Supplementary Material for: "Shifts in colour morph frequencies along an urbanisation gradient in the ground beetle *Pterostichus madidus*"

Maxime Dahirel, Hélène Audusseau, Solène Croci

S1 - Model description

For the present study, our response variable is $B_{i,j}/N_{i,j}$, the proportion of black-legged *P. madidus* beetles captured in a given woodland i during a given sampling session j, with $B_{i,j}$ the number of black-legged beetles and $N_{i,j}$ the number of individuals of both morphs. Note that we naturally only include woodland $\times$ session combinations with $N_{i,j} > 0$. We ran the models below for each of the 8 possible urbanisation metrics described in the main text.

We initially built binomial models as follows:

$$B_{i,j} \sim \text{Binomial}(N_{i,j}, p_{i,j}),$$
$$\text{logit}(p_{i,j}) = \beta_0 + (\beta_1 + \eta_j) \times x_i + \alpha_i + \gamma_j,$$

with x_i the (centered and scaled) urbanisation metric at site i, β_0 and β_1 the fixed effects (intercept and urbanisation slope respectively), α_i the site-specific random-effect intercepts, and γ_j and η_j the session-specific random intercept and slope. Random effects are distributed as follows:

$$\alpha_i \sim \text{Normal}(0, \sigma_\alpha),$$
$$\begin{bmatrix} \gamma_j \\ \eta_j \end{bmatrix} \sim \text{MVNormal}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{\Omega} \right),$$

where $\boldsymbol{\Omega}$ is the covariance matrix for the session-specific random effects, which can be decomposed into its constituent standard deviations and correlation matrix $\boldsymbol{R}$ in this way:

$$\boldsymbol{\Omega} = \begin{bmatrix} \sigma_\gamma & 0 \\ 0 & \sigma_\eta \end{bmatrix} \boldsymbol{R} \begin{bmatrix} \sigma_\gamma & 0 \\ 0 & \sigma_\eta \end{bmatrix}.$$

We used weakly informative priors mostly inspired by McElreath (2020). We used $\text{Normal}(0, 1.5)$ priors for the fixed effect intercept β_0, which corresponds to the logit of a proportion, and $\text{Normal}(0, 1)$ for the fixed effect slope β_1. We used $\text{Half} - \text{Normal}(0, 1)$ priors for all standard deviations σ, and a $\text{LKJ}(2)$ prior for the correlation matrix $\boldsymbol{R}$.

Evaluations of these models revealed slight but consistent evidence of overdispersion (see archived data analysis code). We therefore fitted the equivalent beta-binomial models to account for that overdispersion (Harrison 2015):

$$B_{i,j} \sim \text{BetaBinomial}(N_{i,j}, p_{i,j}, \phi),$$

where ϕ is an overdispersion parameter with prior $1/\phi \sim \text{Half} - \text{Normal}(0, 1)$ (based on the similar parameterisation of the negative binomial distribution; seee e.g. https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations#story-when-the-generic-prior-fails-the-case-of-the-negative-binomial). The remainder of the models is the same as in the binomial case.

S2 - Model performance comparisons

We used two metrics to compare our beta-binomial models. We first used K-fold cross-validation (with $K = 10$) to evaluate these models based on their overall pointwise predictive accuracy (Vehtari, Gelman, and Gabry 2017). We then also compared them based on the proportion of (logit scale) among-site variance explained by fixed effects $\frac{\sigma_\beta^2}{\sigma_\beta^2+\sigma_\alpha^2}$, where the variance explained by fixed effects σ_β^2 is estimated as in Nakagawa and Schielzeth (2013). This is because since all models include a site random effect, comparing them on their overall performance may not reflect the way urbanisation specifically explains average among-site differences, as the site random effect will "absorb" any among-site variation not explained by the urbanisation metric.

All eight models had very similar overall predictive performance based on cross-validation results (**Table S2-1**). The proportion of among-site variance explained by the effect of urbanisation was highest at the 100 m scale and decreased as buffer size increased, with the lowest value for the fixed effect of distance to the urban centroid (**Fig. S2-1**). However, there is actually limited support for differences between models, with posteriors all largely overlapping (**Fig. S2-1**).

