

Butterfly Monitoring Analyses

Reto Schmucki

2024-07-19

Table of contents

Preface	1
NOTE - this book project is a work in progress that only started.	1
 I Counting Butterflies	 3
Butterfly Monitoring	5
 II From Counts to FLight Curves	 7
Butterfly Counts	9
Simulation	9
Non Gaussian flight curve	16
Simple trend case	19

Preface

This project aims to collate documentation of methods used to analyze and work with Butterfly Monitoring Schemes (BMS) data. We focus on methods used to analyze count data associated with Pollard transects that represent the core of the eBMS database.

**NOTE - this book project is a work in progress
that only started.**

Reto Schmucki, July 2024

Part I

Counting Butterflies

Butterfly Monitoring

Content under construction

Part II

From Counts to FLight Curves

Butterfly Counts

Simulation

To demonstrate and assess the method used to compute butterfly abundance indices and the Grassland Butterfly Indicator which involves calculating collated abundance indices from multiple sites and estimating the population trend from a set of indicator species, the best option is to use simulated datasets that provide realistic data with known parameters. The simulation approach allows us to assess the methods' performance and enables us to control individual parameters and conduct rigorous sensitivity analysis. This will provide useful insight into the method and enable rigorous assessment of its power and limitations. To simulate butterfly count across sites and over multiple iterations, we will use the function developed by Collin Edwards and available in the [R packages butterflyGamSims](#). We will use the simulation to test the GAM method implemented in the [R package rbms](#).

```
if(!require("data.table")) install.packages("data.table")
if(!require("ggplot2")) install.packages("ggplot2")
if(!require("devtools")) install.packages("devtools")
if(!require("rbms")) devtools::install_github("RetoSchmucki/rbms")
if(!require("butterflyGamSims")) devtools::install_github("cbedwards/butterflyGamSims")

flc_col <- 'orange'
cnt_col <- 'darkcyan'
missing_col <- 'red'
GAM_col <- '#483d8b'
```

