# How to Walk the BeeWalk: Modelling Bumblebee Citizen Science Data

Fabian R. Ketwaroo, Eleni Matechou, Richard Comont

Tuesday 18th April, 2023

# Bumblebees

- Bumblebees provide pollination services and function as indicator species for changes in climate or land use.

- Declines, either of distribution or abundance, are thus of serious concern from agricultural and economic viewpoints as well as from a conservation point of view.

# Life cycle

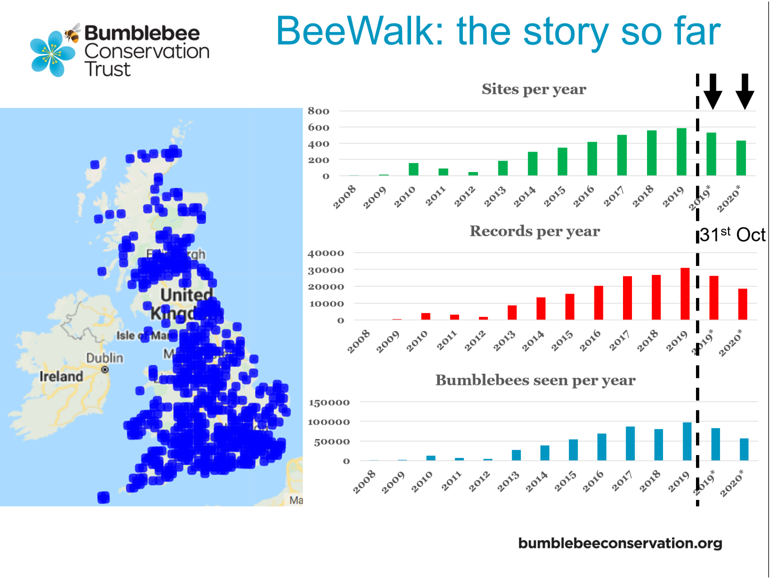
- "Old queens" emerge from hibernation in early spring and establish nests.
- Workers emerge throughout spring and summer and help support the nest.


- "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

- Several species of bumblebee are known to be declining.
- 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 the caste they detect.

# BeeWalk


Bumblebee Conservation Trust
BeeWalk: the story so far
Edinburgh
United Kingdom
Isle of Man
Dublin
Ireland
Sites per year
800
600
400
200
0
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2019*
2020*
31st Oct
Records per year
40000
30000
20000
10000
0
Bumblebees seen per year
150000
100000
50000
0
bumblebeeconservation.org

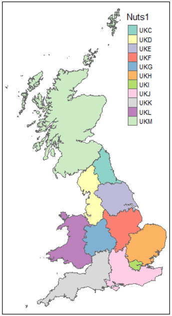
# THE DATA

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

| 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.
- Temporally, we group records to weeks, and spatially, to the NUTS1 regions of the UK excluding Northern Ireland.




[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 extend this model to allow for spatiotemporal modelling in a Bayesian framework.

- We account for caste-specific and area-specific emergence patterns, productivity parameters with caste-specific 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 area and caste using a normal pdf with area and caste-specific mean ($\mu_{ycs}$) and caste-specific variance ($\sigma_c^2$) so that

$$\beta_{y(t-1)cs} = F_{ycs}(t) - F_{ycs}(t-1)$$

To ensure $\sum_{t=1}^{T} \beta_{y(t-1)cs} = 1 \;\; \forall \; y \; c \; s$, we treat the first and last intervals as open-ended and set $\beta_{y0cs} = F_{ycs}(1)$ and $\beta_{y(T-1)cs} = 1 - \sum_{t=1}^{T-1} \beta_{y(t-1)cs}$.

