Package 'bumbl'

July 22, 2025

Title Tools for Modeling Bumblebee Colony Growth and Decline

Version 1.0.3

Description Bumblebee colonies grow during worker production, then decline after switching to production of reproductive individuals (drones and gynes). This package provides tools for modeling and visualizing this pattern by identifying a switchpoint with a growth rate before and a decline rate after the switchpoint. The mathematical models fit by bumbl are described in Crone and Williams (2016) <doi:10.1111/ele.12581>.

License CC BY 4.0

Encoding UTF-8

LazyData true

URL https://github.com/Aariq/bumbl

BugReports https://github.com/Aariq/bumbl/issues

Depends R (>= 3.0)

Suggests testthat (>= 2.1.0), knitr, rmarkdown, covr, car, rsq

Imports dplyr, rlang, tidyr (>= 1.0.0), purrr, broom, glue, MASS, ggplot2, lifecycle

RoxygenNote 7.2.3

VignetteBuilder knitr

Config/testthat/edition 3

NeedsCompilation no

Author Eric R. Scott [aut, cre] (ORCID: <https://orcid.org/0000-0002-7430-7879>)

Maintainer Eric R. Scott <scottericr@gmail.com>

Repository CRAN

Date/Publication 2023-03-08 19:40:09 UTC

Contents

---

autoplot.bumbldf   *Plot observed and fitted results from bumbl()*

Description

Plots observed (points) and fitted (red line) values from the model implemented by bumbl(), faceted by colony.

Usage

```
## S3 method for class 'bumbldf'
autoplot(object, ..., colony = NULL, x = deprecated())
```

Arguments

object	a dataframe produced by [bumbl()](bumbl)
...	other arguments passed to autoplot()
colony	a character vector of colony IDs to plot
x	**[Deprecated]** x has been re-named to object for consistency with other autoplot() functions

Value

invisibly returns a ggplot object

Examples

```
bombus_subset <- bombus[bombus$colony %in% c("17", "104", "20", "24"), ]
results <- bumbl(bombus_subset, colonyID = colony, t = week,
                 formula = mass ~ week)
library(ggplot2)
autoplot(results)
```

bombus	Bombus vosnesenskii *colony growth and estimated floral resources*

Description

This is a subset of the data from Crone and Williams, 2016. *Bombus vosnesenskii* colonies were placed in sites in a mixed agricultural–natural area in California, USA and weighed weekly.

Usage

```
bombus
```

Format

A data frame with 178 rows and 10 columns

site The site ID

colony The colony ID

wild The proportion of the habitat for that site classified as wild using GIS

habitat The habitat type. Either wild (W), organic agriculture (O), or conventional agriculture (C)

date The calendar date the measurements were taken on

week Number of weeks since the initial mass was taken

mass Colony mass, in grams

d.mass Difference in mass from the smallest mass for each colony, in grams

floral_resources Floral density in millions of flowers within a 2km radius around the colony, weighted as a function of distance from colony

cum_floral Cumulative floral resources from the start of the study

Details

Floral resources (*e*) are weighted using the equation $w = e^{\wedge}(\text{-}D/a)$ where D is the distance in meters between the resources and the colony and a is the typical flight distance for the species. This is described in further detail in Williams et al. 2012.

References

Crone, E. E., and Williams, N. M. (2016). Bumble bee colony dynamics: quantifying the importance of land use and floral resources for colony growth and queen production. Ecol. Lett. 19, 460–468. doi:10.1111/ele.12581.

Williams, N. M., Regetz, J., and Kremen, C. (2012). Landscape-scale resources promote colony growth but not reproductive performance of bumble bees. Ecology 93, 1049–1058. doi:10.1890/11-1006.1.

---

bumbl   *Estimate colony growth, switch point, and decay parameters*

Description

Fits generalized linear models that assume bumblebee colonies will switch from growth to gyne production at some point, τ. This allows for a different switchpoint (τ) for each colony, chosen by maximum likelihood methods.

Usage

```
bumbl(
  data,
  t,
  formula,
  family = gaussian(link = "log"),
  colonyID = NULL,
  augment = FALSE,
  keep.model = FALSE,
  tau_optim_maxit = 100,
  ...
)
```

Arguments

data	a dataframe or tibble with a column for colony ID (as a `character` or `factor`), a column for some measure of time (`numeric`), and a column for some measure of colony growth (`numeric`), at minimum.
t	the unquoted column name of the time variable.
formula	a formula with the form `response ~ time + covariates` where response is your measure of colony growth, time is whatever measure of time you have (date, number of weeks, etc.) and covariates are any optional co-variates you want to fit at the colony level.
family	a description of the error distribution and link function. This is passed to [glm()](#) except in the case of `family = "negbin"`, which causes [MASS::glm.nb()](#) to be used to fit a negative binomial GLM.
colonyID	the unquoted column name of the colony ID variable. This is required, so to run `bumbl()` on a single colony, you must create a dummy column with a colony ID.
augment	when FALSE, `bumbl` returns a summary dataframe with one row for each colonyID. When TRUE, it returns the original data with additional columns containing model coefficients. Cannot be used in conjunction with `keep.model = TRUE`.
keep.model	If TRUE, then the output will contain a list-column with the models for each colony. This may be useful for extracting statistics and performing model diagnostics not provided by `bumbl()`. Learn more about working with list columns with `vignette("nest", package = "tidyr")`.

tau_optim_maxit

> passed to optim() which is used to find the optimal change point, tau. Mostly used for testing purposes (to force convergence errors), but could be increased if optimal switchpoint doesn't converge.

