
400
Fitted values

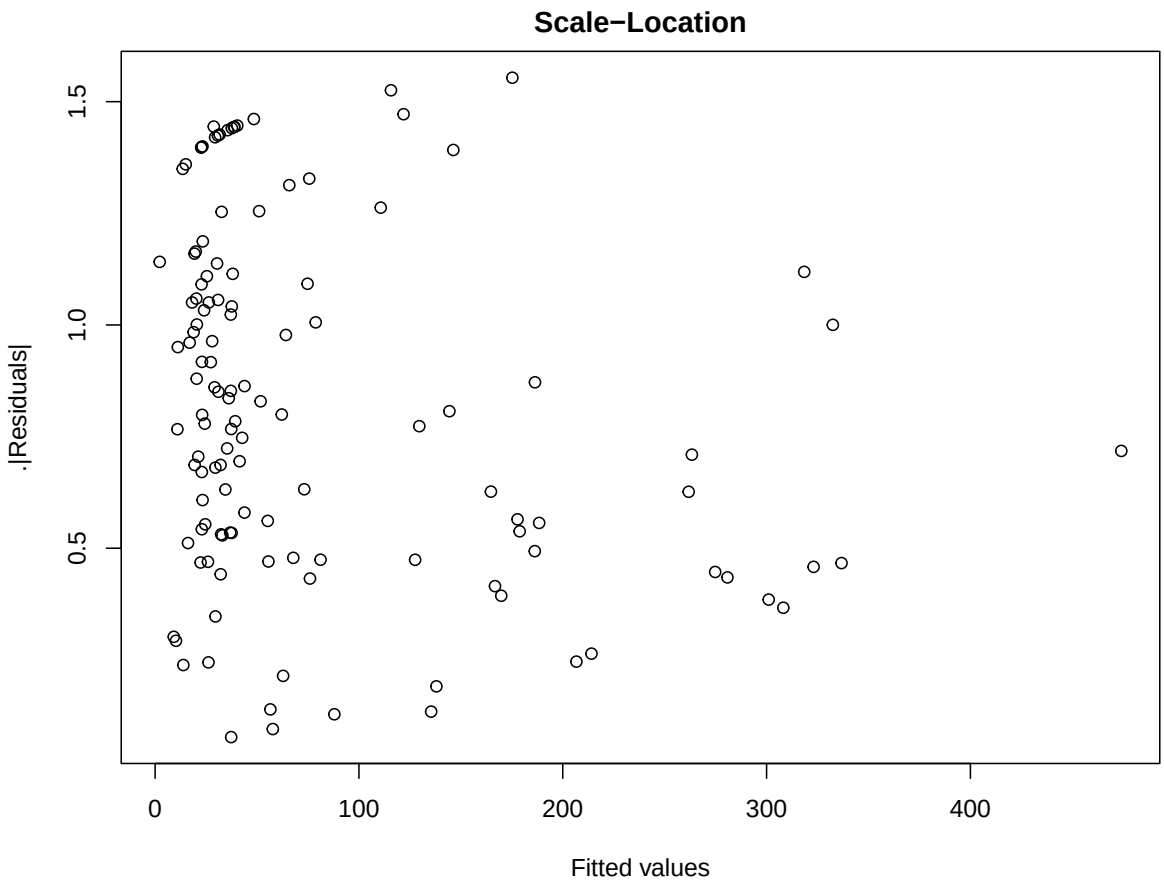


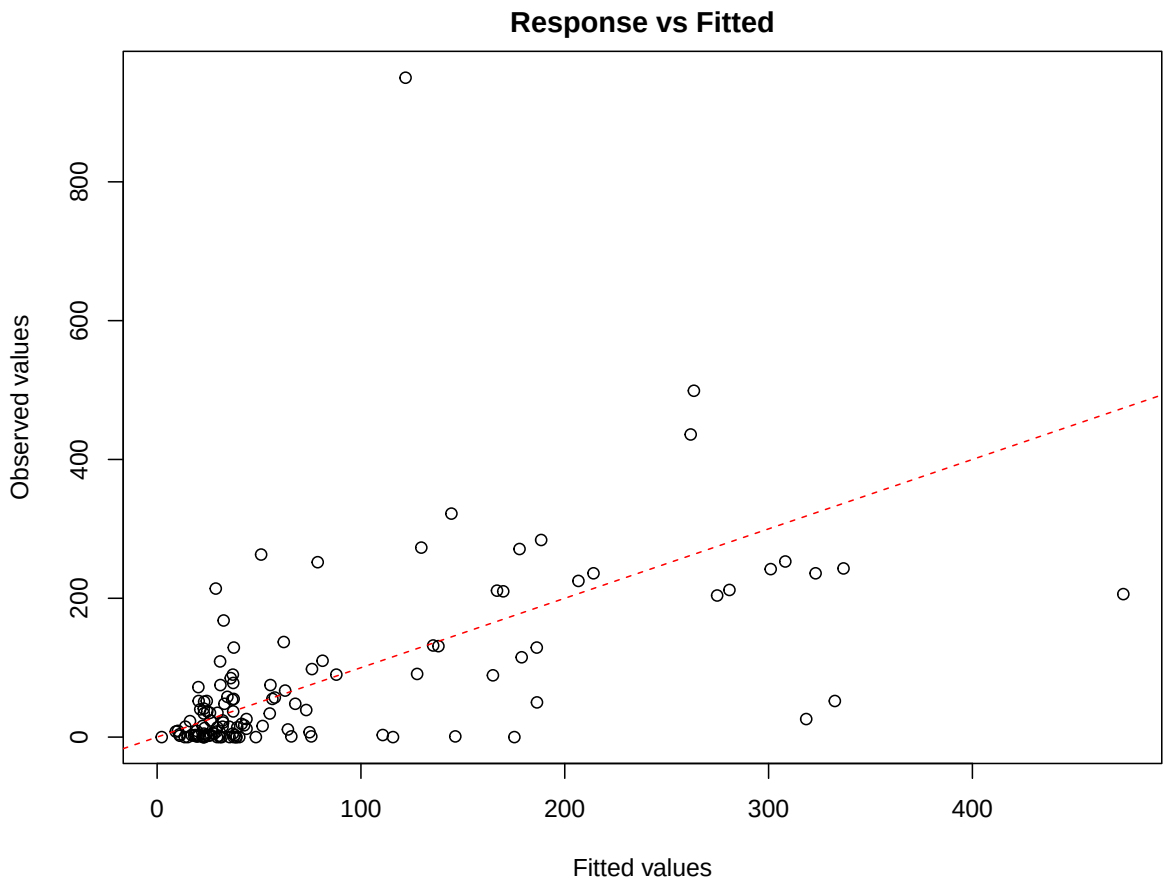

```
### Model 10 : 4b. Wind x Sunlight interaction

Model data dimensions after removing NAs: 125 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(0.496)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    sunlight_proportion) + s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)   3.8441     0.1355   28.36   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                            edf Ref.df Chi.sq  p-value
s(abundance_t_minus_1)                 2.176e+00  2.688 35.799 1.21e-06 ***
s(mean_wind_speed)                     2.655e+00  3.269  6.188  0.11445
s(mean_temp)                           1.000e+00  1.000  1.059  0.30352
