How to Walk the BeeWalk: Modelling Bumblebee Citizen Science Data

Fabian Ketwaroo, Eleni Matechou

November 6, 2022

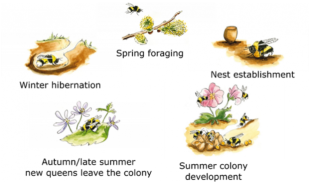
Outline

BUMBLEBEES

- There are 24 species of bumblebee in the UK, 18 of which are social.
- They feed exclusively on pollen and nectar and are cold-adapted.
- Several species of bumblebee are known to be declining. A striking example is the great yellow bumblebee, which used to be distributed throughout the UK but now can only be found on the north coast of Scotland.

Life cycle

- "Old queens" emerge from hibernation in early spring and establish nests.
- Workers emerge throughout spring and summer and help support the nest.



- "New queens" and males emerge towards the end of the season and mate.
- New queens go into hibernation and (some of them) emerge the following year as old queens and the cycle starts again.

BeeWalk

- Changes in abundance typically are an early warning of changes in distribution that are yet to come.
- The BeeWalk `www.beewalk.org.uk` was established to monitor abundance of UK bumblebees.
- Volunteers walk a monthly transect March-October and record the number of bumblebees, their species and caste, where possible, they detect.

THE DATA

- Not all sites are visited each month/year and visits can be made on any day of the month.
- Old and new queens are indistinguishable so they can only be identified as queens, whereas any caste can be classed as "unidentified", which creates four groups in terms of the observation process "queens", "workers", "males", "unknown".

Site	Time	Queens	Workers	Males	Unknown	Queens	Workers	Males	Unknown
A	1	2	5	0	1	NA	NA	NA	NA
B	1	1	10	0	2	3	0	0	1
A	2	NA	NA	NA	NA	0	6	0	2
B	2	0	12	0	5	NA	NA	NA	NA
A	3	0	15	0	1	0	8	0	0
B	3	0	20	0	0	0	15	0	2

ISSUES WITH DATA

- Ideally, the models would be built at site and day level, as in similar work for butterflies[1], but the data are too sparse to allow for such fine scale modelling.
- Instead, temporally, we group records to weeks, and spatially, to the whole of the UK.

[1] Matechou, E., Dennis, E. B., Freeman, S. N., and Brereton, T. (2014). Monitoring abundance and phenology in (multivoltine) butterfly species: a novel mixture model. Journal of Applied Ecology, 51(3), 766-775.

The model

- We model each part of the underlying latent process: emergence, survival, reproduction and we also model the observation process: identification.
- The model is based on the Matechou et al. 2018[2] classical model, but we employ a Bayesian approach here and do not have any deterministic parts in our model.
- We account for caste-specific emergence patterns, productivity parameters and identification probabilities and a stochastic separation of "queens" to old and new queens and of "unknown" to all castes.

[2] Matechou, E., Freeman, S. N., and Comont, R. (2018). Caste-specific demography and phenology in bumblebees: modelling BeeWalk data. Journal of Agricultural, Biological and Environmental Statistics, 23(4), 427-445.

Model description
Emergence pattern

- We model the emergence pattern of each caste using a normal pdf with caste-specific mean and variance so that

$$\beta_{y(t-1)c} = F_{yc}(t) - F_{yc}(t-1)$$

To ensure $\sum_{t=1}^{T} \beta_{y(t-1)c} = 1 \;\; \forall \; y \; c$, we treat the first and last intervals as open-ended and set $\beta_{y0c} = F_{yc}(1)$ and $\beta_{y(T-1)c} = 1 - \sum_{t=1}^{T-1} \beta_{y(t-1)c}$.

MODEL DESCRIPTION
EMERGENCE AND PRODUCTIVITY

- E_{ytc}: number of individuals emerging from caste c, in year y, between week t and $t-1$

$$E_{1tQ_0} \sim \text{Poisson}(v \times \beta_{1(t-1)Q_0})$$
$$E_{ytQ_0} \sim \text{Poisson}(N_{(y-1)Q_n} \times \xi_{y-1} \times \beta_{y(t-1)Q_0})) \;\; \text{for } y = 2, \ldots, Y$$
$$E_{ytc} \sim \text{Poisson}(N_{yQ_0} \times \rho_{yc} \times \beta_{y(t-1)c}) \;\; \forall\, y,\; c = W, M, Q_n$$

Model description
Survival

- S_{ytc}: number of individuals from caste c in year y that survive (apparently!) from week $t-1$ to t

$$S_{ytc} = \text{Binomial}(M_{y(t-1)c}, \phi_{yc}) \;\; \forall \; y, \; c, \; t = 2, \ldots, T$$

- M_{ytc}: number of individuals "around" from caste c in year y and week t

$$M_{y1c} = E_{y1c} \;\; \forall \; y, \; c$$
$$M_{ytc} = S_{ytc} + E_{ytc} \;\; \forall \; y, \; c, \; t = 2, \ldots, T$$

Model description
Identification

- A_{ytc}: number of individuals from caste c detected and identified in year y, week t.

$$A_{ytQ_0} \sim \text{Binomial}(M_{ytQ_0}, \psi_{yQ_0}) \,\forall\;\, y,\; t$$
$$A_{ytW} \sim \text{Binomial}(M_{ytW}, \psi_{yW}) \;\,\forall\, y\; t$$
$$A_{ytM} \sim \text{Binomial}(M_{ytM}, \psi_{yM}) \;\,\forall\, y\; t$$
$$A_{ytQ_n} \sim \text{Binomial}(M_{ytQ_n}, \psi_{yQ_n}) \,\forall\;\, y,\; t$$

FINALLY...