# Model description
# Emergence and Productivity

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

$$E_{1tQ_0s} \sim \text{Poisson}(v \times \beta_{1(t-1)Q_0s})$$
$$E_{ytQ_0s} \sim \text{Poisson}(N_{(y-1)Q_ns} \times \xi_{y-1} \times \beta_{y(t-1)Q_0s})) \text{ for } y = 2, \dots, Y$$
$$E_{ytcs} \sim \text{Poisson}(N_{yQ_0s} \times \rho_{ycs} \times \beta_{y(t-1)cs}) \;\; \forall\, y, s, \; c = W, M, Q_n$$

where $N_{ycs} = \sum_{t-1}^{T} E_{ytcs}$ and $v$ is the relative abundance of $Q_o$ per site in year 1.

# Model description
# Emergence and Productivity

- We use a space-time separable model to model the spatial and temporal correlation among areas and years.

$$
\begin{aligned}
log(\mu_{ycs}) &= \alpha_{\mu c} + SP_{\mu_{cs}} + TM_{\mu_{yc}} \\
SP_{\mu_{c,1:S}} &\sim ICAR(W_{sp}, \sigma^2_{\mu_c}) \\
TM_{\mu_{yc}} &\sim RW1(\sigma^2_{\mu_{TM}})
\end{aligned}
$$

$$
\begin{aligned}
log(\rho_{ycs}) &= \alpha_{\rho c} + SP_{\rho_{cs}} + TM_{\rho_{yc}} \\
SP_{\rho_{c,1:S}} &\sim ICAR(W_{sp}, \sigma^2_{\rho_c}) \\
TM_{\rho_{yc}} &\sim RW1(\sigma^2_{\rho_{TM}})
\end{aligned}
$$

# Model description
# Survival

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

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

- $M_{ytcs}$: number of individuals "around" from caste $c$ in area $s$, year $y$ and week $t$

$$M_{y1cs} = E_{y1cs} \ \ \forall\, y,\ c$$
$$M_{ytcs} = S_{ytcs} + E_{ytcs} \ \ \forall\, y,\ c,\ t = 2, \ldots, T$$

# Model description
# Identification

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

$$A_{ytQ_0s} \sim \text{Binomial}(M_{ytQ_0s}, \psi_{yQ_0}) \ \forall \ \ y, \ t \ s$$
$$A_{ytWs} \sim \text{Binomial}(M_{ytWs}, \psi_{yW}) \ \ \forall \ y \ t \ s$$
$$A_{ytMs} \sim \text{Binomial}(M_{ytMs}, \psi_{yM}) \ \ \forall \ y \ t \ s$$
$$A_{ytQ_ns} \sim \text{Binomial}(M_{ytQ_ns}, \psi_{yQ_n}) \ \forall \ \ y, \ t \ s$$

# FINALLY...

- $\kappa_{ytgs}$: number of individuals assigned to group $g$ in area $s$, year $y$ , time $t$, where

$$\kappa_{ytQs} = A_{tyQ_0s} + A_{tyQ_ns} \forall \ y, \ t \ s$$
$$\kappa_{ytWs} = A_{ytWs} \ \forall \, y, \ t \ s$$
$$\kappa_{ytMs} = A_{ytMs} \ \forall \, y, \ t \ s$$
$$\kappa_{ytUs} = ((M_{yt1s} + M_{yt4s}) - \kappa_{yt1}) + (M_{yt2s} - \kappa_{yt2s}) + (M_{yt3s} - \kappa_{yt3s})$$

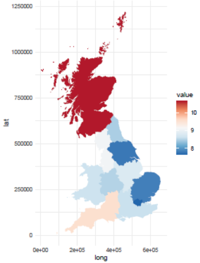
- We model $x_{ytgs}$, the aggregate of counts collected in area $s$, year $y$, time $t$ for each group $g$ as:

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

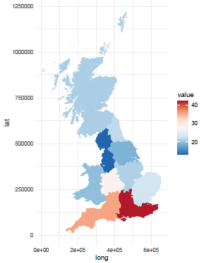
- $\lambda_{ytgs}$: expected number of individuals detected and assigned to group $g$ in area $s$, year $y$, time $t$. Let $n_{yts}$ be the total number of sites visited in area $s$, year $y$, time $t$. Hence,

$$\lambda_{ytgs} \propto \kappa_{ytgs} \times n_{yts} \, \forall \, y, t, g, s$$

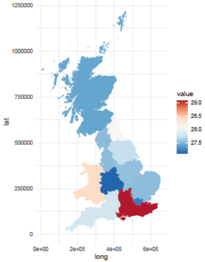
# Pascuorum- Emergence Maps



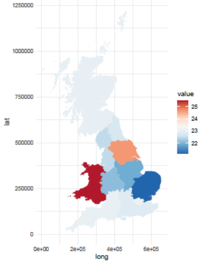

(a) Old Queens




(b) New Queens

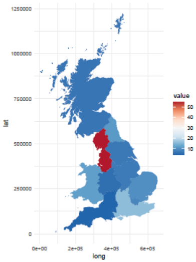



(c) Males

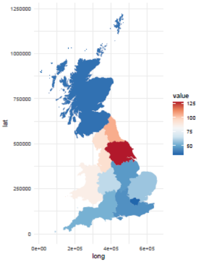



(d) Workers

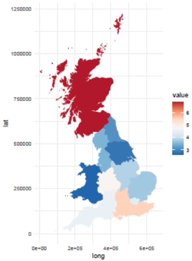
# Pascuorum - Productivity Maps




(a) Males

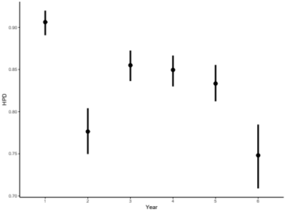



(b) Workers

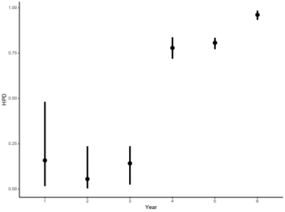



(c) New Queens

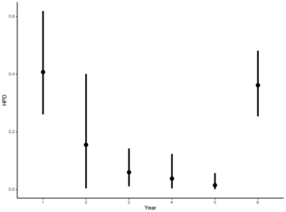
# Pascuorum - Survival



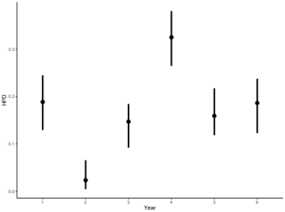

(a) Old Queens




(b) New Queens

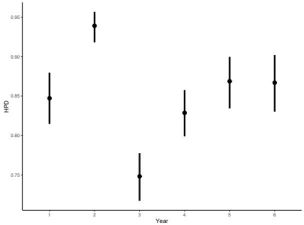



(c) Males

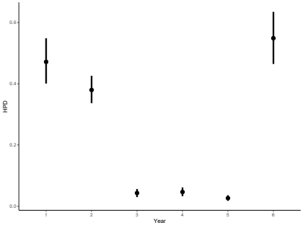



(d) Workers

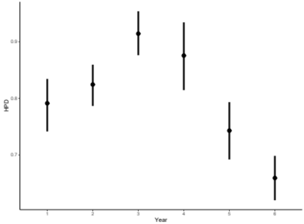
# Pascuorum - Identification



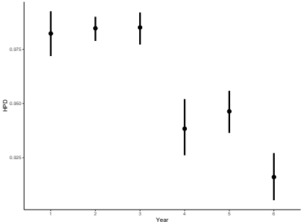

(a) Old Queens




(b) New Queens




(c) Males




(d) Workers

# SUMMARY

- We present a novel Bayesian dynamic mixture model for *Beewalk* citizen science data that accounts for sparsity and enables spatio-temporal modelling.

- This framework produces invaluable information on caste-specific and area-specific demographic parameters such as phenology and relative abundance, amongst others.

# Thank you!
# Any questions/comments?

Contact: **fk231@kent.ac.uk**