s(sunlight_proportion)                 1.000e+00  1.001  6.908  0.00859 **
ti(mean_wind_speed,sunlight_proportion) 2.026e+00  2.469  2.797  0.28661
s(view_id)                             9.369e-05  8.000  0.000  0.71690
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.202   Deviance explained =   29%
-REML = 606.57  Scale est. = 1         n = 125


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
2
1
0
−1
−2
0
100
200
300
400

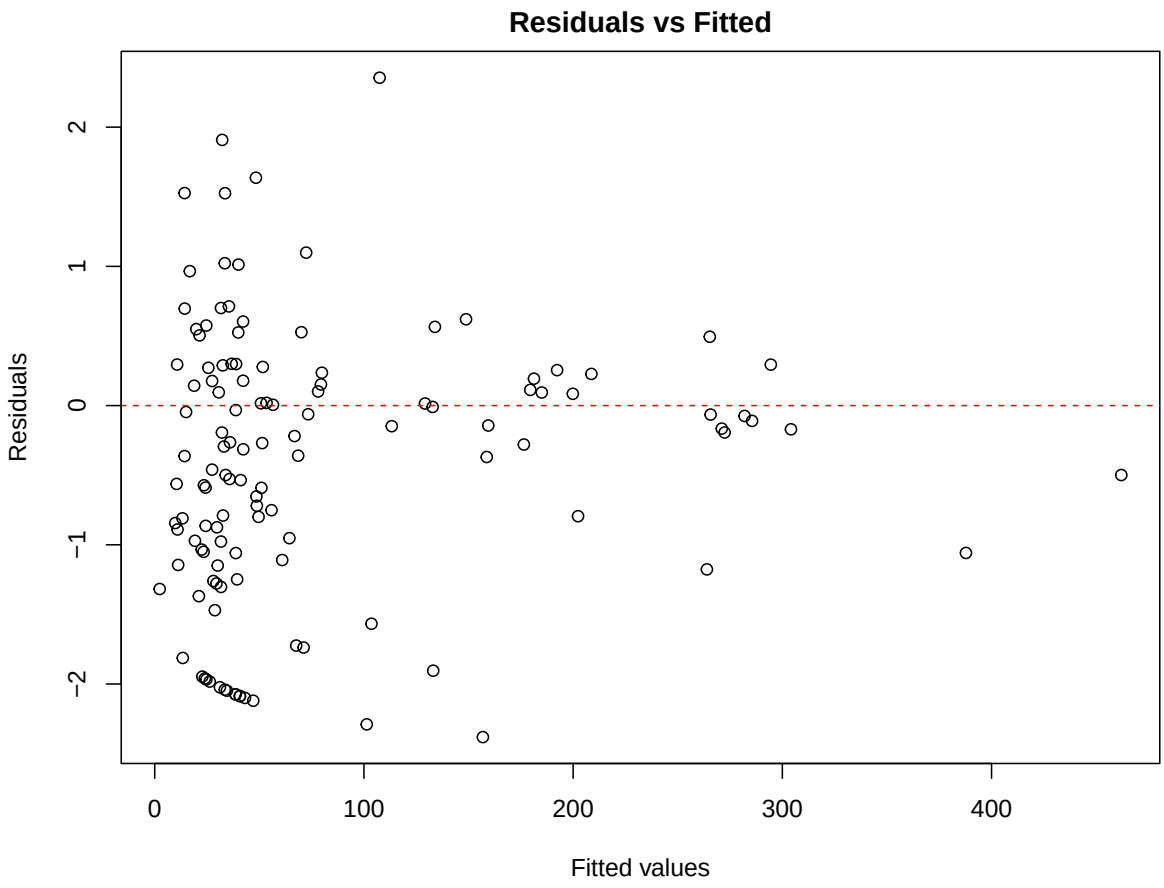

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−2
−1
0
1
2

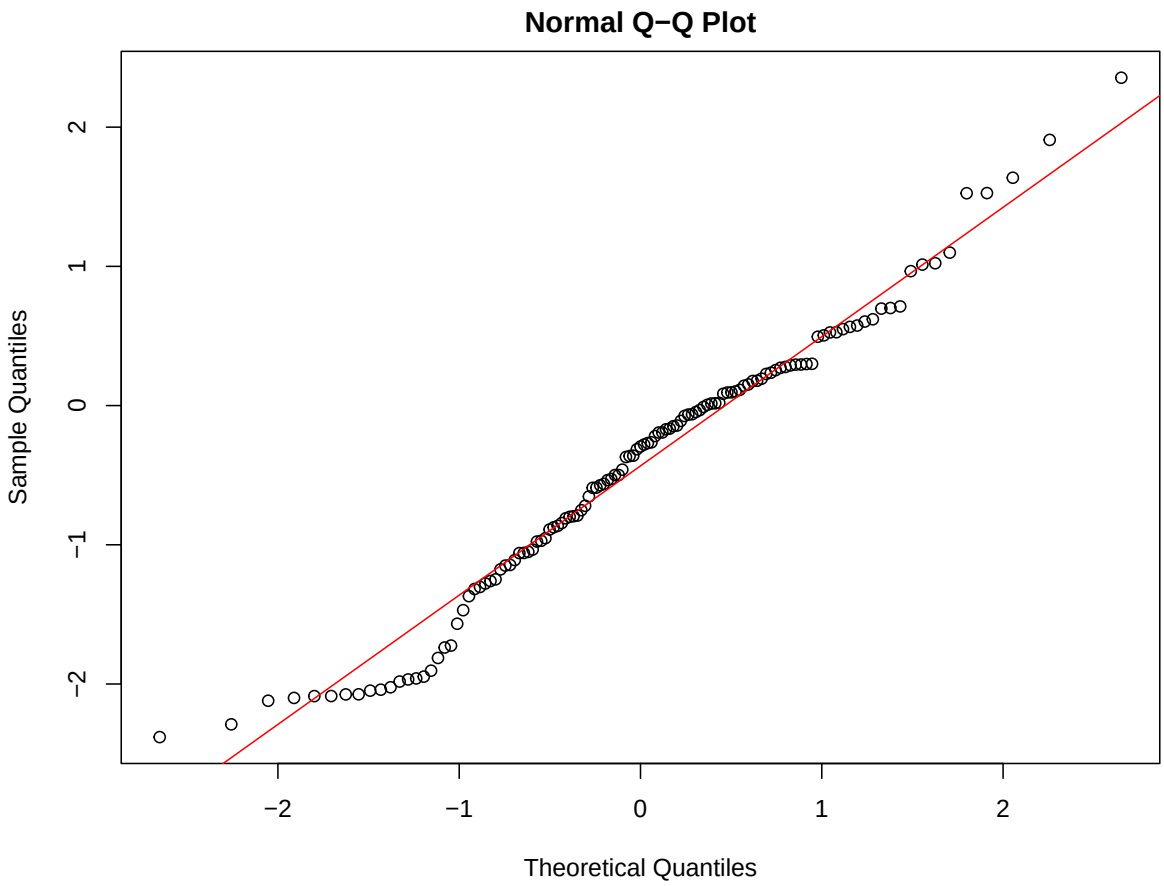

Scale−Location
√|Residuals|
0.5
1.0
1.5
0
100
200
300
400
Fitted values

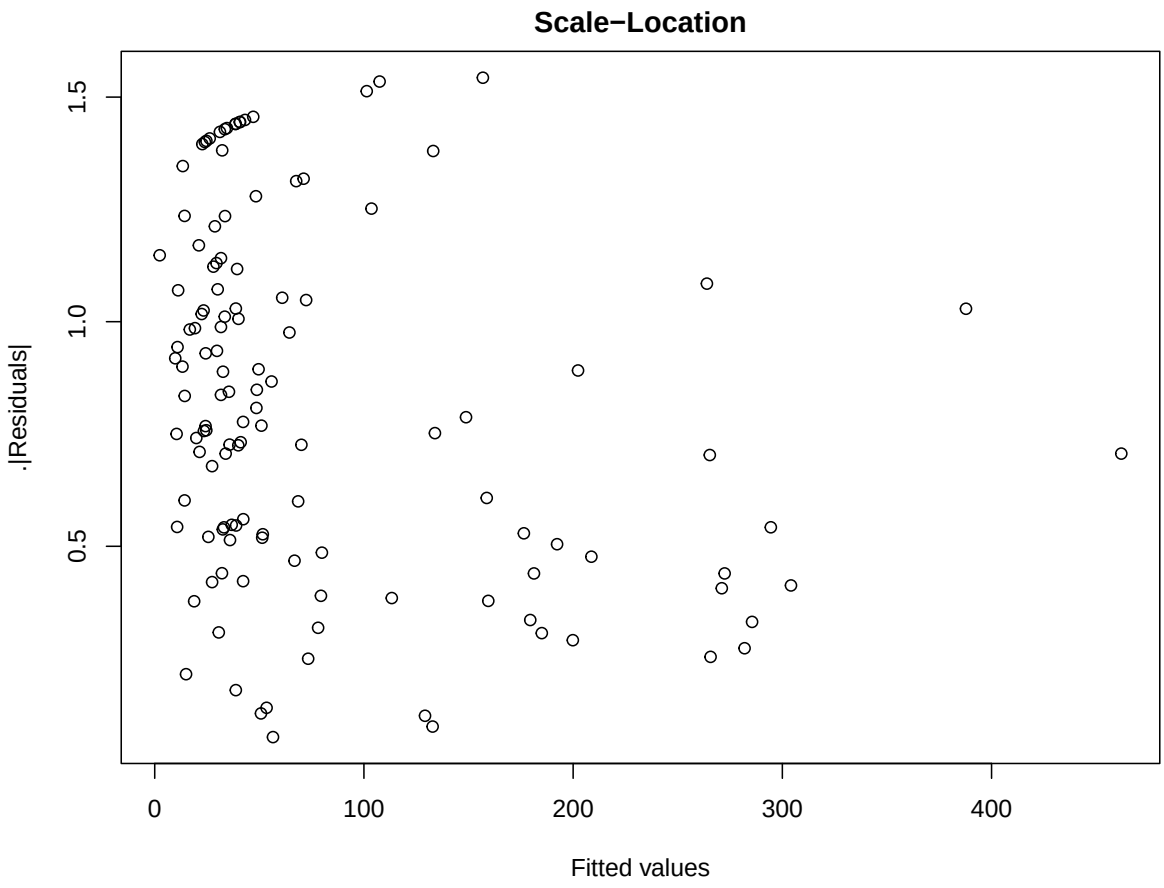


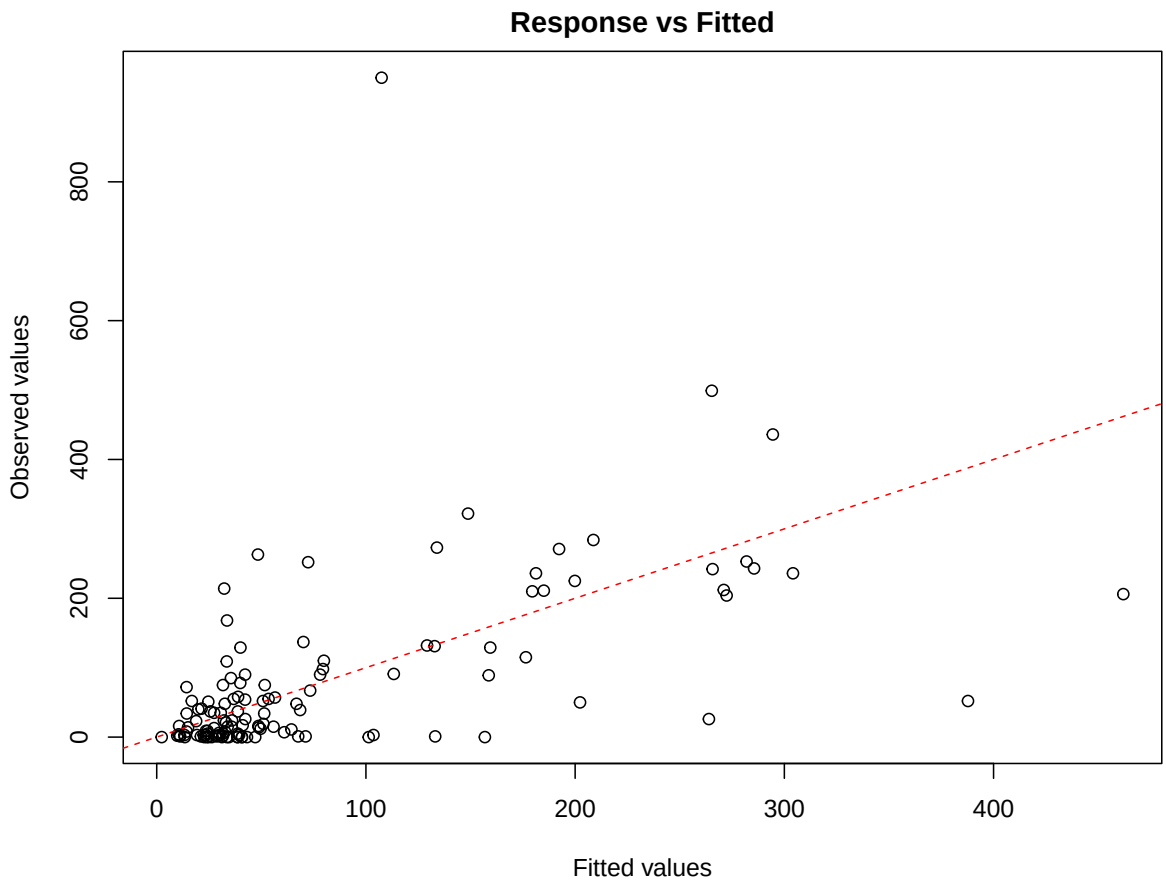