...   additional arguments passed to glm() or MASS::glm.nb().

Details

Colony growth is modeled as increasing exponentially until the colony switches from producing workers to producing reproductive individuals (drones and gynes), at which time the workers die and gynes leave the colony, causing the colony to decline. The switch point, τ, may vary among colonies. bumbl() finds the value of τ that maximizes likelihood and this "winning" model is used to calculate statistics returned in the output. This function works by fitting generalized linear models (GLMs) to modified colony growth data. Because of this, the assumptions for GLMs apply, namely independence and homogeneity of variance. See vignette("bumbl", package = "bumbl") for more details on the underlying math of the model.

Value

A data.frame with the additional class bumbldf containing a summary of the data with a row for every colony and the following columns:

- converged indicates whether the winning model converged.
- tau is the switchpoint, in the same units as t, for each colonyID. The colony grows for τ weeks, then begins to decline in week $\tau + 1$.
- logN0 is the intercept of the growth function. It reflects actual initial colony size, if the colony initially grows exponentially. It would also be lower if there were a few weeks lag before growth started in the field.
- logLam is the average (log-scale) colony growth rate (i.e., rate of weight gain per unit t) during the growth period.
- decay reflects the rate of decline during the decline period. Equivalent to $\ln(\delta) - \ln(\lambda)$ (see vignette for more in-depth explanation).
- logNmax is the maximum weight reached by each colony. It is a function of tau, logN0 and logLam
- Additional columns are coefficients for any covariates supplied in the formula

When augment = TRUE, the original data are returned with these columns as well as fitted values (.fitted) residuals (.resid) and standard error (.se.fit). When keep.model = TRUE a list-column with the glm models for each colony is returned as well.

Note

This function *assumes* there is a switchpoint and does not test whether the switchpoint model is significantly better than a log-linear model. As a result, it may estimate a switchpoint even if the data do not represent a true switchpoint. See the vignette for an example of how to extract the GLMs—you could compare them to a simpler log-linear model without the switchpoint by AIC or a likelihood ratio test to test the significance of the switchpoint.

References

Crone EE, Williams NM (2016) Bumble bee colony dynamics: quantifying the importance of land use and floral resources for colony growth and queen production. Ecology Letters 19:460–468. https://doi.org/10.1111/ele.12581

See Also

`plot.bumbldf()`

Examples

```
bumbl(bombus, colonyID = colony, t = week, formula = d.mass ~ week)
```

---

plot.bumbldf   *Plot observed and fitted results from bumbl()*

Description

Creates one plot per level of colonyID showing the observed (points) and fitted (red line) values from the model implemented by bumbl().

Usage

```
## S3 method for class 'bumbldf'
plot(x, ..., colony = NULL)
```

Arguments

x	a dataframe produced by `bumbl()`.
...	other arguments not used by this method.
colony	optional vector of colony ID's (character) or indexes (numeric) to plot. If not supplied, all colonies will be plotted.

Value

invisibly returns a list of dataframes used for building the plots.

See Also

`bumbl()`, `autoplot.bumbldf()`

Examples

```
set.seed(687)
colonyID_subset <- sample(bombus$colony, 10)
colony_subset <- bombus[bombus$colony %in% colonyID_subset, ]
results <- bumbl(colony_subset, colonyID = colony, t = week,
                 formula = mass ~ week)
plot(results)
```

---

test_df *Simulated colony growth data for internal tests*

Description

Simulated data including two colonies (6 and 7) that should always cause errors for `bumbl()`

Usage

```
test_df
```

Format

An object of class `tbl_df` (inherits from `tbl`, `data.frame`) with 140 rows and 4 columns.

Details

Colonies 1-5 are generated with `sim_colony()` using the following parameters:

colony	tau	n0	lambda	delta
1	11.0	30	1.42	0.30
2	11.5	24	1.35	0.31
3	9.0	49	1.40	0.29
4	12.0	37	1.34	0.48
5	10.6	29	1.37	0.32

Colonies 6 and 7 are generated manually. Colony 6 contains negative numbers and should result in the warning "Warning, cannot find valid starting values: please specify some for colonyID '6' Omitting from results." Colony 7 has all the data pushed up to week 25 and therefore should result in a convergence error when searching for the optimal switchpoint and the warning, "Warning: search for optimal switchpoint did not converge for colonyID '6'. Omitting from results."

Index