



# No effect of model distribution on long-term trends, even with underdispersion

Thomas Schneider-Bruchon<sup>a,\*</sup>, Sabrina Gaba<sup>b,c</sup>, Vincent Bretagnolle<sup>a,c</sup>

<sup>a</sup> UMR 7372 Centre d'Etudes Biologiques de Chizé, CNRS & La Rochelle Université, 79360 Villiers-En-Bois, France

<sup>b</sup> USC 1339 Centre d'Etudes Biologiques de Chizé, INRAE, CNRS & La Rochelle Université, 79360 Villiers-En-Bois, France

<sup>c</sup> LTSER "Zone Atelier Plaine & Val de Sèvre", 79360 Villiers-En-Bois, France

## ARTICLE INFO

### Keywords:

General linear mixed model  
Long-term trend  
Model distribution  
Underdispersion  
Bayesian information criterion  
Skylark

## ABSTRACT

While many studies have illustrated the decline of animal populations—particularly of farmland birds—the statistical analyses, design, and protocols used have raised some concerns and criticism. Using a 27-year dataset (1996–2022) based on recording the number of skylarks (*Alauda arvensis*) at 160 longitudinal count points, our study confronts two approaches commonly used to model long-term trends. The first uses a single model (based on a priori ecological knowledge), while the second is an a posteriori approach that relies on a multi-model selection among candidate models that account for probability distributions to describe the error structure. Here we investigate whether the statistical distribution of modelled variables and the method of including covariates in the model affect trend estimates. With a large amount of data and in the case of underdispersion, we found that the model distribution used had no impact on the estimation of the long-term trend. Moreover, adding confounding covariates did not change or improve the trend estimation, at least when data were obtained from a well-designed protocol (our case). In contrast to other studies reporting an effect of the model's distribution on long-term trends, especially in the presence of overdispersion, our results offer a new perspective on the presence of underdispersion, where simple models perform equally well as complex ones. Further research is now needed on multiple species data or on smaller data sets to check the generality of our findings.

## 1. Introduction

Over the past 30 years, numerous studies have highlighted the global decline of animal abundance worldwide, which underpins a general biodiversity crisis (Chapin III et al., 2000; Vitousek et al., 1997). Such changes, detected through long-term monitoring, concern all taxa, including insects (Forister et al., 2019; Wagner, 2020), birds (Bowler et al., 2019; Rosenberg et al., 2019), mammals (WWF, 2022) and freshwater megafauna (He et al., 2019). However, detecting a trend by repeating noisy local abundance measurements over time can be tricky; consequently, several modelling techniques have been developed to detect the signature of environmental changes and anthropogenic impacts on animal populations. Several assumptions surround these modelling techniques, some of which may not be respected over time, which is a potential issue when dealing with long-term trends. For instance, survey techniques or design may have changed throughout studies using long-term datasets, leading to missing data that are usually ignored or removed from the analysis (Łopucki et al., 2022). These

changes may affect the statistical power of the models used to detect long-term trends and may lead to biased estimation of parameters by violating the assumptions underlying the modelling approach. As a result, weaknesses in statistical analyses (Buschke et al., 2021; Desquilbet et al., 2020; Fraser et al., 2018) or in data collection designs (Kamp et al., 2016) have been highlighted and used to question the validity or robustness of some population declines.

Much attention has been given to the key issue of data distribution, as specific assumptions are made in models in this regard, and failure to meet these assumptions can lead to a biased estimation of parameters (Freckleton, 2009; Grafen and Hails, 2002). Count data are traditionally associated with the Poisson distribution (Fokianos, 2012; James and Sahir, 2022; Zuur et al., 2007), which has a single parameter ( $\lambda$ ) and assumes equal variance and mean (i.e., equidispersion). However, count or density data may show either over/underdispersion when the variance is higher or lower than the mean, respectively. Differences between the observed and expected variance may result from deviations from complete spatial randomness (Lynch et al., 2014; Xekalaki, 2006), e.g.

\* Corresponding author at: Centre d'Etudes Biologiques de Chizé – CNRS, 79360 Beauvoir-sur-Niort, France.

E-mail address: [thomas.schneider.bruchon@gmail.com](mailto:thomas.schneider.bruchon@gmail.com) (T. Schneider-Bruchon).

territoriality resulting from intra- and/or interspecific competition leads to a regular spatial distribution of individuals or pairs, and thus underdispersed data (Cornulier and Bretagnolle, 2006; Svårdson, 1949). Conversely, clustered distributions of individuals generally produce overdispersed data (Hilbe, 2011). Another source of deviation from the basic assumptions is the measurement error, which generally leads to overdispersion of the observed data (Guo and Li, 2002), though small measurement error compared to the abundance may yield underdispersion (Lynch et al., 2014). Ignoring over/underdispersion can result in poor interpretations and misleading conclusions (Campbell, 2021; Richards, 2008; Stoklosa et al., 2022; Harrison, 2014). First, confidence intervals in overdispersed data tend to be too narrow and the  $p$ -value overoptimistic, leading to the detection of a false positive effect (i.e., an inflated Type I error). The opposite is true for underdispersion, where the main problem is a loss of power which may lead to the non-detection of a positive effect (i.e., a Type II error) when it does actually exist (Forthmann and Doebler, 2021; Hartig, 2022). In the particular case of trend estimates, failure to take into account overdispersion can lead to distortions in their evaluation (Tirozzi et al., 2022). Underdispersion is perceived as less problematic due to its conservative inference (resulting from overestimated standard errors), and has thus been less of an issue in the literature (Forthmann and Doebler, 2021). However, the inability to detect truly significant trends is a critical issue for the design of efficient management strategies for conservation-dependent species (Tirozzi et al., 2022).

Therefore, the presence of over/underdispersion suggests that a change in modelled distribution should be applied. Classical alternatives for handling overdispersion are the quasi-Poisson and negative binomial (NB) distributions (Hoef et al., 2007; Johnson, 2012; Lindén and Mäntyniemi, 2011), while the generalized Poisson (GP) and the Conway-Maxwell-Poisson (CMP) distributions are commonly suggested to deal with underdispersion (Harris et al., 2012; Sellers and Shmueli, 2010), although they are capable of dealing with overdispersion as well (Sellers and Morris, 2017). However, changing the model distribution can lead to more complex and time-consuming models, a concern given the exponential increase in the amount of data to be processed (Farley et al., 2018). Furthermore, large data sets will produce, on average, smaller  $p$ -values (Johnson, 1999; White et al., 2014), and thus the dispersion test on the model residuals will yield significant rejection of the null hypothesis (i.e., equidispersion) as it is unlikely to have a perfectly fitted model (Hartig, 2022). This raises the question of when over/underdispersion becomes an issue and needs to be addressed, and in particular whether the cost of using more complex distributions offsets the cost of accuracy loss in trend estimates (Gilad-Bachrach et al., 2003).

In addition, modelling often involves the use of confounding covariates in the statistical model to avoid bias. The way in which covariates are included in the model is also subject to two different approaches, either based on a priori ecological assumptions or a posteriori statistical analyses. In the former, covariates are fitted with their supposed effects based on assumptions about the ecological mechanisms at play. A single a priori model is therefore tested, regardless of its fit to the data, leading to major differences depending on the study model, authors, and hypotheses. For example, some studies deliberately ignore certain confounding variables, such as weather (Harrison et al., 2014), while others include each variable as a linear covariate (Bas et al., 2008; Ellis and Taylor, 2018). Such an approach also reinforces the use of covariates that are often already well known, which can lead to a circular arguments (Adde et al., 2023). In the alternative approach, modellers typically perform a covariates model selection by comparing models' parsimony, i.e., the balance between simplicity and goodness of fit (Aho et al., 2014), to identify the "best" subset of all possible combinations. The choice can be automated or performed manually and is based on statistical decision procedures such as information criteria (e.g., the Akaike information criterion [AIC], see Akaike, 1973; the Bayesian information criterion [BIC], Schwarz, 1978) (Briscoe et al., 2021; Gabriel

et al., 2022; Kouba et al., 2021; Öberg et al., 2015), or  $R^2$  comparison (Vergara and Pablo, 2007). These different methods may lead to different results when applied the same dataset: for example, the BIC will select simpler models than the AIC as its penalty factor is larger (Heinze et al., 2018). There is no simple rule for choosing between the available options, since the optimal model for the purpose of prediction may be different from the selected "true" or "correct" model (Chakrabarti and Ghosh, 2011; Shmueli, 2010). In addition, some covariates may have nonlinear effects that require modelling using various alternative shapes (e.g., with several polynomial orders), making selection even more difficult and time consuming (Tredennick et al., 2021). Consequently, there is a trade-off between model parsimony (and presumably, robustness) and computation time.