```
### Model 11 : 4c. Temp x Sunlight interaction

Model data dimensions after removing NAs: 125 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(0.492)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_temp, sunlight_proportion) +
    s(view_id, bs = "re")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)   3.9357     0.1293   30.44   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                       edf Ref.df Chi.sq  p-value
s(abundance_t_minus_1)           2.028e+00  2.514 28.898 2.58e-06 ***
s(mean_wind_speed)               2.534e+00  3.143  6.236   0.1058
s(mean_temp)                     1.000e+00  1.000  2.179   0.1400
s(sunlight_proportion)           1.000e+00  1.000  6.226   0.0126 *
ti(mean_temp,sunlight_proportion) 1.001e+00  1.001  3.560   0.0593 .
s(view_id)                       3.882e-05  8.000  0.000   0.6120
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.224   Deviance explained = 27.7%
-REML = 606.64  Scale est. = 1         n = 125


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
2
1
0
−1
−2
0
100
200
300
400

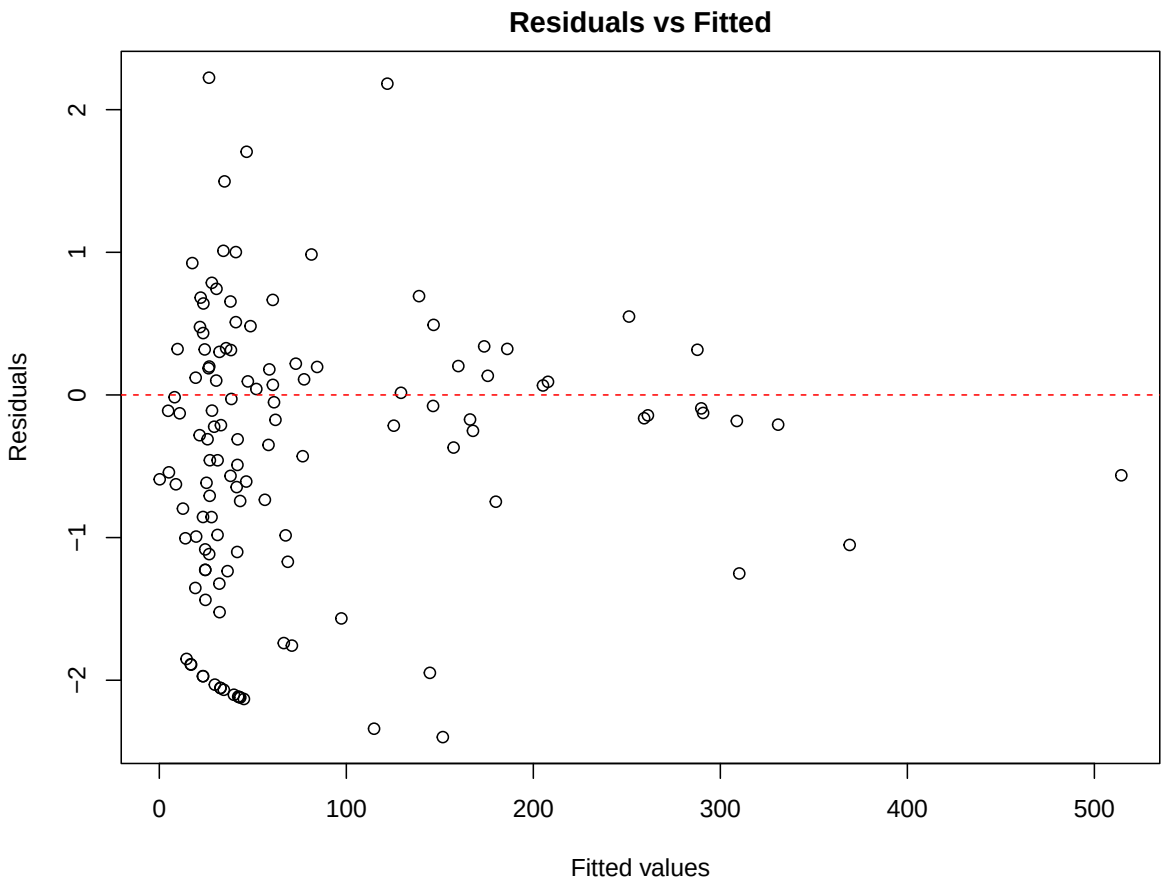

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−2
−1
0
1
2

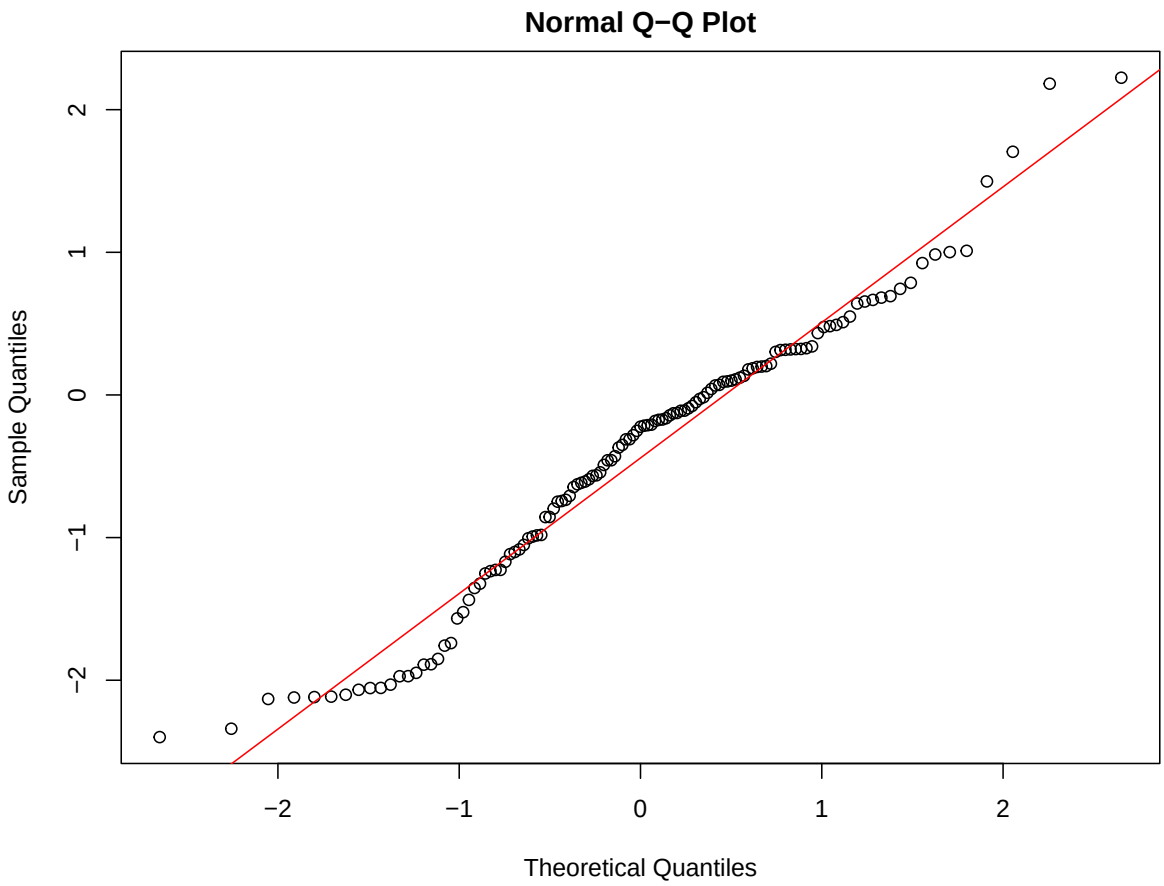

Scale−Location
.|Residuals|
0.5
1.0
1.5
0
100
200
300
400
Fitted values

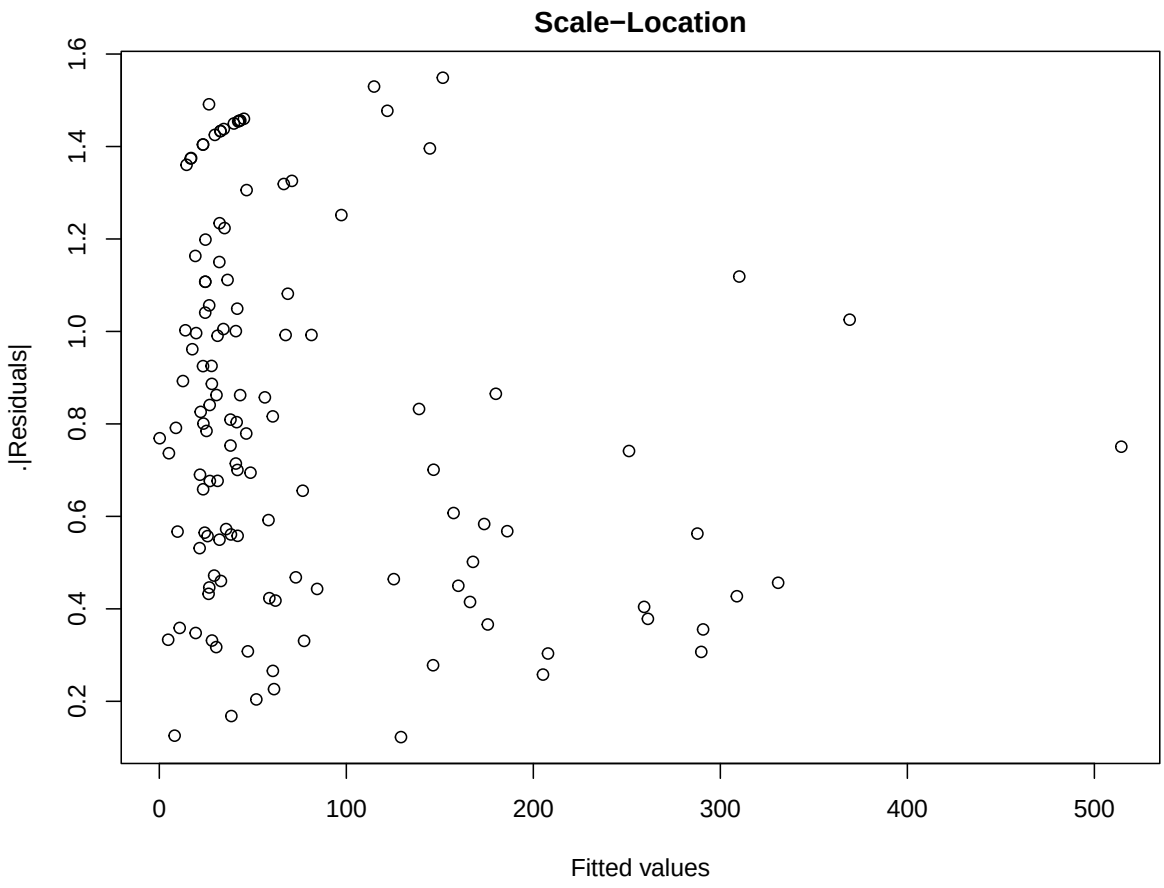


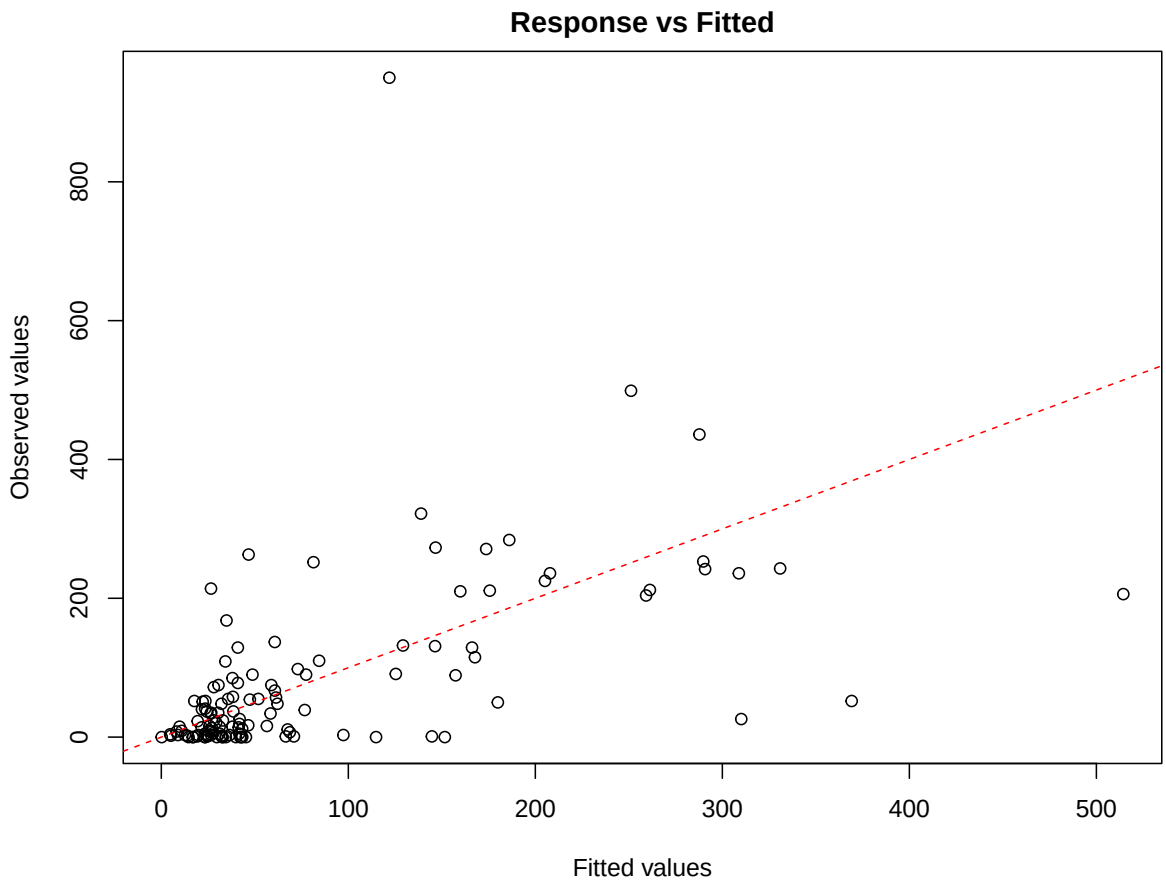

```