Table S2-1. Expected log pointwise predictive density (elpd) for each of the 8 beta-binomial models, expressed as differences ± SE from the elpd of the "best" model in the set.

Urbanisation metric	elpd deviation from "best" model ± SE
IMD (600m)	0.00 ± 0.00
IMD (100m)	-0.01 ± 1.52
IMD (300m)	-0.33 ± 0.50
IMD (900m)	-0.38 ± 0.45
IMD (1500m)	-0.50 ± 0.52
IMD (1800m)	-0.60 ± 0.58
IMD (1200m)	-0.69 ± 0.48
distance to urban centroid	-1.70 ± 1.31

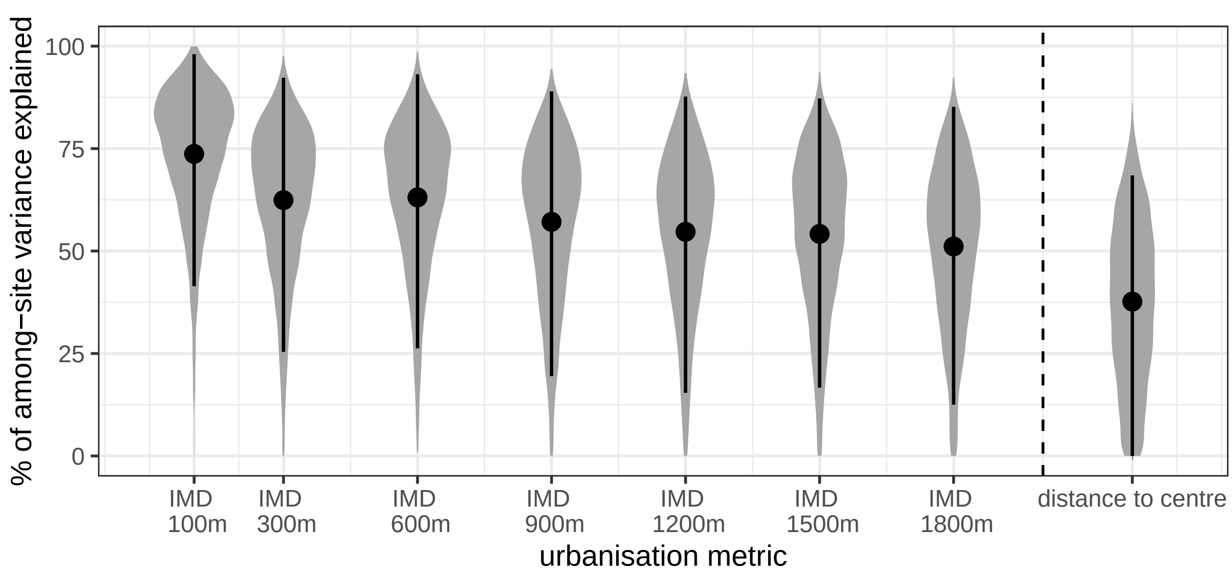


Figure S2-1. Posterior proportion of (logit-scale) site-level variance in morph frequencies explained by urbanisation, depending on the urbanisation metric used in the model. All models are the beta-binomial implementations; black dots are posterior means, segments the 95% Highest Posterior Density Intervals.

S3 - Model effect of urbanisation comparison

The standardised effect of urbanisation β_1 is quite consistent whether distance to city centroid or local Imperviousness Density is used as urbanisation metric, and regardless of the spatial scale at which Imperviousness Density is estimated (**Fig. S3-1**).