Here, we used a long-term data set to contrast these two opposite strategies for modelling trends, namely a single model approach (based on a priori ecological knowledge) vs. a posteriori multi-model comparison approach. We used birds as study models since bird monitoring schemes are widespread and often extend over a long period (e.g., above 50 years) (Sauer et al., 2013). We focused on the European skylark (*Alauda arvensis*), a species for which the probability of detection (Newson et al., 2013) is not an issue. Focusing on a single species further allowed us to avoid the difficulties involved in modelling trends for several species with different abundances (Freckleton, 2009; Leung et al., 2020). We estimated the long-term trend for skylark abundance in a 450 km<sup>2</sup> study area in which abundance data have been obtained for this species since 1996 (i.e., for 27 years) using general linear mixed models (GLMM). The single model approach endorsed a classical Poisson distribution for model counts, used covariates fitted with their supposed a priori effects, and was expected to be faster in computing time. The multi-model approach compared several distributions (i.e., Poisson, NB, GP, and CMP) to account for possible over/underdispersion and included several confounding variables (tested up to their cubic polynomial order) to select the most parsimonious model among the candidates, based on the BIC (Schwarz, 1978). The main aim of this study is to assess how modelling choices affect trend estimates and their level of precision (measured by standard error) considering computation time and the modelling skills required.

## 2. Materials and methods

### 2.1. Study site and species

This study was conducted in the Long-Term Social Ecological Research site Zone Atelier 'Plaine & Val de Sèvre' (LTSER-ZAPVS) (South of Department Deux-Sèvres, Central Western France, 46°14'N, 0°24'W). The area is an agricultural landscape of 450 km<sup>2</sup> mainly dedicated to the intensive production of cereals. It is an open landscape with a few hedges and forest fragments, and about 10% of the surface is covered by small to medium size villages (Bretagnolle et al., 2018). Species of the EU Bird Directive are present in the LTSER-ZAPVS, which led to the designation of half of the study area (207.6 km<sup>2</sup>) as a Specially Protected Area in 2004 (Plaine de Niort Sud-Est, FR5412007).

Our study focused on the skylark, Europe's most abundant farmland bird (Koleček et al., 2015). The skylark is a widespread farmland bird which lives in open habitats, especially in mixed arable fields and meadows, where its population has strongly declined in Europe (−58%, 1980–2021; PECBMS 2022). The species is sensitive to farmland management and is often used as an indicator species (Wakeham-Dawson, 1995; Csikós and Péter, 2020). It usually breeds twice each spring, with an early nesting in April and a late nesting between the end of May and the beginning of June (Delius, 1965). Both population size and singing activity are therefore expected to display a strong seasonal pattern in this species (Hoffmann et al., 2016). Our dataset should be able to capture such seasonality since the counting period extends over 91 days from April 2 to July 2, thus covering the entire skylark breeding season at this latitude (Delius, 1965). Moreover, skylark individuals have a very

loud, detectable, and identifiable song (Chamberlain and Gregory, 1999; Donald, 2004), so there is very little chance of missing a singing skylark if it is present at a site or of confusing it with other species; no other birds in our study site sing in flight, except the tree pipit (*Anthus trivialis*) and the whitethroat (*Sylvia communis*) which do not sing high in the sky and favour hedges, which are avoided by skylarks.

## 2.2. Sampling scheme and protocol of count data

Since 1995, skylark populations have been monitored using a network of 160 permanent point counts (Bretagnolle et al., 2018). The sampling design divided the study area into eight sectors, each with two transects (Brodier et al., 2014) bearing ten points separated by at least 500 m (Fig. 1).

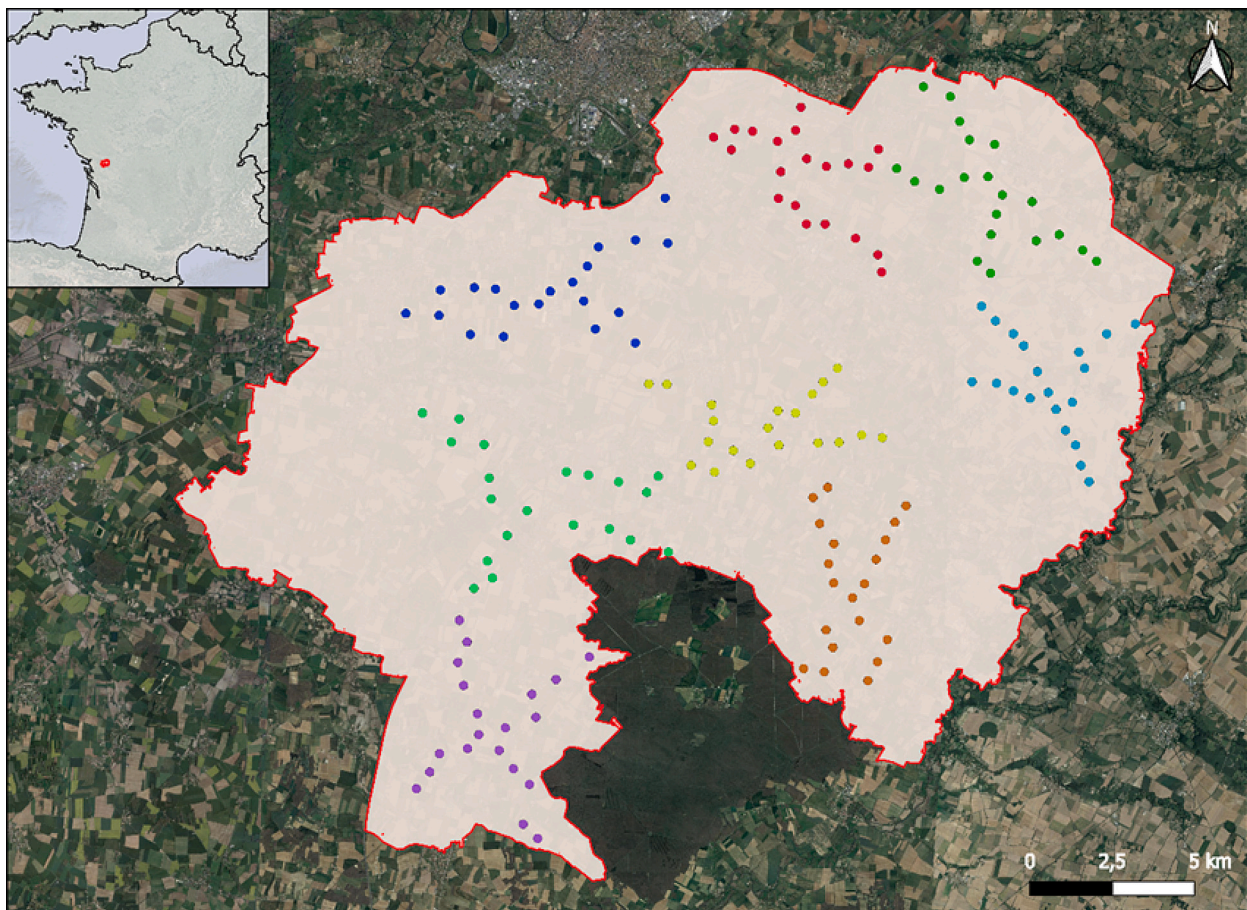
On each survey point, trained ornithologists counted birds within a precise radius of 200 m. This radius was selected to reduce detectability biases and avoid observations overlapping between two neighbouring points. The approximate position and behaviour (singing, flying, or on the ground) of each bird was noted on survey maps. Rainfall (0 or 1), wind speed (0 to 3), and cloud cover (0 to 8) were also recorded. However, observers did not carry out counts in very unfavourable conditions (fog or heavy rain, or constant wind above 10 m/s). Counts were performed in the morning (95% of the counts between 7 a.m. and 11 a.m.) once in the breeding season (usually between April 10 and May 10) or twice since 2006, with a second survey added between May 10 and June 25. Due to various technical and funding constraints, protocols were slightly modified over the course of the study period: in addition to changes in the number of surveys per year, the point count duration was

