Supplementary material 1. Details on the red squirrel spatial distribution and land cover

We investigated potential biases in the opportunistic collection of red squirrel carcasses for genetic sampling on the Isle of Wight, focusing on the relationship between carcass collection locations and land cover types. The analysis involved dividing the island into 145 equally sized squares and collecting data on squirrel carcass counts in relation to five land cover categories: arable, grasslands, woodland, urban, and other. We aimed to assess how environmental factors and random chance may influence carcass collection, including the risk of "false absences," where squirrels are present, but no carcasses are found. Using a negative binomial regression model, the results revealed a higher likelihood of squirrel carcasses being collected in urban and forested areas, suggesting that these habitats are more heavily utilised by the population. While the dataset may not fully represent the entire squirrel population, given the inherent biases of opportunistic sampling, the findings help evaluate the consistency of the sampling process and its alignment with expected population patterns. This analysis contributes to understanding the potential biases in estimating squirrel distribution, providing insights into the challenges of using opportunistic data for ecological studies.

Methods

To assess potential biases from the opportunistic collection of red squirrel carcases for genetic sampling, the relationship between carcass collection locations and land cover on the Isle of Wight was investigated. The island was systematically divided into 145 equally sized squares (3 km^2). Data were collected from each square, with squirrel carcass counts as the dependent variable and the proportions of various land cover types as independent variables. Sampling biases may arise from environmental factors, ecological processes, random chance, or the fact squirrels in certain areas were not hit by vehicles. Additionally, there is a risk of false absences, situations where squirrels are present, but no carcasses were collected (“excess zeroes”), a common issue in ecological studies relying on opportunistic data (MacKenzie et al. 2006; Lewin et al. 2010). This means estimates of squirrel distribution based on carcass collections could be biased, necessitating statistical controls to account for these potential false absences (O'Hara and Kotze, 2010; Sirén et al., 2024), as imperfect detection can significantly skew conventional estimators of species distribution (Chen et al. 2013).

The land cover data used in the analysis were derived from the UKCEH LCM2021 (Marston et al. 2023). The analysis included five broad classes: (1) Arable; (2) Grasslands (all grassland sub-types); (3) Woodland (broadleaf and coniferous); (4) Urban and suburban; (5) Other (which combined all remaining categories, primarily coastal littoral, saltwater, and saltmarsh). Coastal squares smaller than 1 km² that yielded no carcasses were excluded from the analysis. Multicollinearity among predictors was assessed, with all Variance Inflation Factor (VIF) values below 1.6, indicating no serious multicollinearity issues. Model selection was based on the Akaike Information Criterion (AIC), where lower AIC values indicate a better model, as AIC penalises models with more parameters to prevent overfitting and encouraging simplicity.

To assess whether the data followed a Poisson distribution, a one-sample Kolmogorov-Smirnov test was conducted. The test compared the empirical data against a theoretical Poisson distribution, without assuming normality. The R packages MASS, pscl, and car, were used to facilitate this comparison to determine the best fit for the data, with results showing statistical confidence at the $p < 0.05$ level (Venables and Ripley 2002; Zeileis et al. 2008; Fox and Weisberg 2019).

To address overdispersion and excess zeroes, previous studies have employed Negative Binomial (NB) models and Zero-Inflated Negative Binomial (ZINB) models, as well as Hurdle models (Balderama et al. 2016; Lee et al. 2023; He and Huang 2024). Consequently, the performance of the NB, ZINB, and Hurdle model were evaluated using the AIC.

$$Y \sim NB(\lambda,\theta) \text{ where } \log(\lambda)=\beta_0+\beta_1 Grassland+\beta_2 Urban+\beta_3 Forest+\beta_4 Arable+\beta_5 Other$$

Results

The one-sample Kolmogorov-Smirnov test for the Poisson distribution was non-significant, suggesting that a Poisson model could be suitable. However, overdispersion was observed in the Poisson regression model (variance = 1.8, mean = 1.02, Pearson Chi-square value/df = 1.459), a common issue in ecological count data (O'Hara and Kotze 2010).

The analysis revealed that the Negative Binomial (NB) model (AIC = 371.55) and the Zero-Inflated Negative Binomial (ZINB) model (AIC = 370.59) demonstrated comparable performance, while the Hurdle model performed less favourably (AIC = 393.41). Residuals analysis for both NB and ZINB models revealed two outliers with exceptionally high residuals, which were subsequently removed from the dataset, leaving 137 squares for further analysis. Upon re-running both models, the NB and ZINB continued to perform similarly. However, the ZINB model's intercept and coefficients were not statistically significant ($p > 0.05$), indicating that landscape variables did not significantly influence the excess zeros in our data. This implies that the absence of carcass collections is not systematically related to land cover variables included in the model.

The NB regression analysis revealed significant associations between land cover types and squirrel carcass collections. Urban land cover exhibited the strongest positive association with squirrel counts (Exp(B) = 4.4349, $p<0.01$), indicating that urban environments had the highest likelihood of squirrel carcass collections among the studied land cover types. Forest areas followed closely (Exp(B) = 4.2115, $p<0.01$), also showing a high likelihood of squirrel carcass collection. Grassland areas (Exp(B) = 2.4891, $p<0.01$) and arable land (Exp(B) = 2.1932, $p<0.05$) also had significant positive associations, although their effects were less pronounced compared to urban and forest areas. In contrast, the "Other area" category did not reach statistical significance, suggesting minimal impact on squirrel carcass collections relative to other land cover types.

