

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.

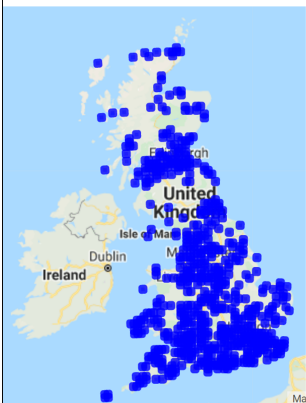


- ▶ “New queens” and males emerge towards the end of the season and mate.
- ▶ New queens go into hibernation and (some of them) emerge the following year as old queens and the cycle starts again.

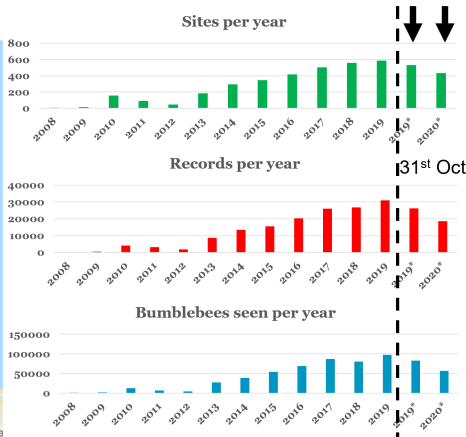
BEEWALK

- ▶ 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



BeeWalk: the story so far



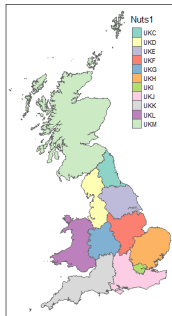
THE DATA

- ▶ 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 butterflies¹, 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.



¹Matechou, 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. 2018² 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.

²Matechou, 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 ($\mu_{y_{cs}}$) and caste-specific variance (σ_c^2) so that

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

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

$$E_{1tQ_0s} \sim \text{Poisson}(v \times \beta_{1(t-1)Q_0s})$$

$$E_{ytQ_0s} \sim \text{Poisson}(N_{(y-1)Q_n s} \times \xi_{y-1} \times \beta_{y(t-1)Q_0s}) \quad \text{for } y = 2, \dots, Y$$

$$E_{ytcs} \sim \text{Poisson}(N_{yQ_0s} \times \rho_{y cs} \times \beta_{y(t-1)cs}) \quad \forall y, s, c = W, M, Q_n$$

where $N_{y cs} = \sum_{t=1}^T E_{ytcs}$ and v is the relative abundance of Q_0 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.

$$\log(\mu_{ycs}) = \alpha_{\mu c} + SP_{\mu cs} + TM_{\mu yc}$$

$$SP_{\mu c, 1:S} \sim ICAR(W_{sp}, \sigma_{\mu c}^2)$$

$$TM_{\mu yc} \sim RW1(\sigma_{\mu TM}^2)$$

$$\log(\rho_{ycs}) = \alpha_{\rho c} + SP_{\rho cs} + TM_{\rho yc}$$

$$SP_{\rho c, 1:S} \sim ICAR(W_{sp}, \sigma_{\rho c}^2)$$

$$TM_{\rho yc} \sim RW1(\sigma_{\rho TM}^2)$$

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

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

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

$$M_{y1cs} = E_{y1cs} \quad \forall y, c$$

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

MODEL DESCRIPTION

IDENTIFICATION

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

$$A_{ytQ_0s} \sim \text{Binomial}(M_{ytQ_0s}, \psi_{yQ_0}) \quad \forall y, t, s$$

$$A_{ytWs} \sim \text{Binomial}(M_{ytWs}, \psi_{yW}) \quad \forall y, t, s$$

$$A_{ytMs} \sim \text{Binomial}(M_{ytMs}, \psi_{yM}) \quad \forall y, t, s$$

$$A_{ytQ_ns} \sim \text{Binomial}(M_{ytQ_ns}, \psi_{yQ_n}) \quad \forall y, t, s$$

FINALLY...