also changed from 5 to 10 min. Details on the protocols and changes per year are given in ESM (Table S1). The first counts started in 1995, but we excluded this first year from the survey because most counts were carried out outside the optimal conditions, particularly very late in the season (July) and in the late morning (90% of the counts after 10 a.m.).

## 2.3. The basic model

All statistical analyses and modelling were conducted with the R software (v. 4.0.2, R Core Team, 2020), using a Lenovo AMD Ryzen 7 quad-core computer.

To estimate trends in skylark abundance between 1996 and 2022, we produced generalized linear mixed effects models (GLMM) using the glmmTMB package (Brooks et al., 2017). All our models are derived from a basic model, with the same random and fixed effects structure. To consider all counts (sessions 1 and 2), we included the site ID as a random factor to account for repeated measurements (Bolker et al., 2009; Zuur et al., 2009), while we further accounted for random variation in trend per count site by nesting the year within the site ID (year|site\_id). Although the observer effects have little impact on long-term trends (Kendall et al., 1996; Eglington et al., 2010), we kept this information by including the observer ID as a random effect. We included a spatial term as a fixed effect in the form of simple X, Y, and X\*Y coordinates to account for spatial autocorrelation (Dormann et al., 2007; Tirozzi et al., 2021; see ESM Appendix A). Counting time was also included as a fixed factor in the models to account for its variation throughout the study period (i.e., it varied from 5 to 10 min across the years), which can affect counts (Bonthoux and Balent, 2011; Fuller and



**Fig. 1.** Location of skylark counting sites within the Long-Term Social Ecological Research site 'Zone Atelier Plaine & Val de Sèvre' in the department of Deux-Sèvres, Western France. Each colour represents one of the eight sectors. In each sector, count points are distributed along two crossing transects, each bearing ten points separated by about 500 m.

Langslow, 1984; Leu et al., 2017). Finally, the year was systematically modelled as a fixed linear trend effect to allow comparison between models.

2.4. Ad hoc comparison between two modelling approaches

The basic model is common to the two modelling approaches, which otherwise differ in the choice of distribution, covariates, and criteria for model selection. The first modelling approach, the a priori single model approach, used the Poisson distribution which is most commonly used for the analysis of count data (Lynch et al., 2014; Ma et al., 2012), to which covariates that could affect the number of skylarks according to their supposed a priori effects were also added. Except for rain, which was included as a factor (presence vs. absence), weather conditions (i.e., cloud cover and wind) were included as numeric linear effects, as is commonly done for weather covariates of long-term trends (Bas et al., 2008; Sokos et al., 2016). Temperature during each count was approximated by retrieving hourly temperatures from the NOAA database using the R ‘rnoaa’ package and using linear interpolation to obtain temperature at the one-minute precision scale (see ESM Appendix B). As warmer-than-average conditions during the breeding season are favourable for skylarks (Brodier et al., 2014), we used a linear effect to model it. To account for a potential relationship between skylark abundance and season, we included the Julian date (where 1 = January 1) as a quadratic term because of the expected seasonal trend in skylark abundance during the breeding season (Hoffmann et al., 2016). Lastly, to standardise the time at which counts were performed throughout the season, we used the difference between the effective counting hour and the sunrise hour (hereafter referred to as ‘hour’) as a linear effect because the number of skylarks counted is expected to be higher at sunrise than in the late morning (Blake, 1992; Verner and Ritter, 1986).

For the multi-model approach, we first assessed the quality of fit of four distributions: Poisson, negative binomial (NB), generalized Poisson (GP), and Conway-Maxwell-Poisson (CMP) distributions. Then we selected the covariates of interest for each of these four models, by comparing the basic model with every possible model, including the covariates in their different forms (linear, quadratic, or cubic). Model selection was performed manually.

For both approaches we centred and standardised all variables in the models to improve the estimation of linear effects in the presence of polynomials and the interpretability of the regression coefficients (Schielzeth, 2010), except for rain which was coded as a factor.

We used the BIC (Schwarz, 1978), rather than the AIC to compare candidate models as the AIC is more appropriate for finding the best model for accurate prediction, whereas the BIC is better suited to comparing a few specific models (Aho et al., 2014; Chakrabarti and Ghosh, 2011; Dziak et al., 2020). We also checked the distribution of the residuals, since this can reveal over/underdispersion or zero-inflation, using the *Dharma* package (v 0.4.5, Hartig, 2022). For the zero-inflation test, the package produces a ratio that compares the observed number of zeros with the zeros expected from simulations: ratios <1 and > 1 indicated that the observed data had fewer or more zeros than expected, respectively. For dispersion, simulation-based tests were performed to compare the observed raw residuals’ variance against the simulated residuals’ variance: a ratio = 1 indicates equidispersion, < 1 indicates underdispersion, and > 1 indicates overdispersion. We further compared the computation time required for the various models, using the R function *difftime* from the base R commands. The two modelling approaches were therefore compared according to their BICs, residual distributions, zero-inflation, and computing times. Finally, we compared both approaches in terms of their trend estimates and standard errors.

3. Results

3.1. Choice of distributions: statistical fit vs. computation time

We found that the model using CMP distribution was the most parsimonious according to BIC ranking (Table 1). The difference in the BIC value with the next best-fit distribution (i.e., GP) was 157. The initial Poisson model was the third best model, with a ΔBIC of 476 compared to CMP. The residuals of each of the four models were underdispersed, but there was a clear difference between the GP and CMP models on one side and the NB and Poisson models on the other (ratio GP & CMP ~0.735 vs. ratio NB & Poisson ~0.615), also for the p-value (GP & CMP ~ 0.02 vs. NB & Poisson <0.001; see Table 1).

In addition, the observed data had fewer zero counts than expected with Poisson and NB distributions (zero-inflation test: ratio = 0.76, p-value <0.001; see Table 1) but not with the other two distributions. Finally, the computation time of the Poisson model was the lowest by far; the computation time of the other models was at least 15 times longer (Table 1). Despite the CMP distribution model exhibiting the best fit, it involved a computation time 650 times longer than that of the Poisson model. The GP distribution may represent a good compromise, as it produced a more parsimonious model (ΔBIC = 319) and maintained a reasonable computation time (only 15 times longer than the Poisson model).

3.2. The effects of using covariates

Adopting the single model approach meant using a Poisson distribution and integrating a set of covariates chosen based on a priori ecological knowledge and assumptions. The calculation time of this model was double that of the initial Poisson model, and it also exhibited decreased fit quality (ΔBIC = 36, see Table 2). With regard to the residuals, the addition of the covariates did not compensate for the underdispersion (dispersion test: ratio = 0.61, p-value <0.001; Tables 1 & 2) or the lack of zero (zero-inflation test: ratio = 0.76, p-value <0.001; Tables 1 & 2) present in the initial model.