In the first case, we will simulate counts for one site for one year, using a Gaussian curve to depict the adult butterflies' seasonal phenology (activity curve). The daily counts are generated from a random Poisson process.

```
btfl_data <- timeseries_sim(nsims=1,
  year = c(2023:2025),
```

```

doy.samples = seq(from=1, to=365, by=1),
abund.type = "exp",
activity.type = "gauss",
sample.type = "pois",
sim.parms = list(growth.rate = 0,
                  init.size = 500,
                  act.mean = 175,
                  act.sd = 15)
)

sim2bms <- function(data, yearKeep = NULL, weeklySample = FALSE, weekdayKeep = NULL,

  btfl_ts <- data.table::data.table(data)[, site_id := paste0("site_",
  if(!is.null(yearKeep)){
    btfl_ts <- btfl_ts[years %in% yearKeep, ]
  }
  btfl_ts[, date := as.Date(doy, origin = paste0(years, "-01-01"))]
  btfl_ts[, week := isoweek(date)]
  btfl_ts[month(date) != 1 | week < 50, weekday := rowid(week), by =
  if(isTRUE(weeklySample)){
    if(!is.null(weekdayKeep)){
      btfl_ts <- btfl_ts[weekday %in% weekdayKeep, ]
    }
    btfl_ts <- btfl_ts[btfl_ts[,.I[sample(.N, 1)]], by = .(week,
  }
  if(!is.null(monitoredSeason)){
    btfl_ts <- btfl_ts[month(date) %in% monitoredSeason, ]
  }
  return(btfl_ts)
}

y <- c(2023)

btfl_ts <- sim2bms(data = btfl_data$timeseries, yearKeep = y)

btfl_fig <- ggplot() +
  geom_point(data=btfl_ts, aes(x=doy, y=count, colour = "count")) +
  geom_line(data = btfl_ts,
  aes(x = doy, y = act, colour = "activity")) +
  xlim(1,365) + ylim(0, max(btfl_ts$count, btfl_ts$act)) +
  scale_colour_manual("",
    breaks = c("count", "activity"),
    values = c(cnt_col, flc_col)) +
  theme_light() +

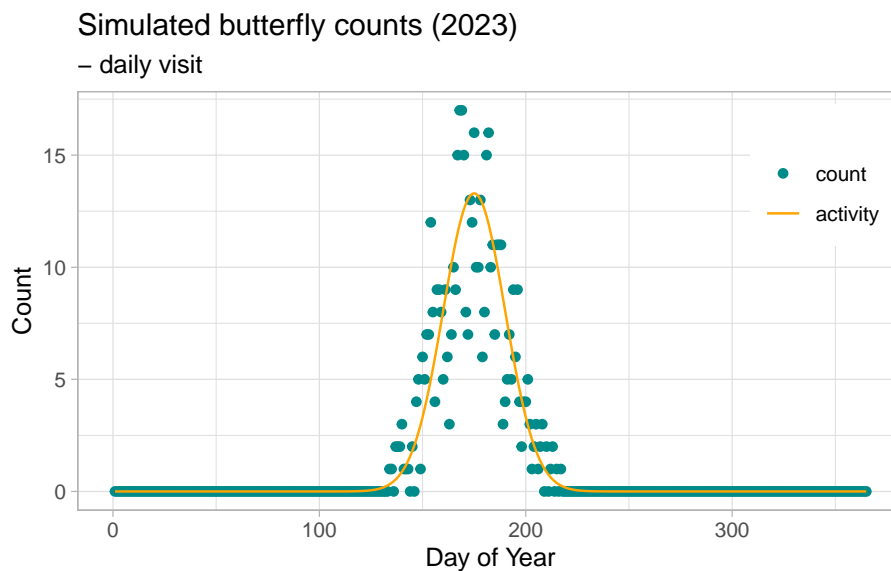
```

```

theme(legend.position = "inside", legend.position.inside = c(0.9, 0.8)) +
labs(title = paste0("Simulated butterfly counts (", y, ")"),
     subtitle = "- daily visit",
     x = "Day of Year",
     y = "Count")

btfl_fig

```



In the example above, the activity curve represented by the line has a Gaussian shape and counts presented by the points along the curve are independent random samples from a Poisson distribution. Because we sampled a count value for 365 days (day-of-year; doy), the counts are representative of the population of active adult butterflies as if the site was visited every. This implies that a proportion of butterflies are counted more than one day as their lifespan exceeds one day. On Pollard transect, this is how butterfly counts are likely to be counted and reported, but with a different frequency as visits are generally weekly, fortnightly, or even monthly. We can replicate this value by resampling the daily count weekly.

```

set.seed(13276)
btfl_week_smpl <- sim2bms(data = btfl_data$timeseries, yearKeep = y,
                          weeklySample = TRUE,
                          weekdayKeep = c(2:5),
                          monitoringSeason = c(4:9))

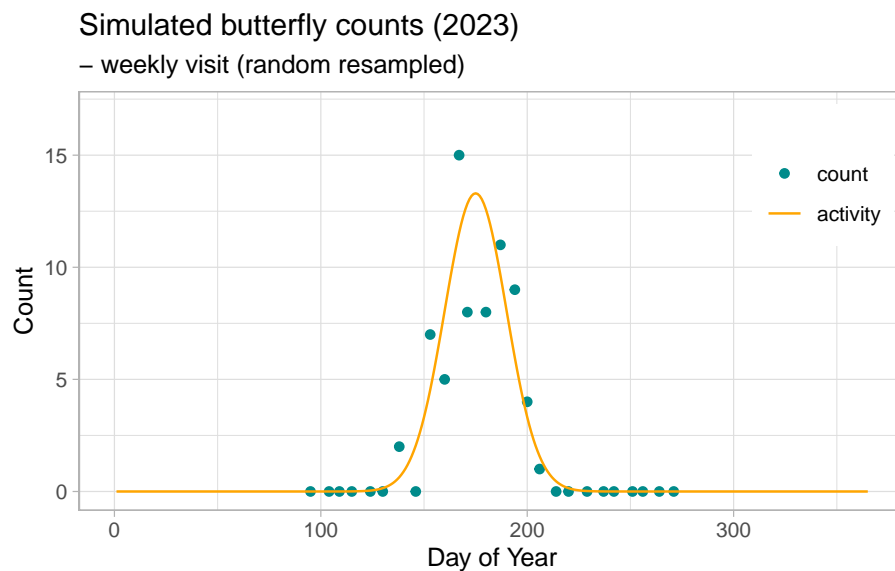
btfl_fig2 <- ggplot() +