### Model 12 : 5a. Wind x Temp + Wind x Sun

Model data dimensions after removing NAs: 125 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(0.504)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + ti(mean_wind_speed, sunlight_proportion) + s(view_id,
    bs = "re")

Parametric coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.8424     0.1371   28.02   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                            edf Ref.df Chi.sq  p-value
s(abundance_t_minus_1)                2.0486540  2.538 30.468 1.56e-06 ***
s(mean_wind_speed)                    2.8229275  3.497  5.449  0.17638
s(mean_temp)                          1.0001581  1.000  0.645  0.42201
s(sunlight_proportion)                1.0002700  1.001  8.411  0.00374 **
ti(mean_wind_speed,mean_temp)         2.2005071  2.514  5.192  0.10797
ti(mean_wind_speed,sunlight_proportion) 1.0003820  1.001  5.305  0.02130 *
s(view_id)                            0.0001237  8.000  0.000  0.71287
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.189   Deviance explained = 30.7%
-REML = 603.89  Scale est. = 1         n = 125


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
2
1
0
−1
−2
0
100
200
300
400
500

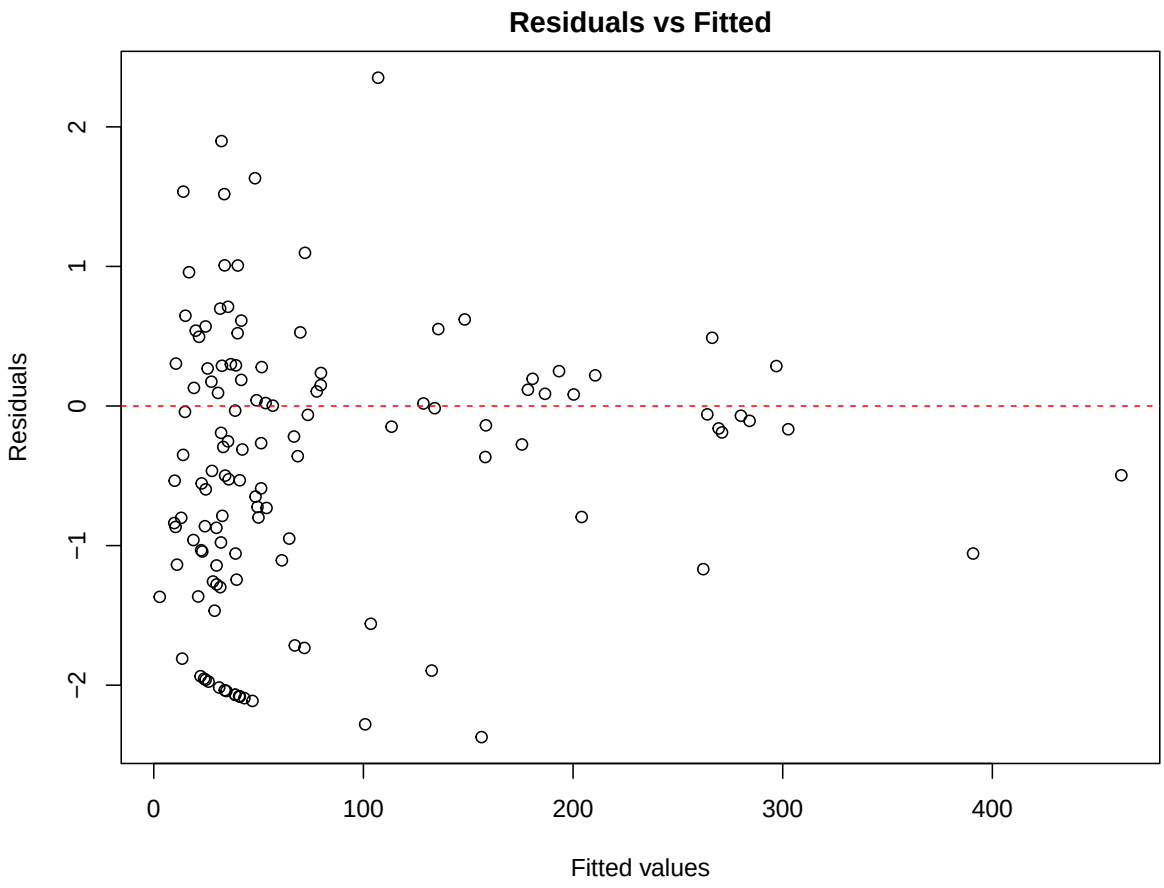

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−2
−1
0
1
2

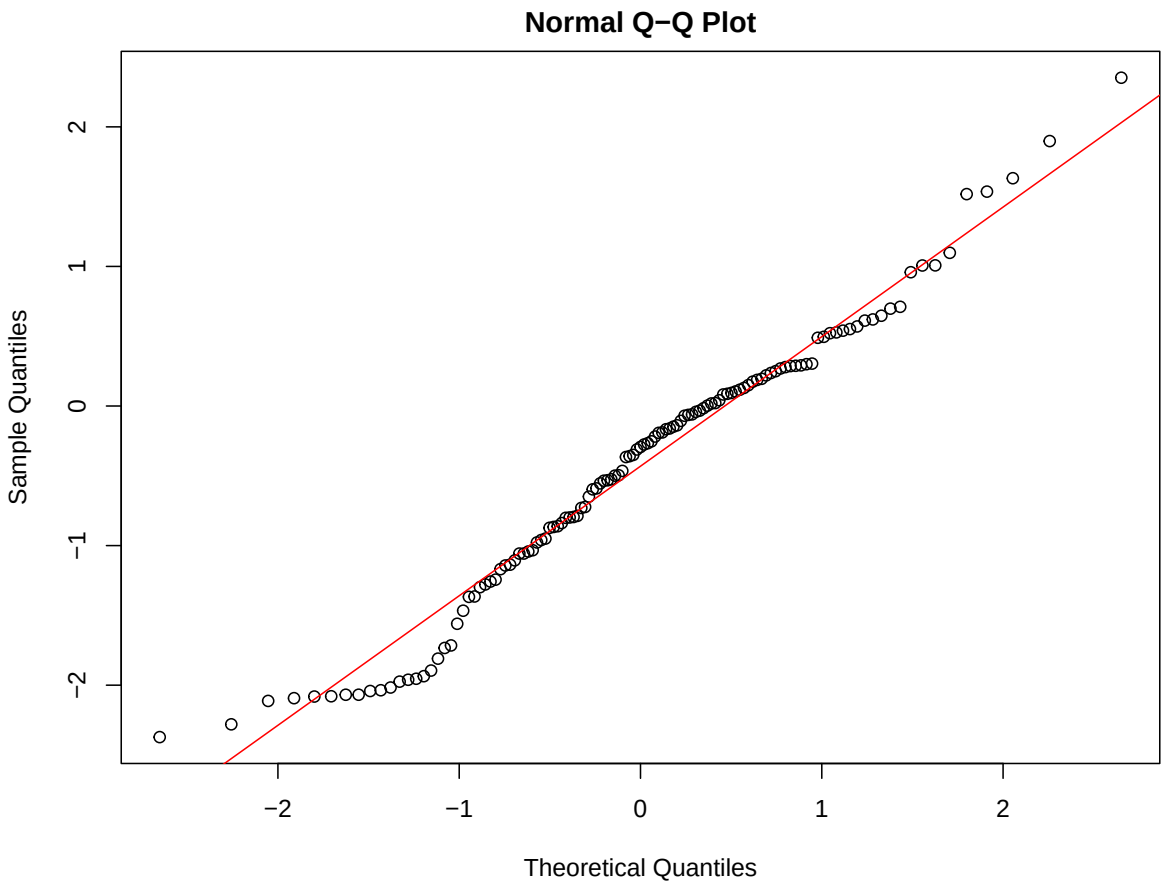

Scale−Location
√|Residuals|
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
0
100
200
300
400
500
Fitted values

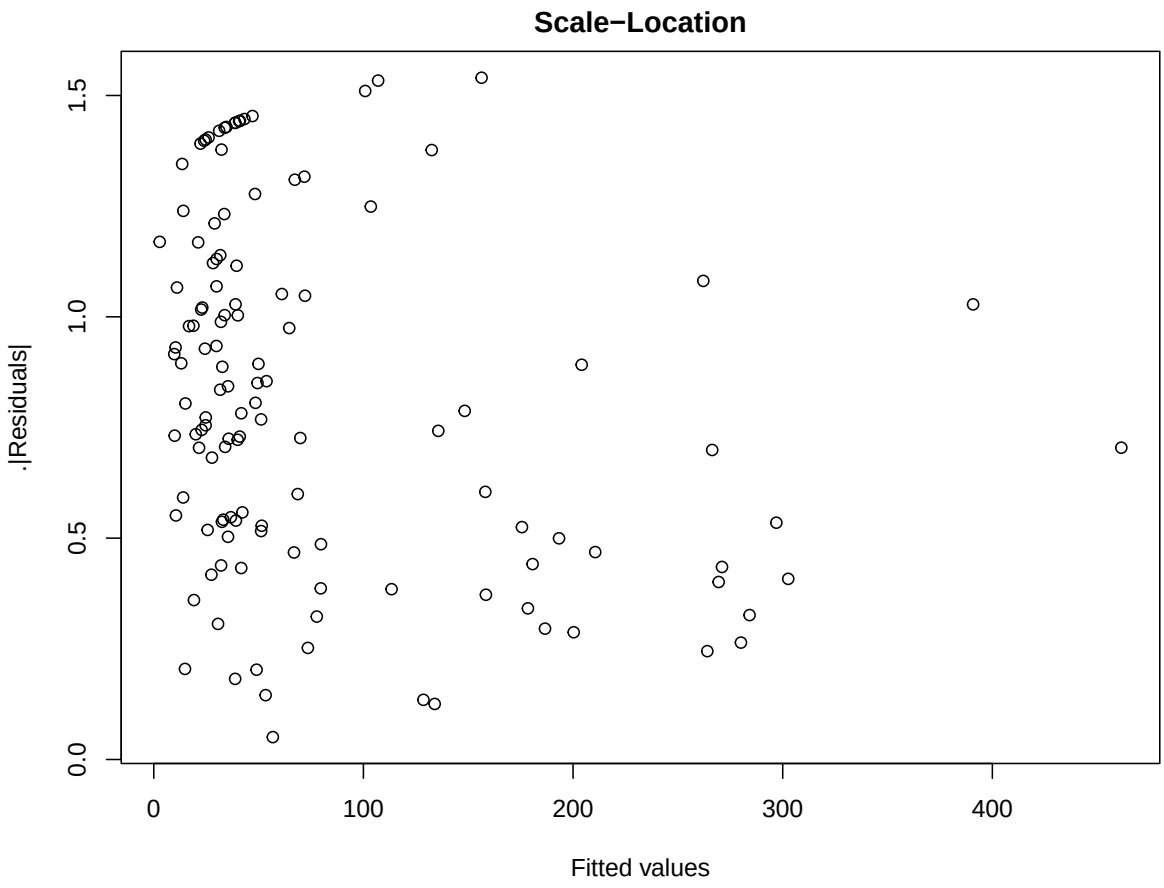


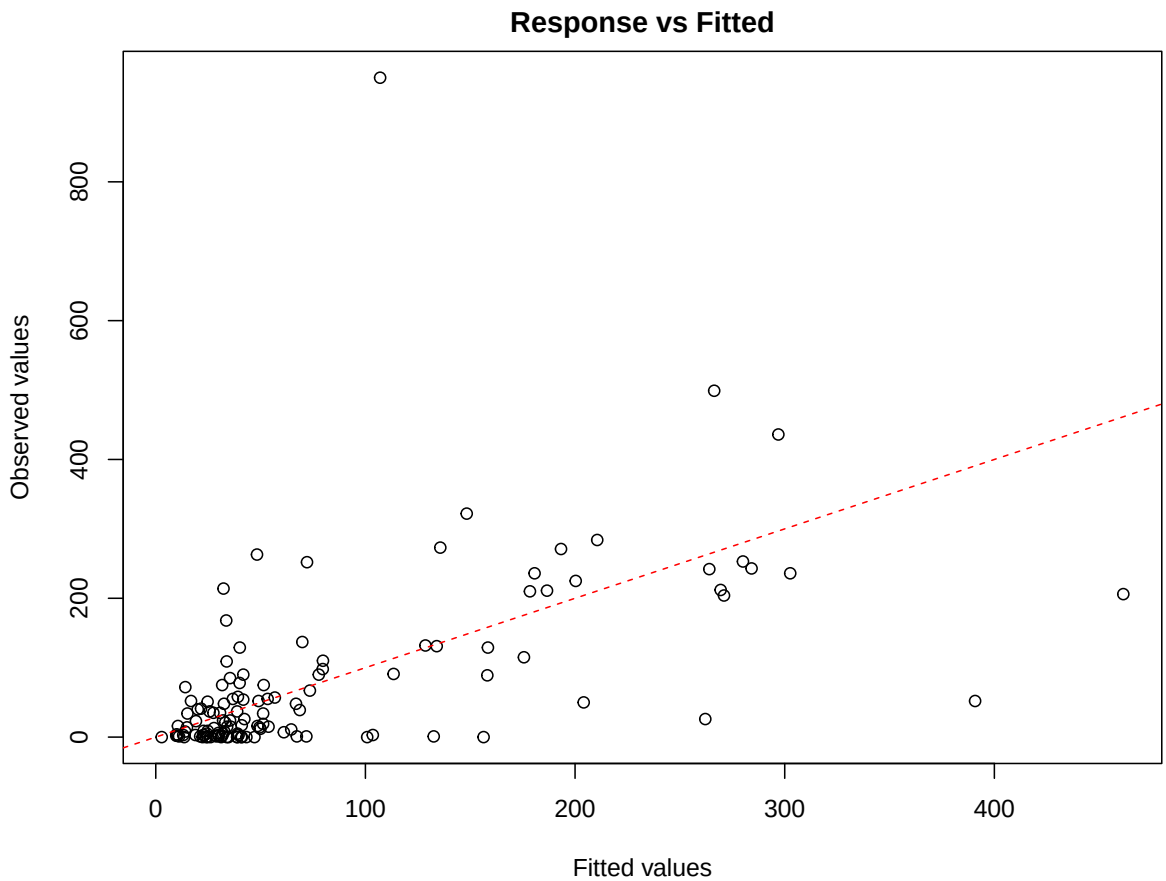