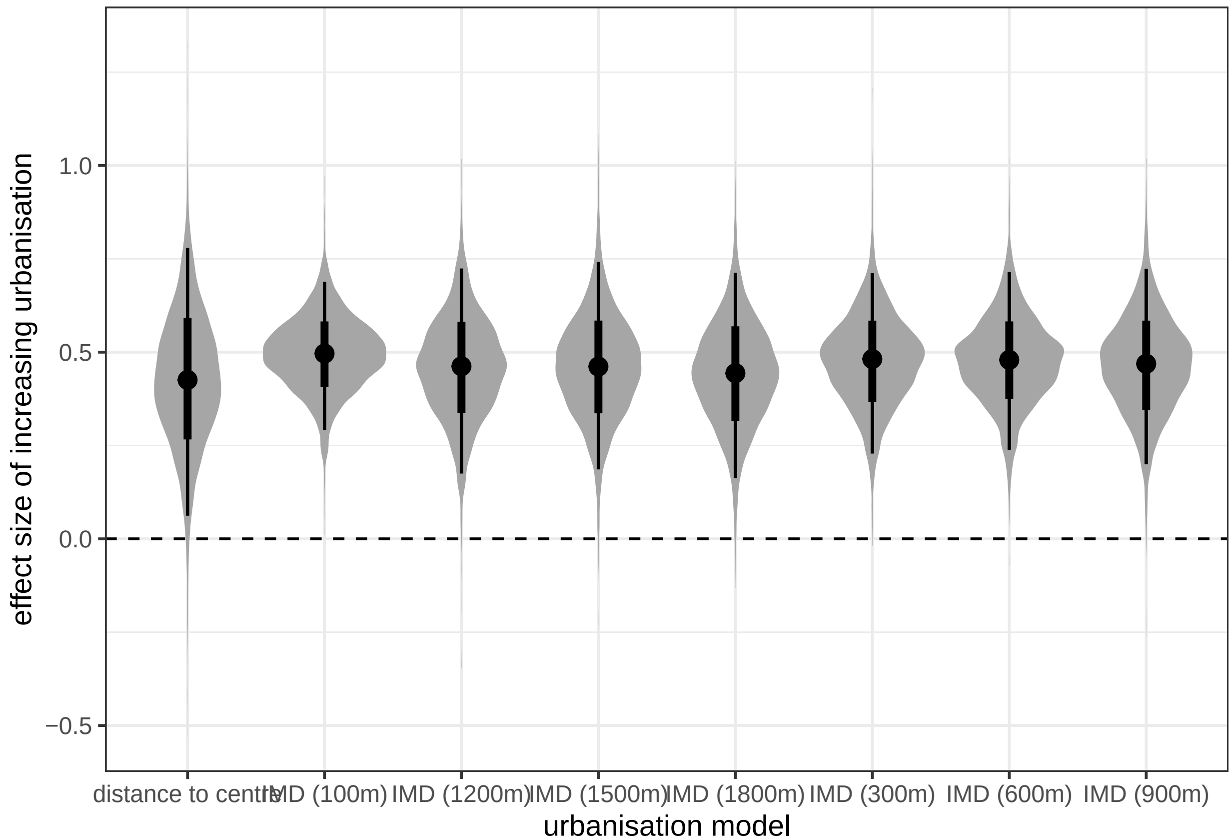


Figure S3-1. Note that the sign of the posterior values for distance to city centroid is inversed to make the comparison with the other posteriors easier (as distances to city centroid decrease when urbanisation increases).

S4 - Seasonal variation in urbanisation effect

As discussed in the main text, between-session variation in the effect of urbanisation is small. This is a strong hint that our results reflect true morph differences and not simply weather-induced seasonal variation in relative morph activity (and trappability).

This seasonal stability can be seen by looking at the predicted urbanisation effect session by session, whether the random effect values themselves (**Fig. S4-1**) or their consequences in terms of session-specific predicted values (**Fig. S4-2**). In either case, no obvious seasonal pattern can be seen.