When adopting the multi-model approach, all covariates up to their third polynomial order were compared with the exception of rain, for which only presence was tested. Running all putative models (i.e., four distributions, five covariates with three shapes and one with a single shape), we thus produced 64 models in total, which required about 11 h of calculation. No weather variables entered the final model, as the BIC was always lowest when these covariates were absent. Only one covariate was systematically retained: the Julian day, as a linear effect. The hour, with a quadratic effect, was also retained, but only in models

Table 1

Summary table of the four initial models with different distributions: Poisson, negative binomial (NB), generalized Poisson (GP), and Conway-Maxwell-Poisson (CMP). The Poisson model is the reference with a BIC of 25,107 and a computation time of 3 s. The ΔBIC is calculated for all models relative to the model reference. Disp indicates the dispersion parameter (values <1 indicate underdispersion); P-val\_D indicates the p-value of the residual dispersion test; Z\_inf indicates the zero-inflation parameter (values <1 indicate that the observed data have fewer zeros than expected; values >1 indicate more zeros than expected). P-val\_Z indicates the p-value of the zero-inflation test; the right column indicates the ratio of calculation time between the model and the reference model.

DISTR.	ΔBIC	R. dispersion		Zero-inflation		Computation Time ratio
		Disp	P-val_D	Z_inf	P-val_Z	
Poisson	0	0.62	< 0.001	0.76	< 0.001	1
NB	+9	0.61	< 0.001	0.77	0.02	15
GP	-319	0.74	0.03	0.98	0.89	15
CMP	-476	0.73	0.01	1.04	0.70	650

**Table 2**

Summary table comparing the effect of including covariates in the two approaches (APP, single model approach and multi-model approach). Covariates are absent from the model (0), or present as a linear effect (1), quadratic effect (2) or cubic effect (3). The initial Poisson model (see Table 1) is the reference, with a BIC of 25,107 and a computation time of 3 s. The  $\Delta$ BIC is calculated for all models relative to this reference. Disp indicates the dispersion parameter (value <1 means underdispersion); P-val\_D indicates the *p*-value of the residual dispersion test; Z\_inf indicates the zero-inflation parameter (value <1 means that the observed data has fewer zeros than expected / value >1 more zero than expected); P-val\_Z indicates the *p*-value of the zero-inflation test; CT indicates the ratio of calculation time between the model and the reference model.

APP.	DISTR.	COVARIATES						$\Delta$ BIC	R. dispersion		Z-inflation		CT
		Rain	Cloud	Wind	Hour	T°	Date		Disp	P-val_D	Z_inf	P-val_Z	
Single model	Poisson	1	1	1	1	1	2	+36	0.61	< 0.001	0.76	< 0.001	2
	Poisson	0	0	0	0	0	1	-5	0.62	< 0.001	0.76	0.02	1
Multi-model	NB	0	0	0	0.0	0	1	+4	0.62	< 0.001	0.76	< 0.001	10
	GP	0	0	0	2	0	1	-328	0.74	0.04	0.98	0.90	5
	CMP	0	0	0	2	0	1	-487	0.73	0.06	1.05	0.72	800

using distributions that managed underdispersion, i.e., GP and CMP (Table 2).

Adding covariates through the model selection procedure improved (decreased) the BIC in all four distributions (Poisson = 5, NB = 4, GP = 9 and CMP = 11, see Tables 1 & 2). Changes in computation time, on the other hand, were variable: it remained stable for the Poisson model, decreased for the NB (by 33%) and GP (66%) models, and increased for the CMP model (~25%, see Tables 1 & 2). With regard to the residuals, the addition of covariates did not compensate for the underdispersion in the Poisson, NB, and GP models (dispersion test: *p*-value <0.05; Table 2) but did so for the CMP model (dispersion test: *p*-value >0.05, Table 2). Finally, the addition of covariates did not affect the lack of zero, which was always present in the Poisson and NB models (zero-inflation test: ratio = 0.76, *p*-value <0.05; Table 2) and always absent in the GP and CMP models (zero-inflation test: *p*-value >0.05; Table 2).

### 3.3. Comparison between the two approaches and consequences on estimated trends

The a priori single-model approach led to a much less parsimonious model than the multi-model approach ( $\Delta$ BIC = 523, Table 2), but it ran 400 times faster than the simple final model of the multi-model approach, and around 20,000 times faster when considering all the models produced in the latter approach with covariates. The multi-model approach resulted in equidispersed residuals (dispersion test: *p*-value >0.05; Table 2), whereas the approach based on a single a priori model led to underdispersed residuals and a lack of zero (dispersion test and zero-inflation test: ratio < 1 and *p*-value <0.001; Table 2).

However, despite these strong differences, both approaches yielded similar and significant negative trends of skylark abundance, with almost identical year coefficient estimates ( $\beta$  estimates: single model =  $-0.0147 \pm 0.0030$  and multi-model =  $-0.0138 \pm 0.0027$ ), with a difference in the standard error of <0.001 between the models (Table 3). Indeed, there was no statistical difference between the estimates (*t*-test: *t* = 0.24, *p*-value > 0.05; see ESM Appendix C). Using predicted values from the GLMM model and contrasting the two approaches, the decrease in abundance between 1996 and 2022 varied from 30.12% to 31.82%, i.e., a variation of one twentieth of the trend (Fig. 2).

**Table 3**

Summary of generalized linear mixed models (GLMM) used to assess skylark population trends under both approaches.  $\beta$  = estimate of the regression coefficient for the explanatory variable 'Year'; SE = standard error of  $\beta$ ; T% = percentage of change in population from 1996 to 2022 according to the prediction of the GLMM model.

APPROACHES	$\beta$	SE	<i>p</i> -value	T%
Single model Poisson Distr.	-0.0147	0.0030	< 0.001	-31.82
Multi-model CMP Distr.	-0.0138	0.0027	< 0.001	-30.12

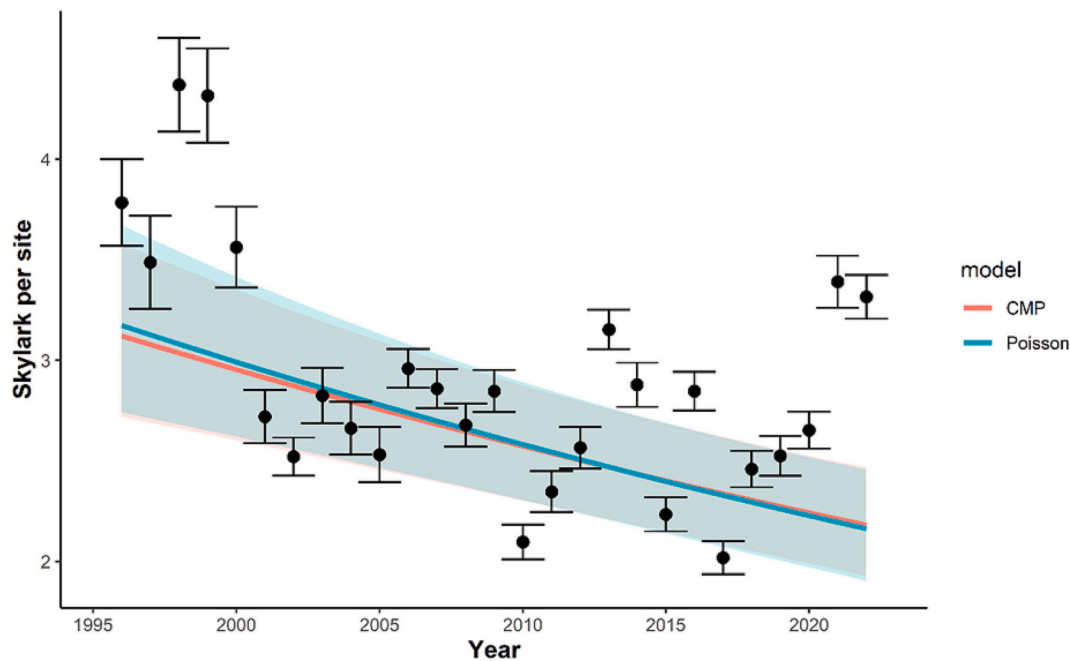
## 4. Discussion

In this study, we investigated whether modelling choices, in particular the method used to select covariates and the model distribution, affected long-term trend estimates for a skylark population and their level of precision. We contrasted two extreme and opposite modelling strategies: one based on a priori ecological assumptions (a single model approach) and the other based on a posteriori statistical properties (a multi-model comparison).