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

$$\kappa_{y t Q s} = A_{t y Q_0 s} + A_{t y Q_n s} \quad \forall y, t, s$$

$$\kappa_{y t W s} = A_{y t W s} \quad \forall y, t, s$$

$$\kappa_{y t M s} = A_{y t M s} \quad \forall y, t, s$$

$$\kappa_{y t U s} = ((M_{y t 1 s} + M_{y t 4 s}) - \kappa_{y t 1}) + (M_{y t 2 s} - \kappa_{y t 2 s}) + (M_{y t 3 s} - \kappa_{y t 3 s})$$

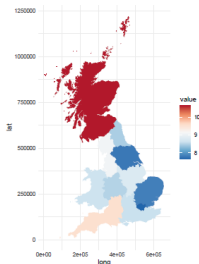
- ▶ We model $x_{y t g s}$, the aggregate of counts collected in area s , year y , time t for each group g as:

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

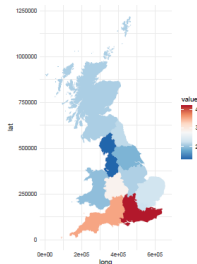
- ▶ $\lambda_{y t g s}$: expected number of individuals detected and assigned to group g in area s , year y , time t . Let $n_{y t s}$ be the total number of sites visited in area s , year y , time t . Hence,

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

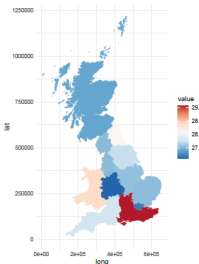
PASCUORUM- EMERGENCE MAPS



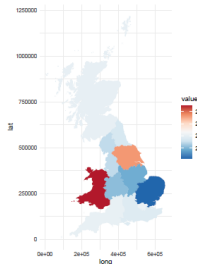
(a) Old Queens



(b) New Queens

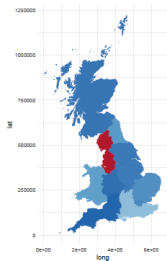


(c) Males

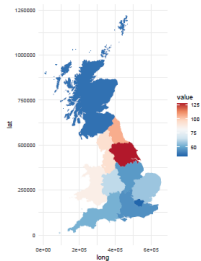


(d) Workers

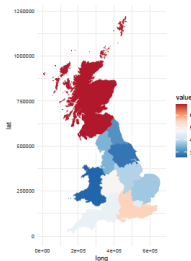
PASCUORUM - PRODUCTIVITY MAPS



(a) Males

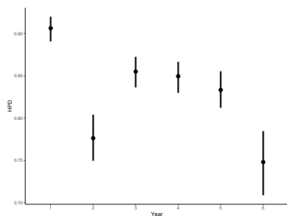


(b) Workers

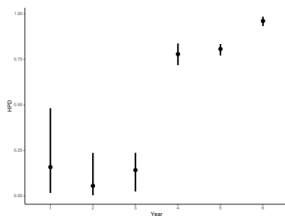


(c) New Queens

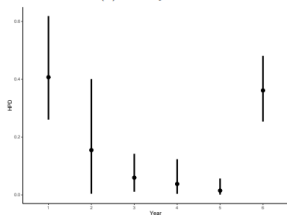
PASCUORUM - SURVIVAL



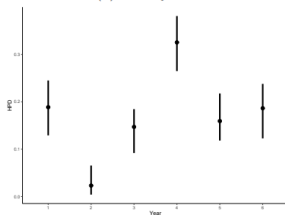
(a) Old Queens



(b) New Queens

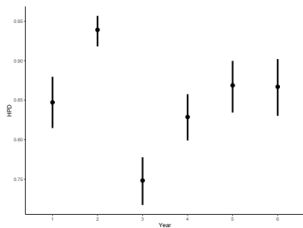


(c) Males

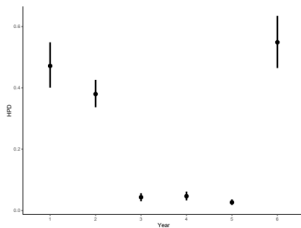


(d) Workers

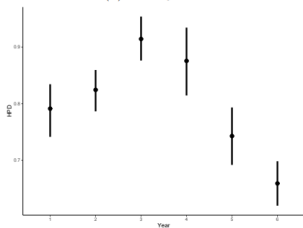
PASCUORUM - IDENTIFICATION



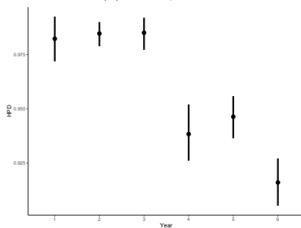
(a) Old Queens



(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