Supplementary material

Habitat preferences of European turtle dove *Streptopelia Turtur Turtur* in the Czech Republic: implications for conservation of a rapidly declining farmland species

Kryštof Korejs, Jan Riegert, Ivan Mikuláš, Jan Vrba & Jan Havlíček*
* Corresponding Author

Table S1. Land cover data used in statistical analyses. The primary category (land cover group) joined similar types of land cover for the purposes of analysing the general environmental effects on turtle dove distribution. The secondary category (land cover type) was used as the input dataset in compositional analysis of habitat use, effectively functioning as habitat types. The tertiary (CLE) category describes which ecosystems (as described by the Consolidated Layer of Ecosystems, AOPK ČR 2012) make up different types of land cover in this study. The final category describes source of information for establishing CLE category, either Habitat Catalogue of Czech Republic (VMP, see Chytrý et al. 2010) and its corresponding GIS layer (VMP, AOPK ČR 2012) or ZBG geographic layer (ČÚZK 2012) or EEA Urban atlas (EEA 2006) or Corine Land Cover data (CLC, EEA 2006). KVES categories were merged into land cover groups based on a preliminary computation of selectivity index for turtle dove inhabitation, by formula recommended by Sunde et al. 2001: mean $\log_2$ (used/available)), as well as based on their general characteristics as described in the CLE layer.

1. Land cover group	2. Land cover (habitat) type	3. CLE category name	Reference	Selectivity
Urban areas	Built-up areas	Continuous urban fabric	Urban atlas, ZBG, CLC	–0.27
		Discontinuous urban fabric	Urban atlas, ZBG, CLC	–0.11
	Parks	Urban nature	Urban atlas, ZBG	–0.82
		Artificial urban green areas – parks, gardens, cemeteries	Urban atlas, ZBG	–1.21
	Industrial	Industrial and commercial units	Urban atlas, ZBG	–1.11
		Dump and construction units	Urban atlas, ZBG	–1.81
	Recreational	Artificial urban green areas – recreation and sport areas	Urban atlas, ZBG	–1.59
	Roads	Transport units	Urban atlas, ZBG	–0.79
	Mines	Artificial rocks, mines, quarries	VMB, ZBG	–0.61
		Natural rocks	VMB, ZBG	–0.42
Agriculture	Orchards & gardens	Orchards and gardens	ZBG	–0.07
	Hop fields	Hop fields	ZBG	–0.04
	Vineyards	Vineyards	ZBG	0.30
	Arable land	Arable land	ZBG	–0.14
	Commercial meadows	Intensive grasslands	ZBG	–0.05
	Semi-natural grasslands	Alpine meadows	VMB	0.15
		Alluvial meadows	VMB	0.11
		Mesic meadows	VMB	0.28
		Dry grasslands	VMB	0.06

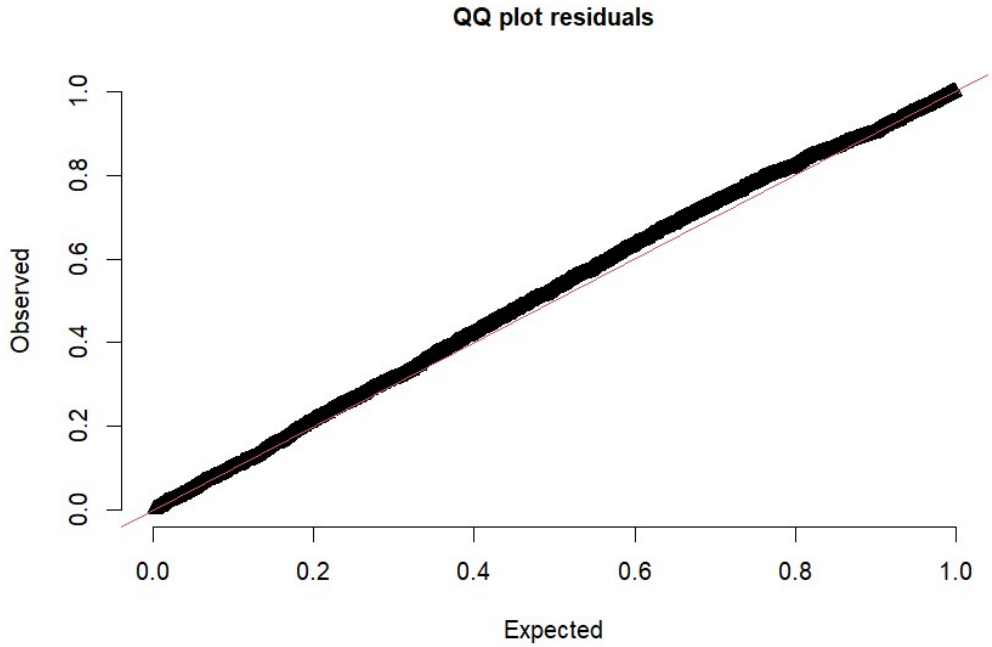
Forests and shrub vegetation	Coniferous tree plantations	Intensive coniferous forests	ZBG, VMB	0.61
	Broadleaf tree plantations	Intensive broad-leaved forests	ZBG, VMB	–0.50
	Mixed tree plantations	Intensive mixed forests	ZBG, VMB	–0.28
	Semi-natural coniferous forests	Pine forests	VMB	0.11
		Spruce forests	VMB	–0.11
	Beech forests	Beech forests	VMB	0.41
	Oak forests	Oak and oak-hornbeam forests	VMB	–0.90
	Riparian forests	Alluvial forests	VMB	0.10
	Natural shrub vegetation	Natural shrub vegetation	VMB	0.01
	Introduced shrub vegetation	Introduced shrub vegetation	ZBG, VMB	0.01
Wetlands	Wetlands	Wetlands and litoral vegetation	VMB	–0.27
		Peatbogs and springs	ZBG, VMB	–0.19
		Swamps	ZBG, VMB	–0.37
		Macrophyte vegetation of water bodies	VMB	–0.40
		Human-influenced water bodies	VMB	–0.48
		Natural water courses	ZBG, VMB	–0.35
		Anthropogenically influenced water courses	ZBG	–0.41