We found a clear trade-off between the parsimony of the model and its complexity, measured here by computation time: the better the model fits the data, the higher the computation time. Indeed, computation time for the single model was up to four orders of magnitude faster than for the multi-model approach. It may be possible to overcome this computation time drawback through the use of high-performance computers (Ruan et al., 2017; Tirozzi et al., 2022) or parallelization techniques (Polanco-Martínez and López-Martínez, 2021). However, this requires access to this type of computer, along with additional modelling skills.

The single-model approach, using a Poisson distribution, revealed the presence of underdispersion. This is likely a result of the regular spatial distribution of the skylarks. Indeed, the study site is an agricultural landscape with few hedges and forest fragments, and counts were carried out at sites that are favourable for skylarks (no counts were carried out in or close to woodland or urban areas), which led to a uniform presence of the species at all counting sites. Indeed, at least one skylark per year was present at 98% of the surveyed sites. As a result, we found a deficit of zero counts in our model. In addition, skylarks are easily identifiable and detectable (Chamberlain and Gregory, 1999; Donald, 2004), which considerably reduces the risk of error and may lead to underdispersion (Lynch et al., 2014). The multi-model approach managed these biases (underdispersion and zero count deficit) by selecting the CMP model and by the addition of relevant covariates.

However, despite differences in BIC, calculation time, underdispersion and zero count deficit, the long-term trend estimates for the skylark population produced by the two approaches differed by <2% in 27 years (-30.1% vs -31.8%, Table 3). This finding contrasts with other studies (with smaller datasets) where the choice of distribution to account for underdispersion was found to be important (Brooks et al., 2019). One explanation may be that our sampling protocol was well planned and sufficient (in terms of repetition, we used two sessions) and therefore enabled us to correctly detect trends over time despite less precise analyses (Sulkava et al., 2007). Another explanation could be that the large amount of data limited the impact on the estimated trend (Campbell, 2021), though previous studies have shown that even in cases of large data sets failure to use the appropriate distribution leads to distortions in trend estimates, at least in the case of overdispersion (Tirozzi et al., 2022). Methods dealing with overdispersed data are already available, since it is a common case in count data. The most popular method is TRIM (Pannekoek and Van Strien, 2005), but other



**Fig. 2.** Population trends of skylarks in the LTSER-ZAPVS between 1996 and 2022. The y-axis indicates the relative abundance of the species at a given site. The red line represents the trends obtained by the multi-model approach with a Conway Maxwell Poisson (CMP) distribution. The blue line illustrates the trends obtained by the ecological single model approach with a Poisson distribution. The raw mean number of skylarks counted per year is represented by a black circle with its associated standard error. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

methods are available, for example the use of hierarchical Bayesian modelling (Sauer et al., 2017). It is also possible to use distributions that take account of overdispersion (e.g., NB, see Le Rest et al., 2015) or to manage residual spatial autocorrelation in order to reduce overdispersion (Le Rest et al., 2013). Thus, our results may not be generalisable to all long-term trend modelling. Moreover, in many bird species, zero-inflation rather than a deficiency of zero counts is the rule, which can require the use of complicated modelling procedures (Blasco-Moreno et al., 2019; Martin et al., 2005). Such difficulties were avoided here because the skylark is an abundant bird species in our study area. Therefore, a similar study using rarer species, possibly leading to overdispersion, is warranted.

Another significant result of our study is that adding covariates did not significantly influence the trend estimate, a result which has already pointed out for the skylark (Bas et al., 2008; Tirozzi et al., 2022). One explanation could be that our data set was actually quite standardised in terms of environmental conditions, since counts were conducted at the optimal time of day (~95% of counts were performed between 7 a.m. and 11 a.m.), only during the peak breeding period (~85% between April and May), and always in favourable weather. In addition, rain and wind, which are factors that could clearly reduce detection probability for the forest counts, have much less influence in open habitat (Ralph et al., 1995). In other words, standardising protocols by restricting counts to optimal conditions simplifies statistical modelling (Brotons and Herrando, 2011; Oakley et al., 2006). Covariate selection may, therefore, be a key concern only when protocols are not standardised or restrictive enough (Sauerbrei et al., 2020). Finally, our study indicates that for underdispersed data obtained using a reasonably standardised protocol, use of the raw data and the Poisson distribution save much computing time with no cost in terms of trend estimate quality.

In our study area, which is mainly dedicated to intensive cereal production, a previous study on the skylark (Brodier et al., 2014) found a 20% decline between 1996 and 2012. Our results confirm this decline, updated to ~31% over 1996–2022. Such a trend is not new for this species and is found throughout Europe, including in France (–25% 2001–2018: Vigie-Nature, 2023), Germany (–55% 1983–2019: NABU,

2023), Switzerland (–52% 1990–2021: Vogelwarte, 2023), Italy (–66%, 2000–2020: Rete Rurale Nazionale and Lipu, 2020), and the UK (–17.5%, 1995–2018: Hughes et al., 2021). Skylarks are highly sensitive to decreased crop diversity, intensive agricultural management (e.g., excessive mowing or use of pesticides and fertilisers) (Chamberlain and Siriwardena, 2000; Chamberlain and Gregory, 1999; Koleček et al., 2015), and urbanisation (Loretto et al., 2019). All these reasons may explain these declines in skylark abundance, as human activity has largely transformed ecosystems over the last century (Ellis and Taylor, 2018; Goudie, 2013).

To conclude, this study aimed to explore how modelling approaches may affect the outcome of long-term monitoring data analysis. Our study suggests that a simple modelling approach using a classical Poisson distribution without adding covariates is appropriate for modelling long-term trends from data collected under well-designed protocols, at least when the data are underdispersed. To check the generality of our findings, further investigation needs to be performed on multiple species data, rarer species, and smaller data sets to check the generality of our findings.

## Funding

This work was supported by the region Nouvelle-Aquitaine, with partial funding through the BioBird project and by the ‘Conseil Départemental des Deux-Sèvres, France’ to TSB PhD grant. VB and SG are funded by CNRS and INRAE, respectively.

## Declaration of Competing Interest

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The data that has been used is confidential.

## Acknowledgements

We would like to thank all the 79 fieldworkers who have contributed to this extensive data collection since 1995, without whom this study would not have been possible. We also thank Elva Fuentes for her help in the Fig. 1 production. The manuscript was read and improved for English by Cambridge proof reading.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecoinf.2023.102222>.