```
### Model 13 : 5b. Wind x Temp + Temp x Sun

Model data dimensions after removing NAs: 125 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(0.487)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + ti(mean_temp, sunlight_proportion) + s(view_id,
    bs = "re")

Parametric coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.9344     0.1301   30.24   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                       edf Ref.df Chi.sq  p-value
s(abundance_t_minus_1)           2.012e+00  2.495 27.320 6.14e-06 ***
s(mean_wind_speed)               2.484e+00  3.078  6.088   0.1092
s(mean_temp)                     1.000e+00  1.000  1.691   0.1934
s(sunlight_proportion)           1.000e+00  1.000  5.808   0.0160 *
ti(mean_wind_speed,mean_temp)    1.001e+00  1.002  0.014   0.9089
ti(mean_temp,sunlight_proportion) 1.000e+00  1.000  3.229   0.0724 .
s(view_id)                       7.648e-06  8.000  0.000   0.6108
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.218   Deviance explained = 27.6%
-REML = 604.69  Scale est. = 1         n = 125


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
2
1
0
−1
−2
0
100
200
300
400

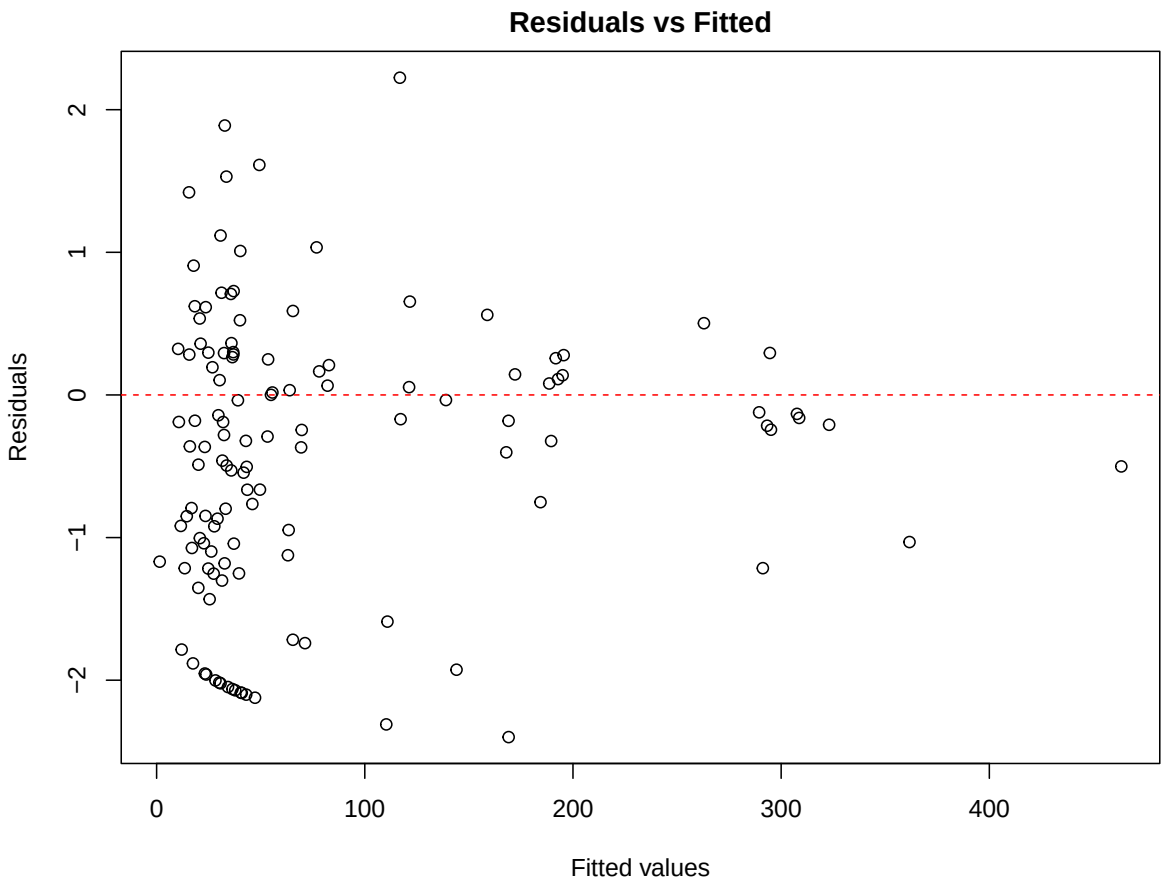

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−2
−1
0
1
2

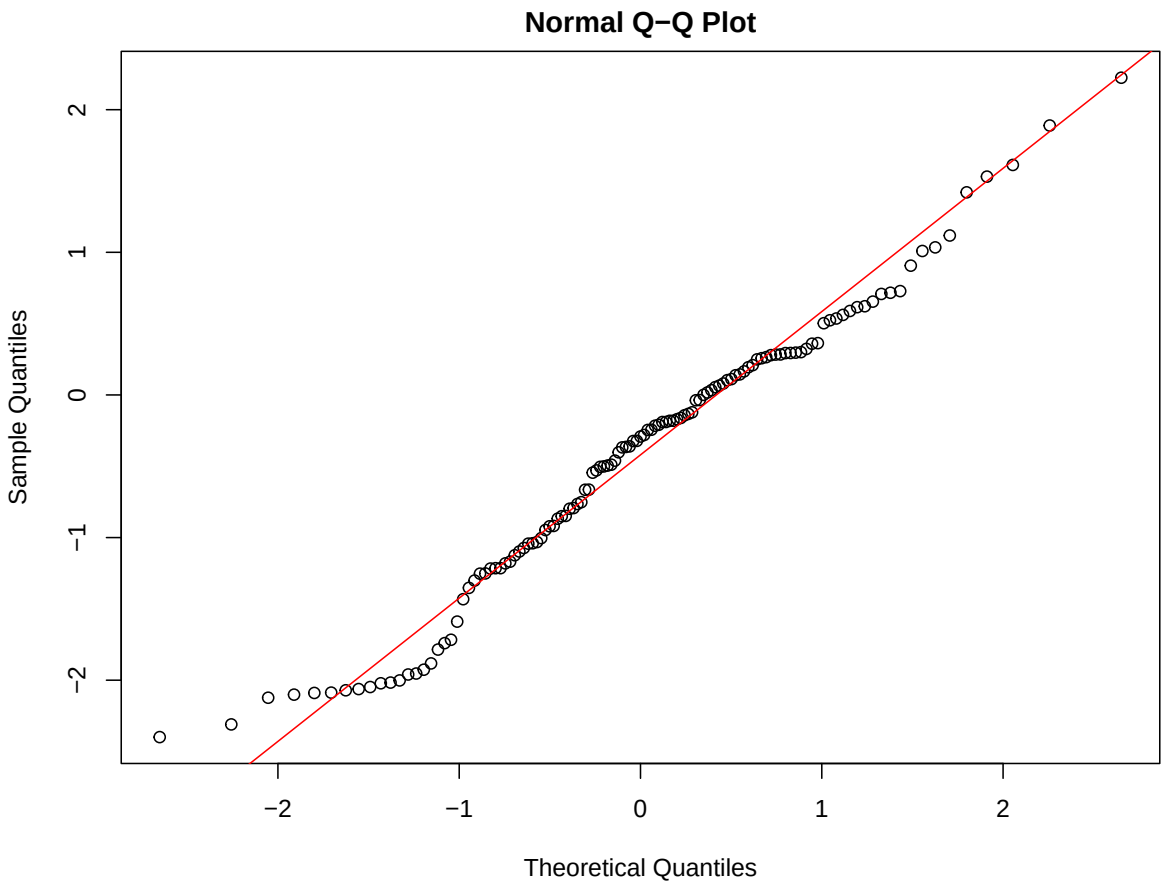

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
0
100
200
300
400
Fitted values