```

```

geom_point(data=btfl_week_smpl, aes(x=doy, y=count, colour = "count"),
geom_line(data = btfl_ts,
aes(x = doy, y = act, colour = "activity")) +
xlim(1,365) + ylim(0, max(btfl_ts$count, btfl_ts$act)) +
scale_colour_manual("",
breaks = c("count", "activity"),
values = c(cnt_col, flc_col)) +
theme_light() +
theme(legend.position = "inside", legend.position.inside = c(0.9, 0.9)),
labs(title = paste0("Simulated butterfly counts (", y, ")"),
subtitle = "- weekly visit (random resampled)",
x = "Day of Year",
y = "Count")
btfl_fig2

```



From the weekly visits, counts outside of the monitoring period will not be informed, in many cases these are ‘zeros’ as we expect the monitoring season to align with butterflies’ activity. Some other weeks might be missing from the time series, potentially due to unsuitable weather conditions for monitoring or recorder’s unavailability. We can include these missing visits by resampling a subset of the weekly counts.

```

missing_prob <- function(data, mu=NULL, alpha = 5, theta = 0.3){
  x_ <- seq_len(nrow(data))
  mu_ <- ifelse(is.null(mu), length(x_) / 2, mu)

```



```

        std_ <- sqrt(mu_ / theta)
        y_ <- abs((alpha * exp(-(x_ - mu_)^2) / std_^2)) - alpha) + alp
        yn_ <- y_ / (sum(y_))
        return(yn_)
    }

sample_missing <- function(data, propMissing = 0.25){

    missing.prob <- data.table::data.table()
    for(i in data[, unique(years)]){
        for(j in data[, unique(site_id)]){
            missing.prob <- rbind(missing.prob, missing_prob(data[years == i & site_id =
        })
    }

    missing.week <- data[sample(seq_len(.N), round(propMissing * .N), prob = unlist(mi

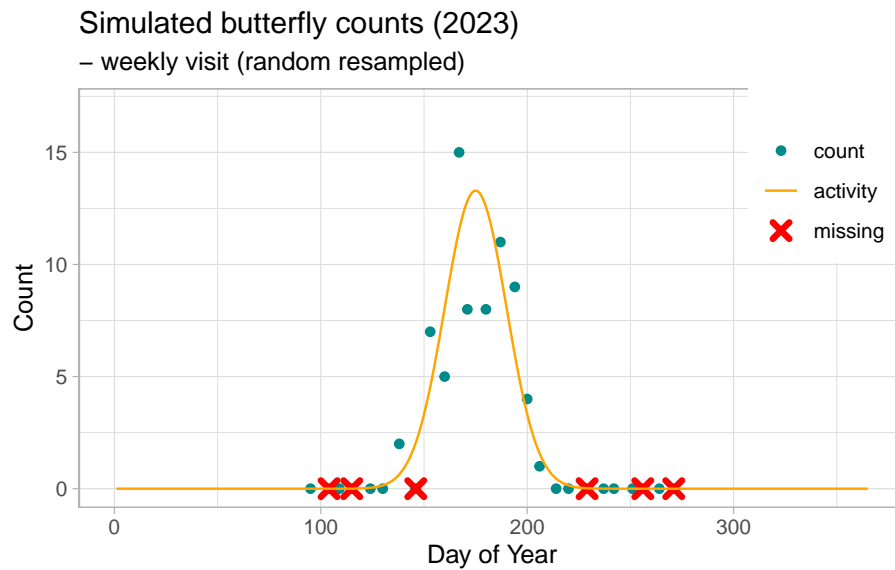
    return(missing.week)
}

btfl_week_missing <- sample_missing(data = btfl_week_smpl, propMissing = 0.25)

btfl_fig3 <- ggplot() +
    geom_point(data=btfl_week_smpl, aes(x=doy, y=count, colour = "count")) +
    geom_point(data=btfl_week_missing, aes(x=doy, y=count, colour = "missing"),
               shape=4, size=2, stroke=2) +
    geom_line(data = btfl_ts,
              aes(x = doy, y = act, colour = "activity")) +
    xlim(1,365) + ylim(0, max(btfl_ts$count, btfl_ts$act)) +
    scale_colour_manual("",
                        breaks = c("count", "activity", "missing"),
                        values = c(cnt_col, flc_col, missing_col)) +
    theme_light() +
    theme(legend.position = "inside", legend.position.inside = c(0.9, 0.8)) +
    labs(title = paste0("Simulated butterfly counts (", y, ")"),
         subtitle = "- weekly visit (random resampled)",
         x = "Day of Year",
         y = "Count")

btfl_fig3