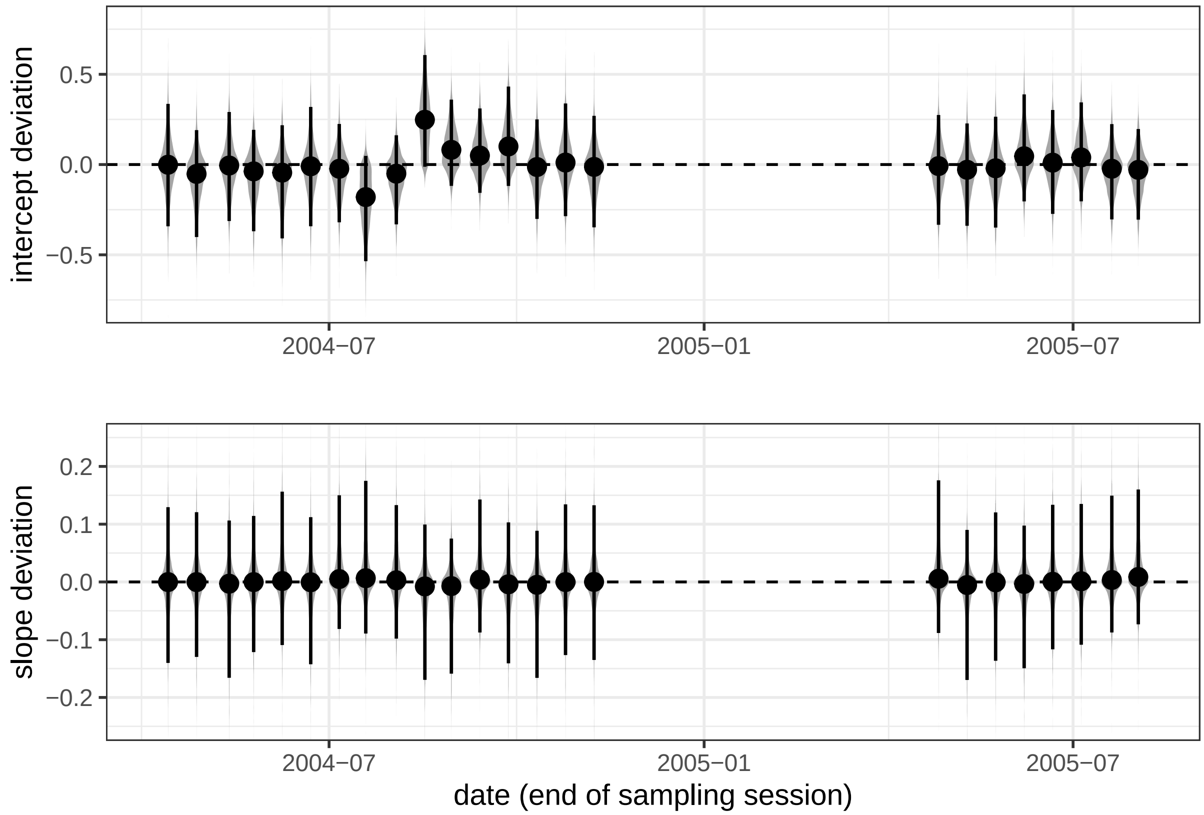

intercept deviation
0.5
0.0
−0.5
2004−07
2005−01
2005−07
slope deviation
0.2
0.1
0.0
−0.1
−0.2
2004−07
2005−01
2005−07
date (end of sampling session)