Table S2. Additional information the final model analysing turtle dove response to general land cover characteristics. The model was performed by the package glmmTMB (Magnusson et al. 2017). Forward selection by AIC was performed, and is elaborated upon in the table. Goodness of fit was tested by the DHARMa package and the test results are explained in the table (Hartig & Hartig 2021).

Model formula	glmmTMB (Presence ~ Agriculture * Altitude + Urban * Altitude + Forest * Altitude + Wetland + exp (pos + 0 | ID), family = binomial (link = 'logit')
Parameters	AIC: 2672; BIC: 2735; LogLik: –1325; Deviance: 2650; Pseudo-R^2: 0.151
Dependent variable	Presence: this variable specifies whether sampled squares were occupied or unoccupied (1/0) by turtle dove.
Fixed effect predictors	Agriculture: log-transformed area of agricultural land use within each sampled square.
	Altitude: the above sea level elevation median of each sampled square. Urban: log-transformed area of urban infrastructure within each sampled square. Forest: log-transformed area of forest land cover within each sampled square. Wetland: log-transformed area of aquatic wetland land cover within each sampled square.
Random effect predictors	exp (pos + 0 | ID): an exponential covariance structure that enables conversion of x y coordinates into a numeric random effect representing spatial autocorrelation. See methods for further information.
Family	family = binomial (link = 'logit'): argument specifying the family function that the model follows based on dependent variable distribution.

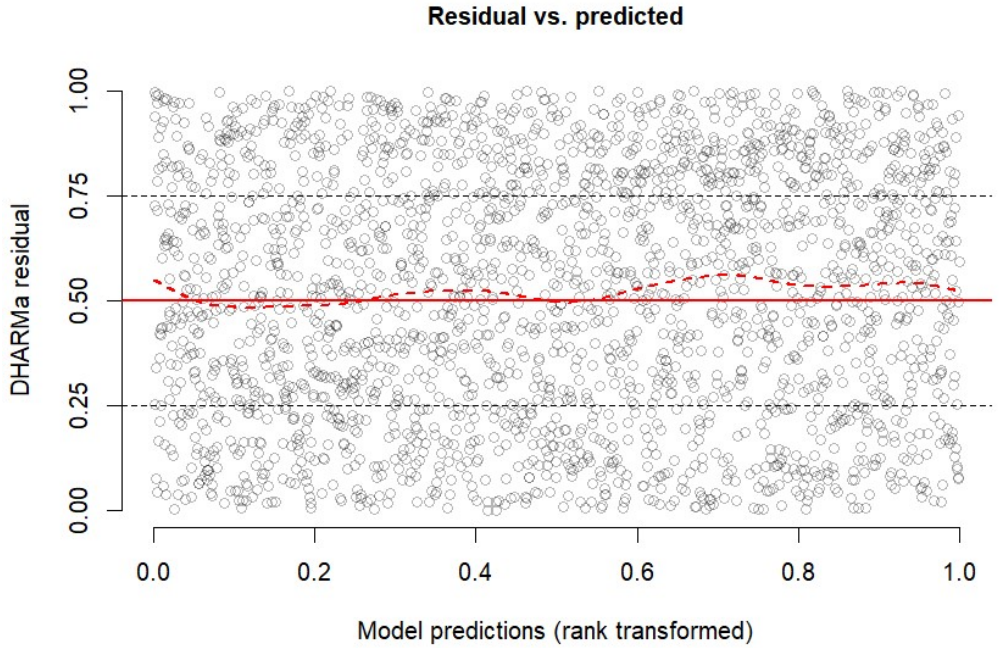
Forward selection	The forward selection involved many comparisons of complex models because of 1) the used spatial covariance structure, 2) the huge size of the input dataset and 3) the high number of tested variables. As such, the whole model selection process has been abridged to only include the final model, and the AIC of models omitting single terms (variables) from it, and the the null model. This has been facilitated by the drop1 function which is part of the basic R installation (R Core Team 2020). As we can see, dropping any terms resulted in an increase of the overall AIC, likely omitting important information that would reduce model complexity. Note that excluding a variable may still mean that it is effect is partially retained within an interaction term, which is why the increase in AIC can be similar as to when merely an interaction term is removed.			
Excluded variable	Model AIC		Excluded variable	Model AIC
1) None	2672		6) Urban	2686
2) Agriculture: altitude	2679		7) Agriculture	2688
3) Altitude	2680		8) Wetland	2693
4) Altitude: urban	2680		9) Forest	2706
5) Altitude: forest	2682		ALL (only random effect kept)	2790
Model goodness of fit	Regression diagnostics of complex glmmTMB models suffer from some statistical constraints. Specifically, our model features a very large dataset, and regression diagnostics for such a model may produce falsely significant results (Hartig & Hartig 2021). Therefore, visual inspection is necessary instead of some statistical tests that are likely to be unreliable for large datasets. Diagnostic plots will be used to visually illustrate model performance based on simulated residuals and tested parameters (Hartig & Hartig 2021).			
Relevant Parameters	Plot and results			