```



Because recorders only report the number of observed butterflies, zeros are generally not reported but can be derived from the visit dates.

```
visit_sim <- btfl_week_smpl[!date %in% btfl_week_missing$date, .(site_id, date, count)]
count_sim <- visit_sim[count>=1,][, species := "sp1"]

names(visit_sim) <- toupper(names(visit_sim))
names(count_sim) <- toupper(names(count_sim))

ts_date <- rbms::ts_dwmy_table(InitYear = 2023, LastYear = 2023, WeekDay1 = 'monday')

ts_season <- rbms::ts_monit_season(ts_date,
  StartMonth = 4,
  EndMonth = 9,
  StartDay = 1,
  EndDay = NULL,
  ComplSeason = TRUE,
  Anchor = TRUE,
  AnchorLength = 2,
  AnchorLag = 2,
  TimeUnit = 'd')

ts_season_visit <- rbms::ts_monit_site(ts_season, visit_sim)

ts_season_count <- rbms::ts_monit_count_site(ts_season_visit, count_sim, sp = "sp1")
```

```

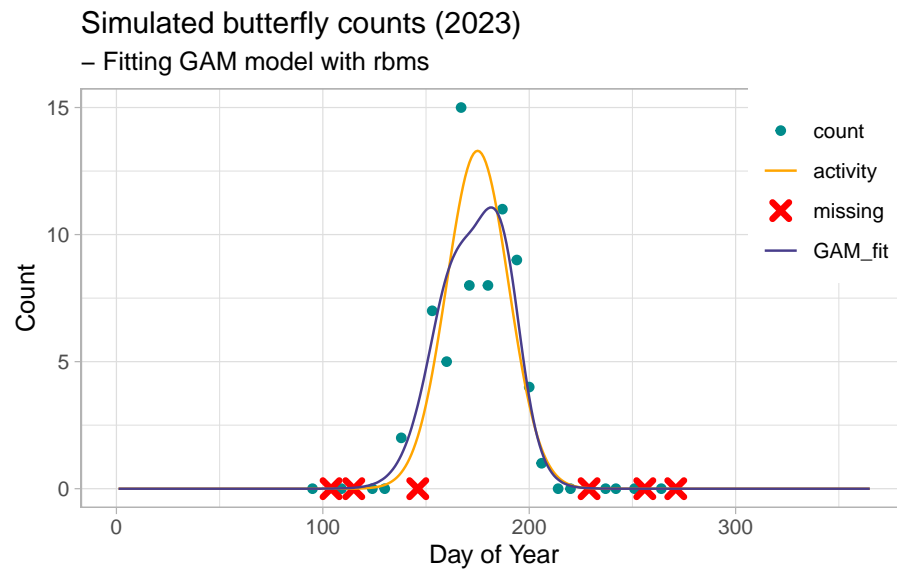
ts_flight_curve <- rbms::flight_curve(ts_season_count,
  NbrSample = 300,
  MinVisit = 5,
  MinOccur = 3,
  MinNbrSite = 1,
  MaxTrial = 4,
  GamFamily = 'nb',
  SpeedGam = FALSE,
  ComplSeason = TRUE,
  SelectYear = NULL,
  TimeUnit = 'd')

pheno <- ts_flight_curve$pheno

btfl_fig4 <- ggplot() +
  geom_point(data=ts_season_count[ANCHOR == 0 & !is.na(COUNT), ], aes(x=DAY_SINCE, y=COUNT)) +
  geom_point(data=btfl_week_missing, aes(x=doy, y=count, colour = "missing"),
    shape=4, size=2, stroke=2) +
  geom_line(data = btfl_ts, aes(x = doy, y = act, colour = "activity")) +
  geom_line(data = pheno,
    aes(x = trimDAYNO, y = btfl_ts[,unique(abund.true)]*NM, colour = "GAM_fit")) +
  xlim(1,365) + ylim(0, max(btfl_ts$act,
    pheno$NM*btfl_ts[,unique(abund.true)],
    btfl_week_missing$count,
    ts_season_count[!is.na(COUNT), COUNT] )) +
  scale_colour_manual("",
    breaks = c("count", "activity", "missing", "GAM_fit"),
    values = c(cnt_col, flc_col, missing_col, GAM_col)) +
  theme_light() +
  theme(legend.position = "inside", legend.position.inside = c(0.9, 0.8)) +
  labs(title = paste0("Simulated butterfly counts (", y, ")"),
    subtitle = "- Fitting GAM model with rbms",
    x = "Day of Year",
    y = "Count")

btfl_fig4