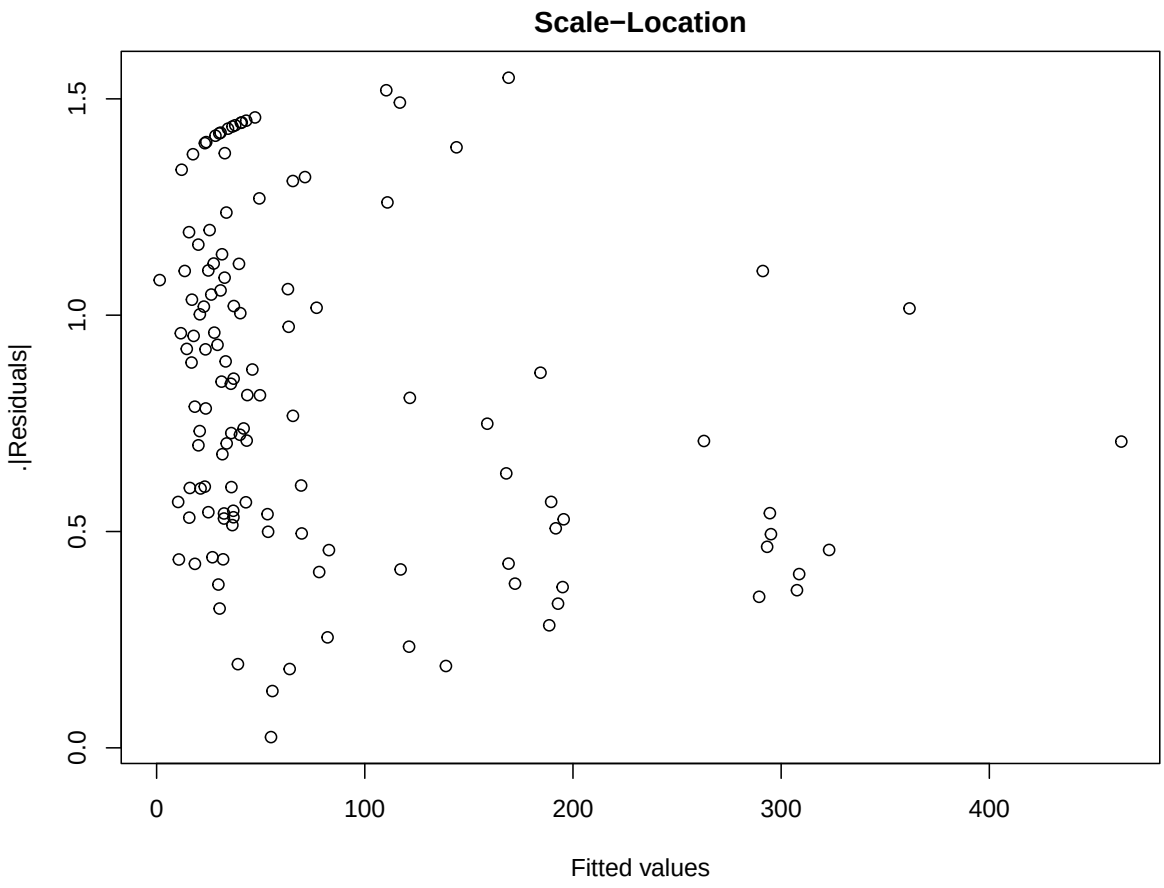


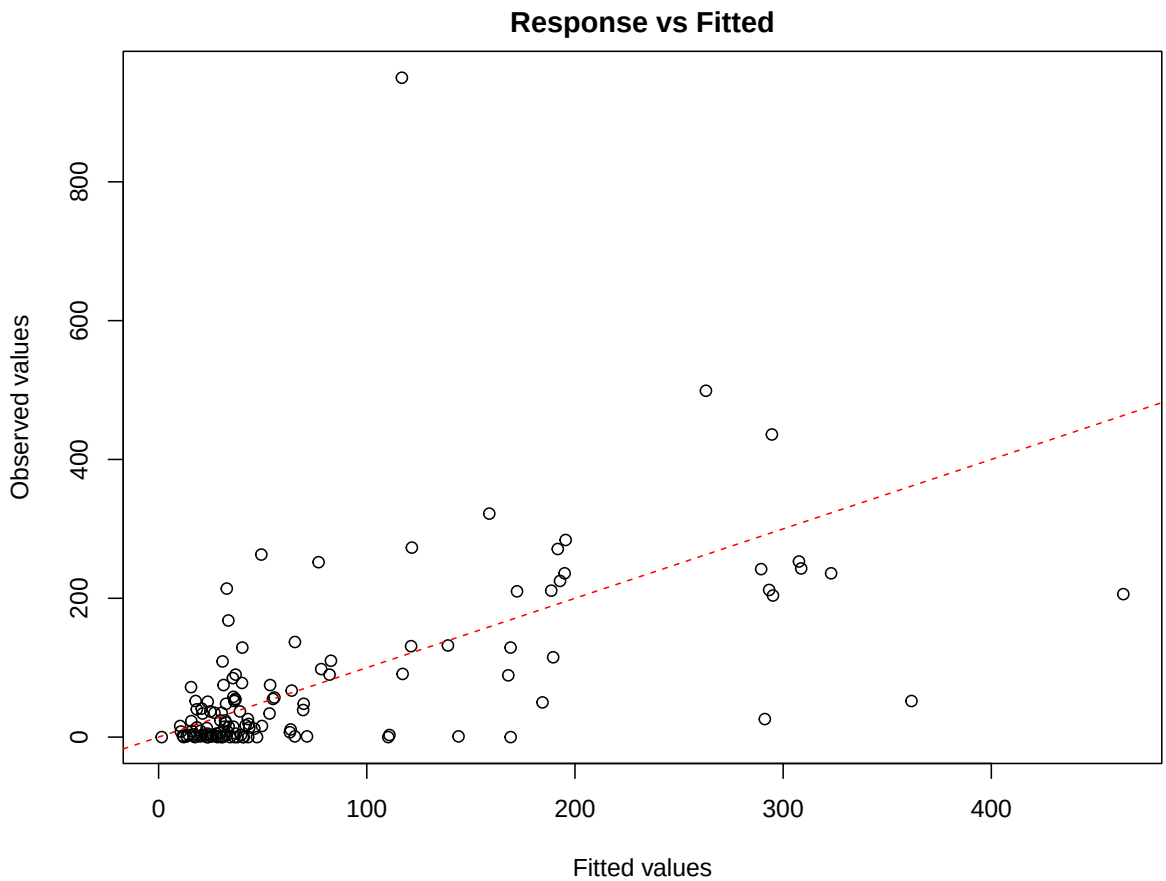

```
### Model 14 : 5c. Wind x Sun + Temp x Sun

Model data dimensions after removing NAs: 125 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(0.493)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    sunlight_proportion) + ti(mean_temp, sunlight_proportion) +
    s(view_id, bs = "re")

Parametric coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.8864     0.1343   28.94   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                            edf Ref.df Chi.sq p-value
s(abundance_t_minus_1)                2.114e+00  2.620 30.115 1.6e-06 ***
s(mean_wind_speed)                    2.645e+00  3.263  6.450 0.10268
s(mean_temp)                          1.000e+00  1.000  2.548 0.11046