## References

- Adde, Antoine, Rey, Pierre-Louis, Fopp, Fabian, Petitpierre, Blaise, Schweiger, Anna K., Broennimann, Olivier, Lehmann, Anthony, et al., 2023. Too many candidates: embedded covariate selection procedure for species distribution modelling with the Covsel R package. *Ecol. Inform.* 75 (July), 102080 <https://doi.org/10.1016/j.ecoinf.2023.102080>.
- Aho, Ken, Derryberry, DeWayne, Peterson, Teri, 2014. Model selection for ecologists: the worldviews of AIC and BIC. *Ecology* 95 (3), 631–636. <https://doi.org/10.1890/13-1452.1>.
- Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, B.F. (Eds.), *Second International Symposium on Information Theory*. Akademiai Kiado, Budapest, pp. 267–281.
- Bas, Yves, Devictor, Vincent, Moussus, Jean-Pierre, Jiguet, Frédéric, 2008. Accounting for weather and time-of-day parameters when analysing count data from monitoring programs. *Biodivers. Conserv.* 17 (14), 3403–3416. <https://doi.org/10.1007/s10531-008-9420-6>.
- Blake, John G., 1992. Temporal variation in point counts of birds in a lowland wet forest in Costa Rica. *Condor* 94 (1), 265–275. <https://doi.org/10.2307/1368816>.
- Blasco-Moreno, Anabel, Pérez-Casany, Marta, Puig, Pedro, Morante, Maria, Castells, Eva, 2019. What does a zero mean? Understanding false, random and structural zeros in ecology. Edited by Robert B. O'Hara. *Methods Ecol. Evol.* 10 (7), 949–959. <https://doi.org/10.1111/2041-210X.13185>.
- Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, H., White, J.S., 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol.* 24 (3), 127–135. <https://doi.org/10.1016/j.tree.2008.10.008>.
- Bonthoux, Balent, 2011. Point count duration: Five minutes are usually sufficient to model the distribution of bird species and to study the structure of communities for a French landscape. *J. Ornithol.* 153 <https://doi.org/10.1007/s10336-011-0766-2>.
- Bowler, Diana E., Heldbjerg, Henning, Fox, Anthony D., Jong, Maaik, Böhnning-Gaese, Katrin, 2019. Long-term declines of European insectivorous bird populations and potential causes. *Conserv. Biol.* 33 (5), 1120–1130. <https://doi.org/10.1111/cobi.13307>.
- Bretagnolle, Berthet, Gross, Gauffre, Plumejeaud, 2018. Towards sustainable and multifunctional agriculture in farmland landscapes: Lessons from the integrative approach of a French LTSEER platform. *Sci. Total Environ.* 622–634, 101016/[j.scitotenv.201801142](https://doi.org/10.1016/j.scitotenv.201801142).
- Briscoe, N.J., Zurell, D., Elith, J., König, C., Fandos, G., Malchow, A.K., Kéry, M., Schmid, H., Guillaera-Arroita, G., 2021 Sep. Can dynamic occupancy models improve predictions of species' range dynamics? A test using Swiss birds. *Glob. Chang. Biol.* 27 (18), 4269–4282. <https://doi.org/10.1111/gcb.15723>.
- Brodier, S., Augiron, S., Cornulier, T., Bretagnolle, V., 2014. Local improvement of skylark and corn bunting population trends on intensive arable landscape: a case study of the conservation tool Natura 2000: Natura 2000 effect on farmland bird trends. *Anim. Conserv.* 17 (3), 204–216. <https://doi.org/10.1111/acv.12077>.
- Brooks, Mollie E., Kristensen, Kasper, van Benthem, Koen J., Magnusson, Arni, Berg, Casper W., Nielsen, Anders, Skaug, Hans J., Mächler, Martin, Bolker, Benjamin M., 2017. GlmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R. J.* 9 (2), 378. <https://doi.org/10.32614/RJ-2017-066>.
- Brooks, Mollie E., Kristensen, Kasper, Darrigo, Maria Rosa, Rubim, Paulo, Uriarte, Maria, Bruna, Emilio, Bolker, Benjamin M., 2019. Statistical modeling of patterns in annual reproductive rates. *Ecology* 100 (7). <https://doi.org/10.1002/ecy.2706>.
- Brottons, Luís, Herrando, Sergi, 2011. Population estimates: towards standardised protocols as a basis for comparability. *Ardeola* 58 (2), 365–370. <https://doi.org/10.13157/arla.58.2.2011.365>.
- Buschke, F.T., Hagan, J.G., Santini, L., et al., 2021. Random population fluctuations bias the Living Planet Index. *Nat. Ecol. Evol.* 5, 1145–1152. <https://doi.org/10.1038/s41559-021-01494-0>.
- Campbell, Harlan, 2021. "The consequences of checking for zero-inflation and Overdispersion in the analysis of count data". Edited by Robert B. O'Hara. *Methods Ecol. Evol.* 12 (4), 665–680. <https://doi.org/10.1111/2041-210X.13559>.
- Chakrabarti, Arijit, Ghosh, Jayanta K., 2011. AIC, BIC and recent advances in model selection. In: *Philosophy of Statistics*. Elsevier, pp. 583–605. <https://doi.org/10.1016/B978-0-444-51862-0.50018-6>.
- Chamberlain, Sirwardena, 2000. The effects of agricultural intensification on Skylarks (*Alauda arvensis*): Evidence from monitoring studies in Great Britain. *Environ. Rev.* 8 (2), 95–113.
- Chamberlain, D.E., Gregory, R.D., 1999. Coarse and fine scale habitat associations of breeding skylarks *Alauda Arvensis* in the UK. *Bird Study* 46 (1), 34–47. <https://doi.org/10.1080/00063659909461113>.
- Chapin III, F., Stuart, Erika S., Zavaleta, Valerie T., Eviner, Rosamond L., Naylor, Peter M., Vitousek, Heather L., Reynolds, David, Hooper, U., et al., 2000. Consequences of changing biodiversity. *Nature* 405 (6783), 234–242. <https://doi.org/10.1038/35012241>.
- Cornulier, Thomas, Bretagnolle, Vincent, 2006. Assessing the influence of environmental heterogeneity on bird spacing patterns: a case study with two raptors. *Ecography* 29 (2), 240–250. <https://doi.org/10.1111/j.2006.0906-7590.04287.x>.