```



Non Gaussian flight curve

The same procedure can be applied to flight curves having more complex shapes. Here we will generate a time series of butterfly counts drawn from a known flight curve (adult activity), using simulation from a Zonneveld model.

```
btfl_data_zn <- timeseries_sim(nsims=10,
  year = c(2023),
  doy.samples = seq(from=1, to=365, by=1),
  abund.type = "exp",
  activity.type = "zon",
  sample.type = "nb",
  sim.parms = list(growth.rate = 0,
    init.size = 500,
    act.mean = 175,
    act.sd = 15,
    theta = 5,
    zon.theta = 50,
    t0 = 100,
    beta = 5,
    alpha = 0.05)
)

btfl_ts <- data.table(btfl_data_zn$timeseries)[, site_id := paste0("site_",sim.id)]
```

```

set.seed(13276)
btfl_ts[, date := as.Date(doy, origin = paste0(years, "-01-01"))-1]
btfl_ts[, week := isoweek(date)]
btfl_ts[month(date) != 1 | week < 50, weekday := rowid(week), by = .(site_id, years)]
btfl_week_smpl <- btfl_ts[weekday %in% c(2,3,4,5), ][btfl_ts[weekday %in% c(2,3,4,5), ][, .I[s
btfl_week_smpl <- btfl_week_smpl[month(date) %in% c(4:9), ]

btfl_week_missing <- data.table()

for(i in btfl_week_smpl[, unique(years)]){
  for(j in btfl_week_smpl[, unique(site_id)]){
    btfl_week_missing <- rbind(btfl_week_missing, sample_missing(data = btfl_week_smp
  }
}

visit_sim <- btfl_week_smpl['date %in% btfl_week_missing$date, .(site_id, date, count)]
count_sim <- visit_sim[count>=1,][, species := "sp1"]

names(visit_sim) <- toupper(names(visit_sim))
names(count_sim) <- toupper(names(count_sim))

ts_date <- rbms::ts_dwmy_table(InitYear = 2023, LastYear = 2023, WeekDay1 = 'monday')

ts_season <- rbms::ts_monit_season(ts_date,
  StartMonth = 4,
  EndMonth = 9,
  StartDay = 1,
  EndDay = NULL,
  ComplptSeason = TRUE,
  Anchor = TRUE,
  AnchorLength = 2,
  AnchorLag = 2,
  TimeUnit = 'd')

ts_season_visit <- rbms::ts_monit_site(ts_season, visit_sim)

ts_season_count <- rbms::ts_monit_count_site(ts_season_visit, count_sim, sp = "sp1")

# mod_k <- "COUNT ~ s(DAY_SINCE, bs = \"cr\", k = 5) + factor(SITE_ID)"

ts_flight_curve <- rbms::flight_curve(ts_season_count,
  NbrSample = 300,
  MinVisit = 5,
  MinOccur = 3,