s(sunlight_proportion)                1.000e+00  1.000  7.813 0.00519 **
ti(mean_wind_speed,sunlight_proportion) 1.001e+00  1.002  0.975 0.32403
ti(mean_temp,sunlight_proportion)     1.000e+00  1.001  1.637 0.20083
s(view_id)                            6.887e-05  8.000  0.000 0.68493
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.213   Deviance explained = 28.5%
-REML = 604.38  Scale est. = 1         n = 125


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
2
1
0
−1
−2
0
100
200
300
400

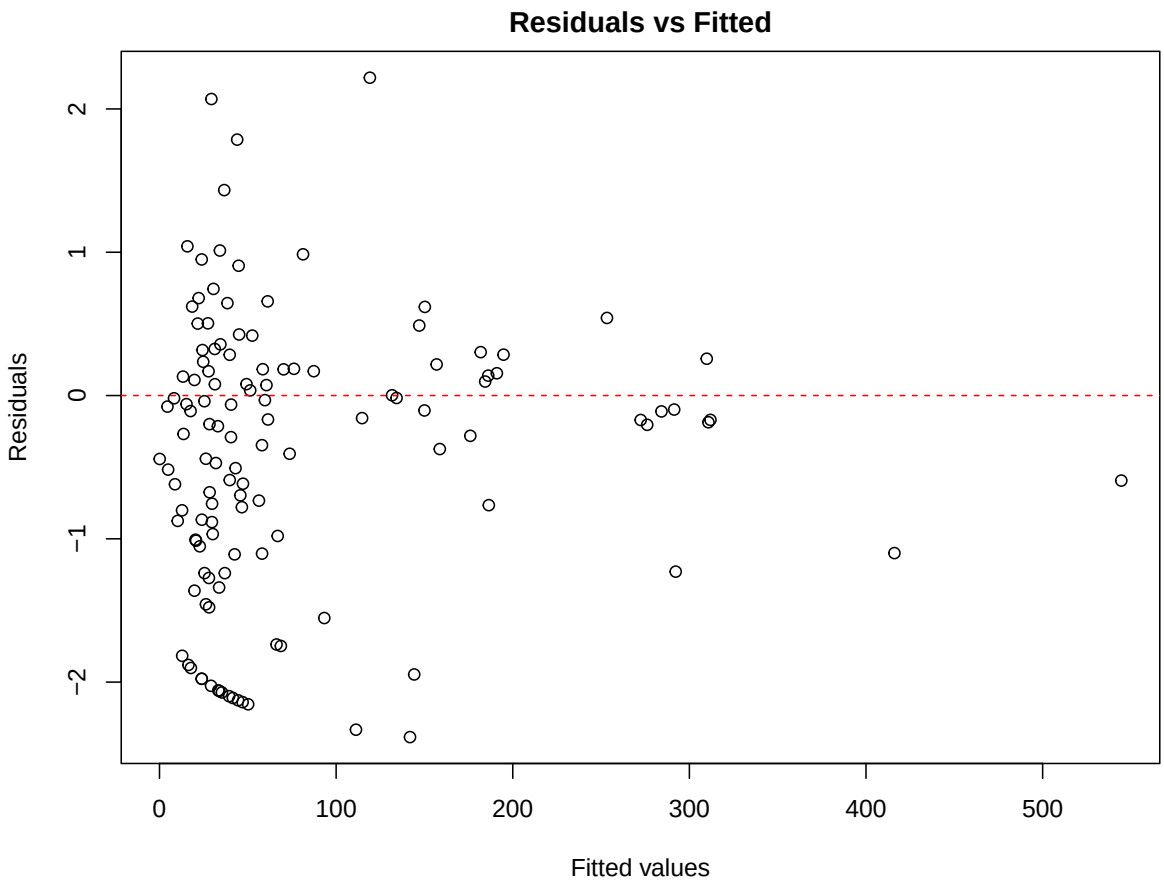

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−2
−1
0
1
2

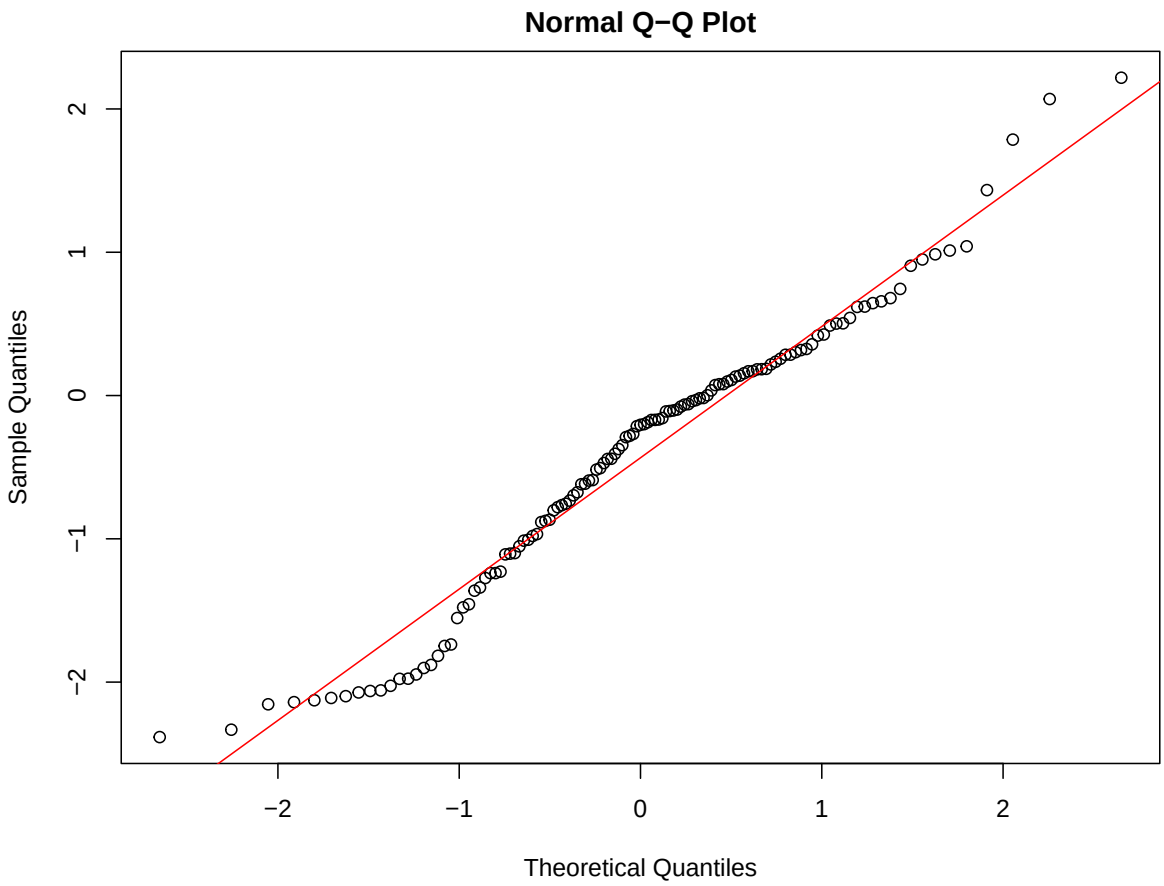

Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
0
100
200
300
400
Fitted values

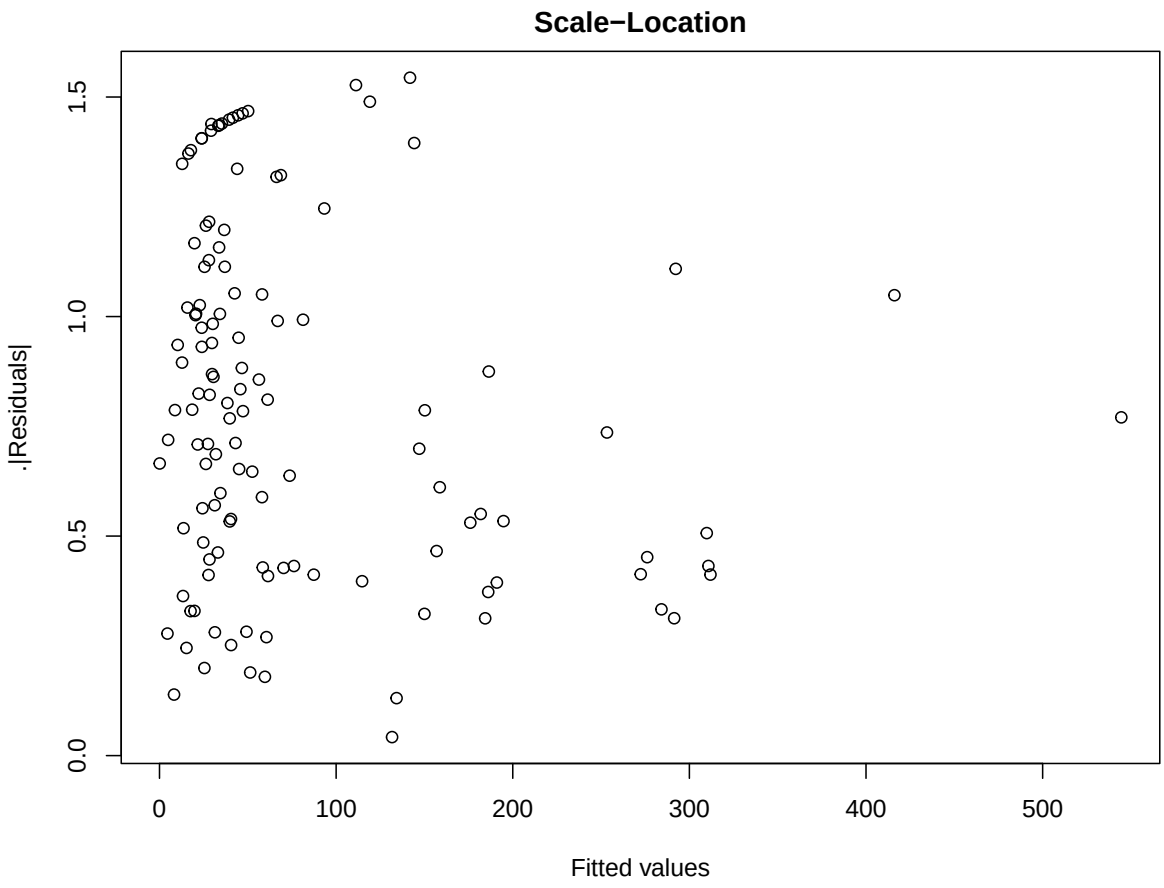


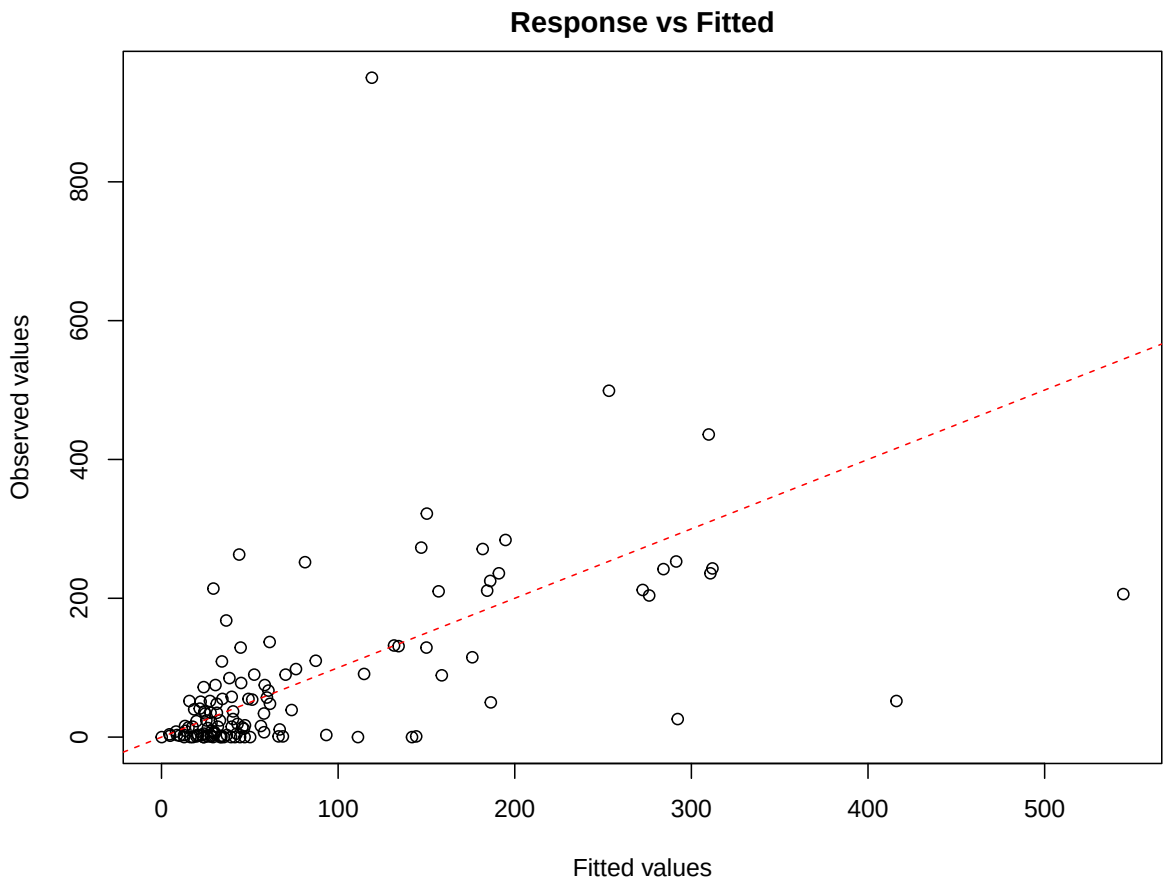