- Csikós, Nándor, Péter, Szilassi, 2020. Impact of Energy Landscapes on the Abundance of Eurasian Skylark (*Alauda arvensis*), an Example from North Germany. *Sustainability* 12 (2), 664. <https://doi.org/10.3390/su12020664>.
- Delius, J.D., 1965. A population study of Skylarks *Alauda arvensis*. *Ibis* 107, 466–492. <https://doi.org/10.1111/j.1474-919X.1965.tb07332.x>.
- Desquilbet, Gaume, Grippa, Céréghino, Humbert, Bonmatin, Cornillon, Maes, Van Dyck, Goulson, 2020. Comment on "Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science* 370 (6523). <https://doi.org/10.1126/science.abd8947>.
- Donald, 2004. *The Skylark*. T & A D Poyser. ISBN:9780713665680.
- Dormann, Carsten F., McPherson, Jana M., Araújo, Miguel B., Bivand, Roger, Bolliger, Janine, Carl, Gudrun, Davies, Richard G., et al., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30 (5), 609–628. <https://doi.org/10.1111/j.2007.0906-7590.05171.x>.
- Dziak, John J., Coffman, Donna L., Lanza, Stephanie T., Li, Runze, Jermiin, Lars S., 2020. Sensitivity and specificity of information criteria. *Brief. Bioinform.* 21 (2), 553–565. <https://doi.org/10.1093/bib/bbz016>.
- Ellis, Murray V., Taylor, Jennifer E., 2018. Effects of weather, time of day, and survey effort on estimates of species richness in temperate woodlands. *Emu - Austral Ornithol.* 118 (2), 183–192. <https://doi.org/10.1080/01584197.2017.1396188>.
- Eglington, S.M., Davis, S.E., Joys, A., Chamberlain, D., Noble, D., 2010. The effect of observer experience on English Breeding Bird Survey population trends. *Bird Study* 57 (2), 129–141. <https://doi.org/10.1080/00063650903440648>.
- Farley, S.S., Dawson, A., Goring, S.J., Williams, J.W., 2018. Ecology as a Big-Data Science. *Current Advances, Challenges, and Solutions*. *Bioscience* 68, 563–576. <https://doi.org/10.1093/BIOSCI/BIY068>.
- Fokianos, K., 2012. Count time series models. *Time Series-Methods and Applications*, 30, 315–347. <https://doi.org/10.1016/B978-0-444-53858-1.00012-0>.
- Forister, Matthew L., Pelton, Emma M., Black, Scott H., 2019. Declines in insect abundance and diversity: we know enough to act now. *Conserv. Sci. Pract.* 1 (8). <https://doi.org/10.1111/csp.280>.
- Forthmann, Boris, Doeblner, Philipp, 2021. Reliability of researcher capacity estimates and count data dispersion: a comparison of Poisson, negative binomial, and Conway-Maxwell-Poisson models. *Scientometrics* 126 (4), 3337–3354. <https://doi.org/10.1007/s11192-021-03864-8>.
- Fraser, H., Parker, T., Nakagawa, S., Barnett, A., Fidler, F., 2018. Questionable research practices in ecology and evolution. *PLOS ONE* 13 (7), e0200303. <https://doi.org/10.1371/journal.pone.0200303>.
- Freckleton, R.P., 2009. The seven deadly sins of comparative analysis. *J. Evol. Biol.* 22 (7), 1367–1375. <https://doi.org/10.1111/j.1420-9101.2009.01757.x>.
- Fuller, R.J., Langslow, D.R., 1984. Estimating numbers of birds by point counts: how long should counts last? *Bird Study* 31 (3), 195–202. <https://doi.org/10.1080/00063658409476841>.
- Gabriel, Hernando, Miguel, Roa, Isabel, Fernández-Gil Juan, Juan, Jara, Fuertes, Benito, Reguera, Bárbara, Revilla, Eloy, 2022. Trends in Weather Conditions Favor Generalist over Specialist Species in Rear-Edge Alpine Bird Communities. *Ecosphere* 13 (4), e3953. <https://doi.org/10.1002/ecs2.3953>.
- Gilad-Bachrach, Ran, Navot, Amir, Tishby, Naftali, 2003. An information theoretic tradeoff between complexity and accuracy. In: Schölkopf, Bernhard, Warmuth, Manfred K. (Eds.), *Learning Theory and Kernel Machines*, 2777. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 595–609. [https://doi.org/10.1007/978-3-540-45167-9\\_43](https://doi.org/10.1007/978-3-540-45167-9_43).
- Goudie, 2013. *The Human Impact on the Natural Environment: Past, Present, and Future*. Wiley. ISBN:9781118688151.
- Grafen, A., Hails, R., 2002. Modern statistics for the life sciences. *Ann Bot* 90 (6), 776–777. <https://doi.org/10.1093/aob/mcf254>.
- Guo, Jie Q., Li, Tong, 2002. Poisson regression models with errors-in-variables: implication and treatment. *J. Stat. Plan. Infer.* 104 (2), 391–401. [https://doi.org/10.1016/S0378-3758\(01\)00250-6](https://doi.org/10.1016/S0378-3758(01)00250-6).
- Harrison, 2014. Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ* 9 (2), e616. <https://doi.org/10.7717/peerj.616>.
- Harris, Tammy, Yang, Zhao, Hardin, James W., 2012. Modeling underdispersed count data with generalized Poisson regression. *Stata J. Promot. Commun. Stat. Stata* 12 (4), 736–747. <https://doi.org/10.1177/1536867X1201200412>.
- Harrison, Philip J., Buckland, Stephen T., Yuan, Yuan, Elston, David A., Brewer, Mark J., Johnston, Alison, Pearce-Higgins, James W., 2014. "Assessing trends in biodiversity over space and time using the example of British breeding birds". Edited by Yann Clough. *J. Appl. Ecol.* 51 (6), 1650–1660. <https://doi.org/10.1111/1365-2664.12316>.
- Hartig, Florian, 2022. DHARMA: Residual Diagnostics for Hierarchical (Multi-Level/Mixed) Regression Models, 8 September 2022. <https://cran.r-project.org/web/packages/DHARMA/vignettes/DHARMA.html>.
- He, Fengzhi, Zarfl, Christiane, Bremerich, Vanessa, David, Jonathan N.W., Hogan, Zeb, Kalinkat, Gregor, Tockner, Klement, Jähnig, Sonja C., 2019. The global decline of freshwater megafauna. *Glob. Chang. Biol.* 25 (11), 3883–3892. <https://doi.org/10.1111/gcb.14753>.