It is important to note that our model suggests that the zero carcass collections in the dataset are primarily true absences when habitat (land cover) is considered as a predictor. However, while the model showed a statistically significant improvement over an intercept-only model, its relatively low explanatory power (McFadden's pseudo R^2=10.5%) indicates that the variables included (land cover) may not fully capture the complexity of factors influencing squirrel presence on the Isle of Wight.

Discussion

We conducted a negative binomial regression on squirrel carcass findings relative to land cover types to examine patterns within our opportunistic sample. The analysis showed a higher likelihood of carcass presence in urban and forest areas, consistent with anticipated habitat and human activity influences. However, due to the lack of recorded absences in our data, our sample is biased toward locations where carcasses were found, which could limit the representativeness of these findings across all habitats. While not necessarily representative of the full squirrel population distribution, these results help assess whether our sampling aligns with expected patterns, thereby supporting the validity of using this sample for evaluating population viability.

Spatial analyses revealed that red squirrel carcasses were predominantly collected in urban and forested areas, suggesting that the population utilises a variety of habitats, including human-modified landscapes. This aligns with studies indicating that red squirrels can persist in diverse environments, particularly when not in competition with grey squirrels (Kenward et al. 1998; Ballingall et al. 2016). The significant associations with urban and forested areas may reflect the squirrels' ability to survive and adapt within fragmented and mixed landscapes.

However, the analysis revealed that 44.8% of the analysed squares yielded no collected carcasses. This may indicate either the absence of squirrels or that squirrels present in these areas were not deceased or hit by vehicles, rather than true population absences. Consequently, this absence of data from certain locations introduces potential biases in our sampling, which may underestimate or misrepresent squirrel distribution in less accessible or less frequently surveyed areas. This potential bias raises the possibility that habitat fragmentation or other factors could be influencing squirrel movement patterns, which reduce their likelihood of crossing roads and being detected through carcass collections.

References

Balderama E, Gardner B, Reich BJ (2016) A spatial–temporal double-hurdle model for extremely over-dispersed avian count data. Spat Stat 18:263–275. https://doi.org/10.1016/j.spasta.2016.05.001

Ballingall KT, McIntyre A, Lin Z, et al (2016) Limited diversity associated with duplicated class II MHC-DRB genes in the red squirrel population in the United Kingdom compared with continental Europe. Conservation Genetics 17:1171–1182. https://doi.org/10.1007/s10592-016-0852-3

Chen G, Kéry M, Plattner M, et al (2013) Imperfect detection is the rule rather than the exception in plant distribution studies. Journal of Ecology 101:183–191. https://doi.org/10.1111/1365-2745.12021

Fox J, Weisberg S (2019) An R Companion to Applied Regression, 3rd edn. SAGE Publications, Inc, Thousand Oaks, CA

He Q, Huang H-H (2024) A framework of zero-inflated Bayesian negative binomial regression models for spatiotemporal data. J Stat Plan Inference 229:106098. https://doi.org/10.1016/j.jspi.2023.106098

Kenward RE, Hodder KH, Rose RJ, et al (1998) Comparative demography of red squirrels (*Sciurus vulgaris*) and grey squirrels (*Sciurus carolinensis*) in deciduous and conifer woodland. J Zool 244:7–21. https://doi.org/10.1111/j.1469-7998.1998.tb00002.x

Lee KH, Pedroza C, Avritscher EBC, et al (2023) Evaluation of negative binomial and zero-inflated negative binomial models for the analysis of zero-inflated count data: application to the telemedicine for children with medical complexity trial. Trials 24:. https://doi.org/https://doi.org/10.1186/s13063-023-07648-8

Lewin W-C, Freyhof J, Huckstorf V, et al (2010) When no catches matter: Coping with zeros in environmental assessments. Ecol Indic 10:572–583. https://doi.org/10.1016/j.ecolind.2009.09.006

MacKenzie DI, Nichols JD, Royle JA, et al (2006) Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence. Elsevier/Academic Press, Burlington, MA

Marston CG, O'Neil AW, Morton RD, et al (2023) LCM2021 – the UK Land Cover Map 2021. Earth Syst Sci Data 15:4631–4649. https://doi.org/10.5194/essd-15-4631-2023

O'Hara R, Kotze J (2010) Do not log-transform count data. Nature Precedings. https://doi.org/10.1038/npre.2010.4136.1

Sirén APK, Berube J, Clarfeld LA, et al (2024) Accounting for missing ticks: Use (or lack thereof) of hierarchical models in tick ecology studies. Ticks Tick Borne Dis 15:102342. https://doi.org/10.1016/j.ttbdis.2024.102342

Venables WN, Ripley BD (2002) Modern Applied Statistics with S, 4th edn. Springer, New York, NY

Zeileis A, Kleiber C, Jackman S (2008) Regression Models for Count Data in *R*. J Stat Softw 27:. https://doi.org/10.18637/jss.v027.i08

Corresponding Author

Alyson E. V. Buchanan, Bournemouth University, Christchurch House, Talbot Campus, Poole, Dorset BH12 5BB, U.K

E-mail abuchanan@bournemouth.ac.uk