```
### Model 15 : 6. Global model

Model data dimensions after removing NAs: 125 x 6
GAM fitted successfully
Model Summary:

Family: Negative Binomial(0.503)
Link function: log

Formula:
total_butterflies ~ s(abundance_t_minus_1) + s(mean_wind_speed) +
    s(mean_temp) + s(sunlight_proportion) + ti(mean_wind_speed,
    mean_temp) + ti(mean_wind_speed, sunlight_proportion) + ti(mean_temp,
    sunlight_proportion) + s(view_id, bs = "re")

Parametric coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.8533     0.1394   27.65   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                            edf Ref.df Chi.sq  p-value
s(abundance_t_minus_1)                1.910e+00  2.373 25.877 8.47e-06 ***
s(mean_wind_speed)                    2.925e+00  3.618  5.628  0.17693
s(mean_temp)                          1.000e+00  1.000  1.509  0.21937
s(sunlight_proportion)                1.000e+00  1.000  9.660  0.00188 **
ti(mean_wind_speed,mean_temp)         2.233e+00  2.567  4.224  0.16456
ti(mean_wind_speed,sunlight_proportion) 1.000e+00  1.000  3.442  0.06358 .
ti(mean_temp,sunlight_proportion)     1.000e+00  1.000  0.798  0.37157
s(view_id)                            1.927e-05  8.000  0.000  0.67589
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.165   Deviance explained =   31%
-REML = 601.79  Scale est. = 1         n = 125


Diagnostic plots:
```


Residuals vs Fitted
Residuals
Fitted values
2
1
0
−1
−2
0
100
200
300
400
500


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−2
−1
0
1
2


Scale−Location
√|Residuals|
0.0
0.5
1.0
1.5
0
100
200
300
400
500
Fitted values




```r
# Model comparison table
comparison_full_day <- data.frame(
  Model = names(aic_values_full_day),
  AIC = round(aic_values_full_day, 2)
) %>%
  arrange(AIC)

cat("\n### Model Comparison (Full Day lag)\n")
```

```
### Model Comparison (Full Day lag)
```

```r
kable(comparison_full_day)
```

	Model	AIC
3b. Wind + Sunlight	3b. Wind + Sunlight	1218.64
2c. Sunlight only	2c. Sunlight only	1219.16
5a. Wind x Temp + Wind x Sun	5a. Wind x Temp + Wind x Sun	1220.02
2a. Wind only	2a. Wind only	1220.14
4c. Temp x Sunlight interaction	4c. Temp x Sunlight interaction	1220.32
3d. All additive	3d. All additive	1220.65
3c. Temp + Sunlight	3c. Temp + Sunlight	1220.93
1. Baseline	1. Baseline	1220.94
4b. Wind x Sunlight interaction	4b. Wind x Sunlight interaction	1221.11
5c. Wind x Sun + Temp x Sun	5c. Wind x Sun + Temp x Sun	1221.16
6. Global model	6. Global model	1221.31
3a. Wind + Temp	3a. Wind + Temp	1221.75
5b. Wind x Temp + Temp x Sun	5b. Wind x Temp + Temp x Sun	1222.31
4a. Wind x Temp interaction	4a. Wind x Temp interaction	1222.37
2b. Temperature only	2b. Temperature only	1223.01

Final Summary

```
cat("Analysis Complete\n")
```

```
Analysis Complete
```

```
cat("=================\n\n")
```

```
=================
```

```
# Combine all AIC comparisons
all_comparisons <- bind_rows(
  comparison_30m %>% mutate(Lag = "30 minutes"),
  comparison_60m %>% mutate(Lag = "1 hour"),
  comparison_120m %>% mutate(Lag = "2 hours"),
  comparison_240m %>% mutate(Lag = "4 hours"),
  comparison_full_day %>% mutate(Lag = "Full day")
) %>%
  select(Lag, Model, AIC) %>%
```

```r
  pivot_wider(names_from = Lag, values_from = AIC)

cat("AIC Values Across All Time Lags:\n")
```

```
AIC Values Across All Time Lags:
```

```r
kable(all_comparisons, digits = 2)
```

Model	30 minutes	1 hour	2 hours	4 hours	Full day
6. Global model	42783.35	45226.15	48104.98	50016.03	1221.31
5b. Wind x Temp + Temp x Sun	42795.36	45242.48	48148.72	50329.92	1222.31
5c. Wind x Sun + Temp x Sun	42820.34	45279.34	48132.33	50235.69	1221.16
4c. Temp x Sunlight interaction	42835.33	45301.73	48178.96	50353.42	1220.32
5a. Wind x Temp + Wind x Sun	42895.82	45322.71	48269.90	50199.82	1220.02
4a. Wind x Temp interaction	42907.80	45356.24	48342.66	50574.97	1222.37
4b. Wind x Sunlight interaction	42919.99	45364.55	48294.28	50425.97	1221.11
3d. All additive	42932.22	45393.15	48365.21	50586.55	1220.65
3c. Temp + Sunlight	42955.46	45423.55	48400.93	50656.47	1220.93
3b. Wind + Sunlight	43058.30	45692.15	48906.77	51273.00	1218.64
2c. Sunlight only	43076.43	45722.21	48958.71	51377.60	1219.16
3a. Wind + Temp	43126.63	45651.52	48761.57	51160.97	1221.75
2b. Temperature only	43137.36	45670.08	48785.36	51212.90	1223.01
2a. Wind only	43234.48	45916.93	49266.06	51779.13	1220.14
1. Baseline	43244.82	45936.70	49308.57	51872.69	1220.94

```r
# Find best model for each lag
best_models <- data.frame(
  Lag = c("30 minutes", "1 hour", "2 hours", "4 hours", "Full day"),
  Best_Model = c(
    comparison_30m$Model[1],
    comparison_60m$Model[1],
    comparison_120m$Model[1],
    comparison_240m$Model[1],
    comparison_full_day$Model[1]
  ),
  Best_AIC = c(
    comparison_30m$AIC[1],
    comparison_60m$AIC[1],
```

```r
    comparison_120m$AIC[1],
    comparison_240m$AIC[1],
    comparison_full_day$AIC[1]
  )
)

cat("\n\nBest Model for Each Time Lag:\n")
```

```
Best Model for Each Time Lag:
```

```r
kable(best_models, digits = 2)
```

Lag	Best_Model	Best_AIC
30 minutes	6. Global model	42783.35
1 hour	6. Global model	45226.15
2 hours	6. Global model	48104.98
4 hours	6. Global model	50016.03
Full day	3b. Wind + Sunlight	1218.64