```

```

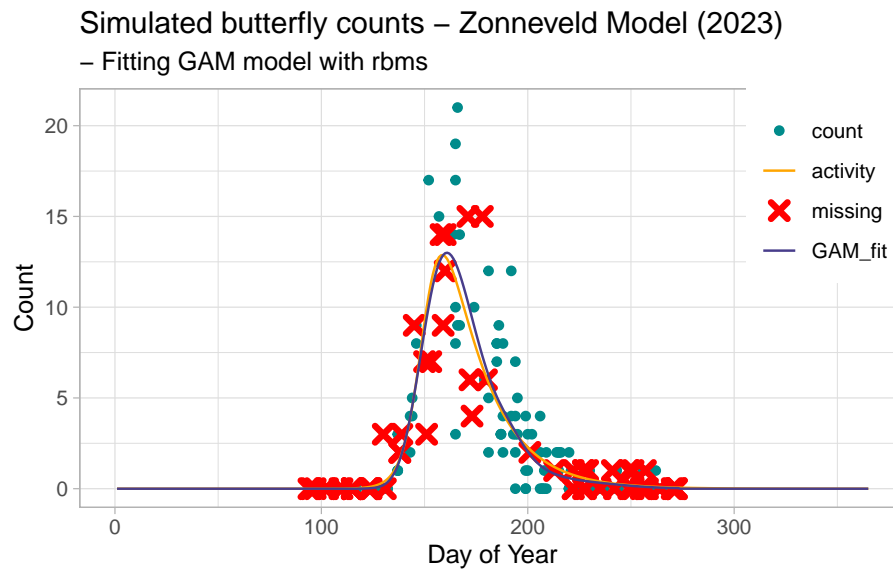
        MinNbrSite = 1,
        MaxTrial = 4,
        GamFamily = 'nb',
        SpeedGam = FALSE,
        CompltSeason = TRUE,
        SelectYear = NULL,
        #mod_form = mod_k,
        TimeUnit = 'd')

pheno <- ts_flight_curve$pheno

btfl_fig5 <- ggplot() +
  geom_point(data=ts_season_count[ANCHOR == 0 & !is.na(COUNT), ], aes(
  geom_point(data=btfl_week_missing, aes(x=doy, y=count, colour = "missing",
    shape=4, size=2, stroke=2) +
  geom_line(data = btfl_ts, aes(x = doy, y = act, colour = "activity"))
  geom_line(data = pheno, aes(x = trimDAYNO, y = btfl_ts[,unique(abund.true)]))
  xlim(1,365) + ylim(0, max(btfl_ts$act,
    pheno$NM*btfl_ts[,unique(abund.true)]),
    btfl_week_missing$count,
    ts_season_count[!is.na(COUNT), COUNT] )) +
  scale_colour_manual("",
    breaks = c("count", "activity", "missing", "GAM_fit"),
    values = c(cnt_col, flc_col, missing_col, GAM_col)) +
  theme_light() +
  theme(legend.position = "inside", legend.position.inside = c(0.9, 0.9))
  labs(title = paste0("Simulated butterfly counts - Zonneveld Model ("),
    subtitle = "- Fitting GAM model with rbms",
    x = "Day of Year",
    y = "Count")

btfl_fig5

```



Simple trend case

In the first scenario, we will apply the method to a simple case where we have one univoltine species that is monitored over 15 years across 100 sites where the populations follow the same trend with a known growth rate.