Distribution uniformity
Overdispersion
Outliers



Distribution uniformity: compared datasets (line: predicted values, dots: residuals) are plotted to visualize if they are from the same distribution. This figure shows little deviation from expected distribution (residuals mostly align with predicted values).
Overdispersion: visual checks can be performed for overdispersion in the variance of the response variable that is not accounted for by the model. As the variance of residuals does not appear to increase in either direction, it is likely that the data are not overdispersed. Testing overdispersion is also likely redundant for a binomial response variable (0/1, Hartig & Hartig 2021).
Outliers: this internal test is used to determine if the number of outliers is larger than expected for the model. As the simulated residuals only featured 18 outliers in 2378 observations, there is likely no effect on model results.

Homogeneity of residuals



Homogeneity of residuals: the plot comparing predicted values (central line) with residuals (dots and red dashed line) shows only slight variation between residuals and predicted values. Any inconsistencies may also be due to the very large size of the original dataset (see above). A secondary issue may occur where DHARMa produces artefacts in correlation of predicted values and residuals of the glmmTMB function (Hartig & Hartig 2021).

Table S3. Collinearity levels of variables in the mixed-effects model, tested by the multicollinearity function, package performance (Lüdecke et al. 2019). The variance inflation factor (VIF) is a measure to analyse the magnitude of multicollinearity of model terms. A VIF less than five indicates a low correlation of that predictor with other predictors. A value between five and ten indicates a moderate correlation, while VIF values larger than ten are a sign for high, not tolerable correlation of model predictors (James et al. 2013). The table also displays the associated confidence intervals, the factor by which the standard error is increased due to possible correlation with other terms, and tolerance values (including confidence intervals), where tolerance = 1/VIF. The interaction terms from the model are not included in the multicollinearity tests, as they inflate the variance inflation factor due to being linear combinations of their parent predictors (Lüdecke et al. 2019).

Variable	VIF	VIF 95% CI	Increased SE	Tolerance	Tolerance 95% CI
Agriculture	1.74	[1.65, 1.85]	1.32	0.65	[0.54, 0.61]
Forest	1.73	[1.64, 1.84]	1.32	0.58	[0.54, 0.61]
Urban	1.51	[1.44, 1.60]	1.23	0.66	[0.62, 0.70]
Altitude	1.49	[1.41, 1.57]	1.22	0.67	[0.64, 0.71]
Wetland	1.03	[1.01, 1.12]	1.02	0.97	[0.89, 0.99]

Literature

AOPK ČR 2012: Biotope mapping layer (electronic georeferenced database). Distribution of natural and near-natural habitats in the territory of the Czech Republic. *Agency for the Protection of Nature and Landscape of the Czech Republic, Praha, Czech Republic. (in Czech)*

Chytrý M., Kučera T. & Kočí M. et al. 2010: Catalogue of habitats of the Czech Republic, 2nd ed. *Nature Conservation Agency of the Czech Republic, Praha, Czech Republic.*

ČÚZK 2012: Fundamental base of geographic data (ZABAGED®) (digital data ESRI Shapefile), (1:10 000). *Czech Geodetic and Cadastral Office, Praha, Czech Republic. https://geoportal.cuzk.cz/ (in Czech)*

EEA 2006: Urban Atlas 2006 (digital data ESRI Shapefile), (1:10 000). *European Commission, Directorate-General Entreprise and Industry, Directorate-General for Regional Policy, Brussels, Belgium. http://www.eea.europa.eu/data-and-maps/data/urban-atlas*

EEA 2006: Corine Land Cover (digital data ESRI Shapefile), (1:100 000). *European Environment Agecy, Kobenhaven, Denmark. http://www.eea.europa.eu/data-and-maps/data/clc-2006-vector-data-version*

James G., Witten D., Hastie T. & Tibshirani R. 2013: An introduction to statistical learning: with applications in R. *Springer, New York, USA.*