- κ_{ytg}: number of individuals assigned to group g in year y , time t, where

$$\kappa_{ytQ} = A_{tyQ_0} + A_{tyQ_n} \; \forall \; y, \; t$$
$$\kappa_{ytW} = A_{ytW} \; \forall \, y, \; t$$
$$\kappa_{ytM} = A_{ytM} \; \forall \, y, \; t$$
$$\kappa_{ytU} = ((M_{yt1} + M_{yt4}) - \kappa_{yt1}) + (M_{yt2} - \kappa_{yt2}) + (M_{yt3} - \kappa_{yt3}) \, \forall y, t.$$

- We model x_{ytg}, the aggregate of counts collected in year y, time t for each group g as:

$$x_{ytg} \sim \text{Poission}(\lambda_{ytg}) \;\; \forall \; y, \; t, \; g$$

- λ_{ytg}: expected number of individuals detected and assigned to group g at all sites visited in year y, time t. Let n_{yt} be the total number of sites visited in year y, time t. Hence,

$$\lambda_{ytg} \propto \kappa_{ytg} \times n_{yt} \, \forall \; y, t, g$$

The app

BeeWalk

Help | Data | Settings | Results | Diagnostics

Background | Data | Settings | Results | Diagnostics

This app implements a Bayesian modelling approach of the methods developed by E., Matechou, S. N. Freeman and R. Comont, in *Caste-Specific Demography and Phenology in Bumblebees: Modelling BeeWalk Data*, Journal of Agricultural, Biological and Environmental Statistics 23.4 (2018): 427-445.

The app can be used to model bumblebee count data collected at S sites in Y seasons or years, with T sampling occasions within each year, assumed to be equally spaced apart, for example taking place weekly. Since these sampling occasions are equally spaced apart, we also refer to them as time, where time = 1 corresponds to the first sampling occasion, time = 2 to the second etc.

Each sampling occasion within each year gives rise to the number of bumblebees counted for a particular species and group, where groups are defined as "queens", "workers", "males" and "unknown", the latter corresponding to the number of bumblebees detected for the species that did not have their caste identified.

We consider the aggregate of counts collected at all S sites at each time t, and hence, the data are summarised in **X** of dimension Y x T x 4 with the third dimension, which we denote by g = 1,2,3,4 denoting the group (queens, workers, males, unknown) to which an individual has been assigned. The model estimates caste-specific parameters, such as phenology, and we denote castes by c=1,2,3,4 with 1 denoting old queens, 2 denoting workers, 3 denoting males and 4 denoting new queens. We note that data are only available on groups but inference is made on castes.

We list the parameters below as they appear in the Matechou et al (2018) paper below for users who are familiar with that model. We note that in the app this notation is not used in the plots produced in the results, and instead we only use the parameter names as they appear below. Also note that even though the general model allows for all parameters to be year and sampling occasion specific (where appropriate), due to the typical sparseness of the data we constrain some of the parameters to be the same across years or sampling occasions, as indicated below.

ρ_{yc}: **Within-season productivity**: mean number of individuals in caste c, c=2, 3, 4, per old queen in year y. Here we allow productivity to vary by caste but not by year.

ξ_{y-1}: **Winter survival probability**: probability a new queen survives the winter in year y-1 and hence is available for detection as an old queen in year y.

$\beta_{y(t-1)c}$: **Emergence probability**: probability that an individual from caste c in year y emerges from the nest or from winter dormancy as appropriate, between times t-1 and t.

φ_{ytc}: **Within-season apparent survival probability**: probability that an individual from caste c in year y that is alive at time t survives to time t+1. Here we allow survival probability to vary by year and caste but not by time.

ψ_{ytc}: **Identification probability**: probability that an individual from caste c in year y that is detected at time t has its caste identified. Here we allow identification probability to vary by year and caste but not by time.

Data

BeeWalk

Help | Data | Settings | Results | Diagnostics

Choose CSV file

Browse... pasc11_16.csv

Upload complete

File Preview

Site	week	queens_11	workers_11	males_11	unknown_11	queens_12	workers_12	males_12	unknown_12	queens_13	workers_13	males_13	unknown_13
29 Shirley Rd.	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
A48 - Blackweir (alternative path to Taff Trail)	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
A48 - North Rd	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Abbotstone Down	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Abercorn-Duddingston	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Aberdaron Headland	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Number of Years

6

Number of Sampling Occasions

40

Settings

BeeWalk

Help Data Settings Results Diagnostics

MCMC settings

Number of iterations

650000

Number of burn-in iterations

50000

Number of thinned iterations

100

Number of chains

1

Run

Cleaning Data. This may take a while..

DISCUSSION

- The model and app have been motivated by the BeeWalk, but it is applicable to any such scheme.
- At the moment, we have to ignore differences in space and assume that emergence patterns are the same across the UK. Bumblebees are not as sensitive to differences in temperature as butterflies, but still it would be nice to relax this assumption, but without having to estimate site-specific relative abundances.
- Computation time increases every year as more data are collected. Using the whole time series in model fitting is the sensible choice but it will soon become too computationally expensive.

Thank you!
Any questions/comments?