Figure S4-1

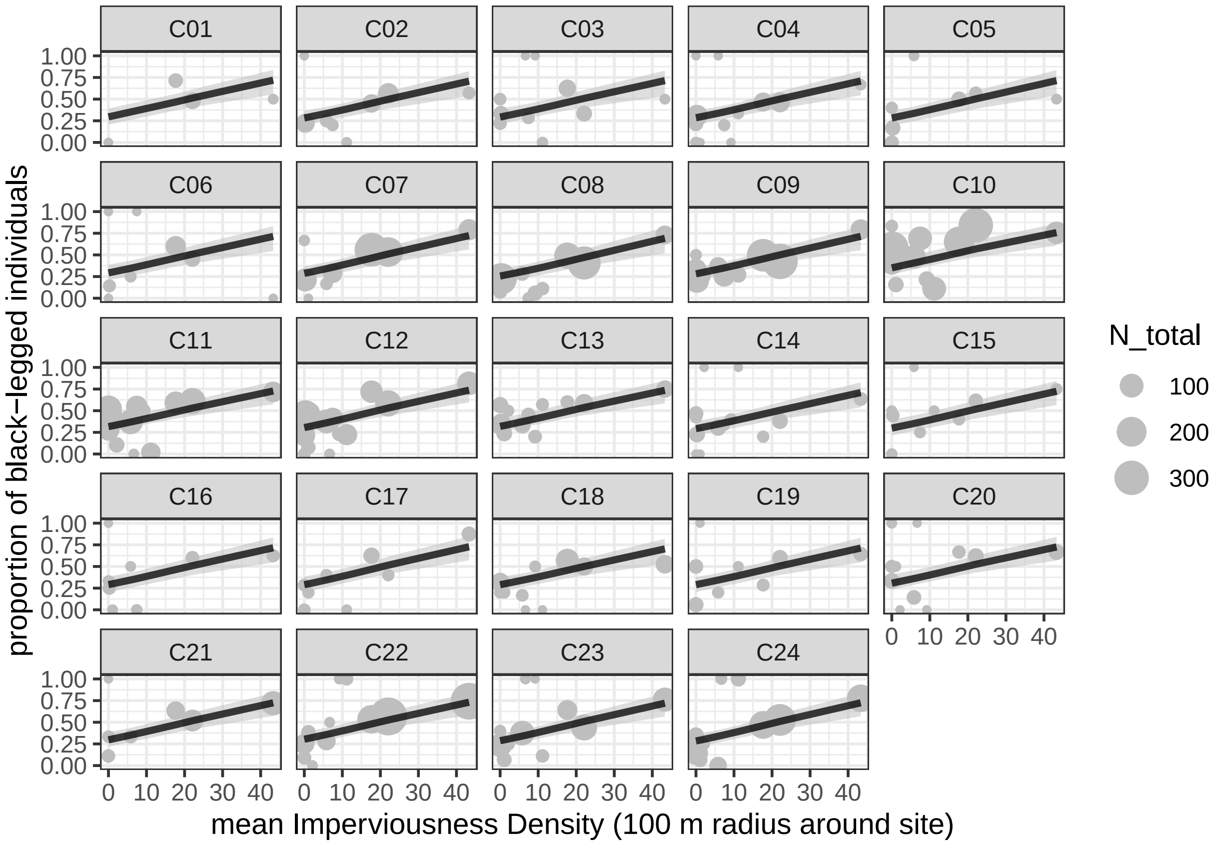


Figure S4-2

S5 - Population size variability along the urbanisation gradient

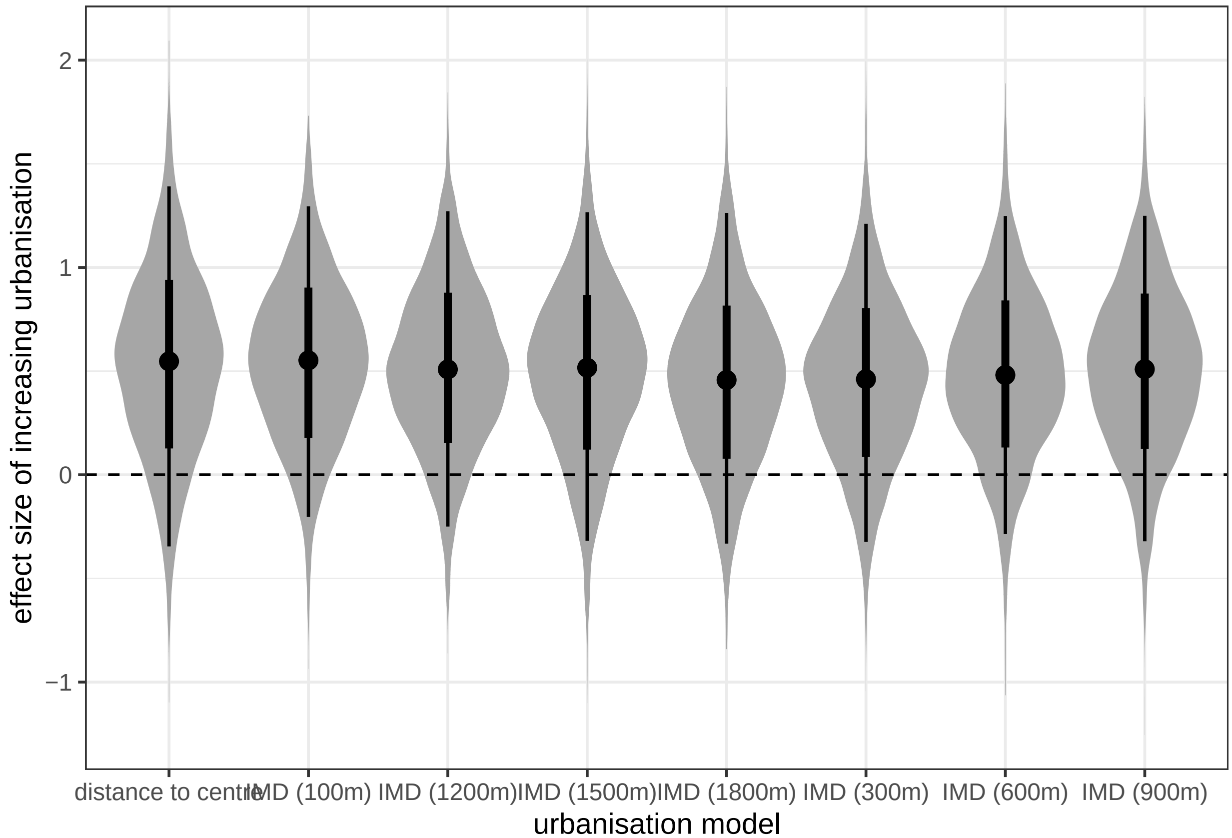
Evolutionary clines may in some cases be caused by non-adaptive forces; this includes genetic drift, in particular if a gradient in population sizes is correlated with the environmental gradient of interest . For a first check on whether our results may be explainable by drift, we looked at whether there was an effect of urbanisation on population sizes of *P. madidus* in our data.

Our models are here essentially structured the same way as the main proportion models (see main text and **S1**), with a few nuances summed up here:

- the response is now $N_{i,j}$ the total number of *P. madidus* caught by woodland i × session j combination;
- importantly, samples with 0 beetles found are **included** and not excluded;
- since data are now counts and not proportions, we use a negative binomial model (with a log link) rather than a (beta-)binomial ones;
- we include an offset or rate term to account for the fact sampling sessions are not the same length;
- the prior for the intercept is changed back to the "standard" Normal$(0, 1)$ from Normal$(0, 1.5)$, as the latter is mostly meaningful in the context of proportional data (). Other priors are left as is.

Keeping the model structure of the main models (with random effects of sessions) allows us to account correctly for the known seasonality of beetle abundances, rather than ignoring it by pooling, which may lead to potential issues.

We find that independently of the metric used, there is no clear evidence that urbanisation has an effect on population sizes (**Fig. S5-1**). We note that if there were to be an effect, the data suggests it would be towards *increased*, not reduced, population sizes in cities.



References

Harrison, Xavier A. 2015. "A Comparison of Observation-Level Random Effect and Beta-Binomial Models for Modelling Overdispersion in Binomial Data in Ecology & Evolution." *PeerJ* 3: e1114. https://doi.org/10.7717/peerj.1114.

McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan.* 2nd edition. Boca Raton, USA: Chapman and Hall/CRC.

Nakagawa, Shinichi, and Holger Schielzeth. 2013. "A General and Simple Method for Obtaining R^2 from Generalized Linear Mixed-Effects Models." *Methods in Ecology and Evolution* 4 (2): 133–42. https://doi.org/10.1111/j.2041-210x.2012.00261.x.

Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2017. "Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC." *Statistics and Computing* 27 (5): 1413–32. https://doi.org/10.1007/s11222-016-9696-4.