- Heinze, Georg, Wallisch, Christine, Dunkler, Daniela, 2018. Variable selection - a review and recommendations for the practicing statistician. *Biom. J.* 60 (3), 431–449. <https://doi.org/10.1002/bimj.201700067>.
- Hilbe, Joseph M., 2011. *Negative Binomial Regression*, 2nd ed. Cambridge University Press. <https://doi.org/10.1017/CBO9780511973420>.
- Hoef, Ver, Jay, M., Boveng, Peter L., 2007. Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology* 88 (11), 2766–2772. <https://doi.org/10.1890/07-0043.1>.
- Hoffmann, Jörg, Wittchen, Udo, Stachow, Ulrich, Berger, Gert, 2016. Moving window abundance – a method to characterise the abundance dynamics of farmland birds: the example of skylark (*Alauda Arvensis*). *Ecol. Indic.* 60 (January), 317–328. <https://doi.org/10.1016/j.ecolind.2015.06.037>.
- James, A. Hanley, Sahir, Bhatnagar, 2022. The “Poisson” Distribution: History, Reenactments, Adaptations. *The American Statistician* 76 (4), 363–371. <https://doi.org/10.1080/00031305.2022.2046159>.
- Johnson, Douglas H., 1999. The insignificance of statistical significance testing. *J. Wildl. Manag.* 63 (3), 763. <https://doi.org/10.2307/3802789>.
- Johnson, Timothy R., 2012. J.M. Hilbe (2011) negative binomial regression, second edition. *Psychometrika* 77 (3), 611–612. <https://doi.org/10.1007/s11336-012-9263-7>.
- Kamp, J., Ooppel, S., Heldbjerg, H., Nyegaard, T., Donald, P.F., 2016. Unstructured citizen science data fail to detect long-term population declines of common birds in Denmark. *Diversity Distrib.* 22, 1024–1035. <https://doi.org/10.1111/ddi.12463>.
- Kendall, W.L., Peterjohn, B.G., Sauer, J.G., 1996. First-Time Observer Effects in the North American Breeding Bird Survey. *The Auk* 113 (4), 823–829. <https://doi.org/10.2307/4088860>.
- Koleček, Jaroslav, Reif, Jiří, Weidinger, Karel, 2015. The abundance of a farmland specialist bird, the skylark, in three European regions with contrasting agricultural management. *Agric. Ecosyst. Environ.* 212 (December), 30–37. <https://doi.org/10.1016/j.agee.2015.06.018>.
- Kouba, M., Bartoš, L., Bartošová, J., et al., 2021. Long-term trends in the body condition of parents and offspring of Tengmalm’s owls under fluctuating food conditions and climate change. *Sci Rep* 11, 18893. <https://doi.org/10.1038/s41598-021-98447-1>.
- Leu, M., Farnsworth, M.L., Fleishman, E., Dobkin, D.S., Scherer, R.D., Noon, B.R., Dickson, B.G., 2017. Effects of point-count duration on estimated detection probabilities and occupancy of breeding birds. *J. Field Ornithol.* 88. <https://doi.org/10.1111/jof.12183>.
- Leung, Brian, Hargreaves, Anna L., Greenberg, Dan A., McGill, Brian, Dornelas, Maria, Freeman, Robin, 2020. Clustered versus catastrophic global vertebrate declines. *Nature* 588 (7837), 267–271. <https://doi.org/10.1038/s41586-020-2920-6>.
- Lindén, Andreas, Mäntyniemi, Samu, 2011. Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology* 92 (7), 1414–1421. <https://doi.org/10.1890/10-1831.1>.
- Łopucki, R., Kiersztyń, A., Pitucha, G., Kitowski, I., 2022. Handling missing data in ecological studies: Ignoring gaps in the dataset can distort the inference. *Ecol. Model.*, 468, Article 109964. <https://doi.org/10.1016/j.ecolmodel.2022.109964>.
- Loretto, M.C., Schöll, E.M., Hille, S., 2019. Occurrence of Eurasian Skylark *Alauda arvensis* territories in relation to urban area and heterogeneous farmland. *Bird Study* 66 (2), 273–278. <https://doi.org/10.1080/00063657.2019.1637816>.
- Lynch, Heather J., Thorson, James T., Shelton, Andrew Olaf, 2014. Dealing with under- and over-dispersed count data in life history, spatial, and community ecology. *Ecology* 95 (11), 3173–3180. <https://doi.org/10.1890/13-1912.1>.
- Ma, Zhihai, Zuckerberg, Benjamin, Porter, William F., Zhang, Lianjun, 2012. Spatial Poisson models for examining the influence of climate and land cover pattern on bird species richness. *For. Sci.* 58 (1), 61–74. <https://doi.org/10.5849/forsci.10-111>.
- Martin, Tara G., Wintle, Brendan A., Rhodes, Jonathan R., Kuhnert, Petra M., Field, Scott A., Low-Choy, Samantha J., Tyre, Andrew J., Possingham, Hugh P., 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations: modelling excess Zeros in ecology. *Ecol. Lett.* 8 (11), 1235–1246. <https://doi.org/10.1111/j.1461-0248.2005.00826.x>.
- Newson, Stuart E., Massimino, Dario, Johnston, Alison, Baillie, Stephen R., Pearce-Higgins, James W., 2013. Should we account for detectability in population trends? *Bird Study* 60 (3), 384–390. <https://doi.org/10.1080/00063657.2013.805729>.
- Oakley, Karen L., Thomas, Lisa P., Fancy, Steven G., 2006. Guidelines for long-term monitoring protocols. *Wildlife Soc. Bull.* (1973–2006) 31 (4), 1000–1003.
- Öberg, M., Arlt, D., Pärt, T., Laugen, A.T., Eggers, S., Low, M., 2015. Jan. Rainfall during parental care reduces reproductive and survival components of fitness in a passerine bird. *Ecol. Evol.* 5 (2), 345–356. <https://doi.org/10.1002/ece3.1345>.
- Pannekoek, Van Strien, 2005. TRIM 3 manual (TRENDS and INDICES for Monitoring data). Statistics, Netherlands.
- Polanco-Martínez, Josué M., López-Martínez, José L., 2021. A non-parametric method to test the statistical significance in rolling window correlations, and applications to ecological time series. *Ecol. Inform.* 64 (September), 101379. <https://doi.org/10.1016/j.ecoinf.2021.101379>.
- Ralph, C. John, Droege, Sam, Sauer, John R., 1995. *Managing and Monitoring Birds Using Point Counts: Standards and Applications*.
- Rest, Le, Kévin, David Pinaud, Bretagnolle, Vincent, 2013. Accounting for spatial autocorrelation from model selection to statistical inference: application to a National Survey of a diurnal raptor. *Ecol. Inform.* 14 (March), 17–24. <https://doi.org/10.1016/j.ecoinf.2012.11.008>.
- Rest, Le, Kévin, David Pinaud, Bretagnolle, Vincent, 2015. Volunteer-based surveys offer enhanced opportunities for biodiversity monitoring across broad spatial extent. *Ecol. Inform.* 30 (November), 313–317. <https://doi.org/10.1016/j.ecoinf.2015.08.007>.
- Rete Rurale Nazionale & Lipu, 2020. Common Breeding Farmland Birds in Italy. Update of Population Trends and Farmland Bird Indicator for National Rural Network 2000–2020. <https://www.reterurale.it/flex/cm/pages/ServeBLOB.php/L/IT/IDPagina/22311>.
- Richards, S.A., 2008. Dealing with overdispersed count data in applied ecology. *J. Appl. Ecol.* 45, 218–227. <https://doi.org/10.1111/j.1365-2664.2007.01377.x>.
- Rosenberg, Kenneth V., Dokter, Adriaan M., Blancher, Peter J., Sauer, John R., Smith, Adam C., Smith, Paul A., Stanton, Jessica C., et al., 2019. Decline of the north American avifauna. *Science* 366 (6461), 120–124. <https://doi.org/10.1126/science.aaw1313>.
- Ruan, Guangchen, Hanson, Paul C., Dugan, Hilary A., Plale, Beth, 2017. Mining Lake time series using symbolic representation. *Ecol. Inform.* 39 (mai), 10–22. <https://doi.org/10.1016/j.ecoinf.2017.03.001>.
- Sauer, John R., Link, William A., Fallon, Jane E., Pardieck, Keith L., Ziolkowski, David J., 2013. The north American breeding bird survey 1966–2011: summary analysis and species accounts. *N. Am. Fauna* 79 (December), 1–32. <https://doi.org/10.3996/nafa.79.0001>.
- Sauer, John R., Pardieck, Keith L., Ziolkowski, David J., Smith, Adam C., Hudson, Marie-Anne R., Rodriguez, Vicente, Berlanga, Humberto, Niven, Daniel K., Link, William A., 2017. The first 50 years of the north American breeding bird survey. *Condor* 119 (3), 576–593. <https://doi.org/10.1650/CONDOR-17-83.1>.
- Sauerbrei, Willi, Perperoglou, Aris, Schmid, Matthias, Abrahamowicz, Michal, Becher, Heiko, Binder, Harald, Dunkler, Daniela, Harrell, Frank E., Royston, Patrick, Heinze, Georg, 2020. State of the art in selection of variables and functional forms in multivariable analysis—outstanding issues. *Diagnost. Prognost. Res.* 4 (1), 3 s41512-020-00074-3. <https://doi.org/10.1186/s41512-020-00074-3>.
- Schwarz, 1978. Estimating the Dimension of a Model. *Ann. Statist.* 6(2) 461 - 464. <https://doi.org/10.1214/aos/1176344136>.
- Schielzeth, Holger, 2010. Simple means to improve the interpretability of regression coefficients: *interpretation of regression coefficients*. *Methods Ecol. Evol.* 1 (2), 103–113. <https://doi.org/10.1111/j.2041-210X.2010.00012.x>.
- Sellers, Kimberly F., Morris, Darcy S., 2017. Underdispersion models: models that are ‘Under the Radar’. *Commun. Stat. Theory Methods* 46 (24), 12075–12086. <https://doi.org/10.1080/03610926.2017.1291976>.
- Sellers, Kimberly F., Shmueli, Galit, 2010. A flexible regression model for count data. *Ann. Appl. Stat.* 4 (2) <https://doi.org/10.1214/09-AOAS306>.
- Shmueli, Galit, 2010. To explain or to predict? *Stat. Sci.* 25, (3). <https://doi.org/10.1214/10-STS330>.
- Sokos, Christos K., Birtsas, Periklis K., Platis, Petros C., Pappaspyropoulos, Konstantinos G., 2016. Weather influence on the abundance of bird species wintering in three Mediterranean ecosystems. *Folia Zool.* 65 (3), 200–207. <https://doi.org/10.25225/fozo.v65.i3.a4.2016>.
- Stoklosa, J., Blakey, R.V., Hui, F.K.C., 2022. An Overview of Modern Applications of Negative Binomial Modelling in Ecology and Biodiversity, 14. *Diversity*, p. 320. <https://doi.org/10.3390/d14050320>.
- Sulkava, Mika, Luysaert, Sebastiaan, Rautio, Pasi, Janssens, Ivan A., Hollmén, Jaakko, 2007. Modeling the effects of varying data quality on trend detection in environmental monitoring. *Ecol. Inform.* 2 (2), 167–176. <https://doi.org/10.1016/j.ecoinf.2007.03.008>.
- Svårdson, Gunnar, 1949. Competition and habitat selection in birds. *Oikos* 1 (2), 157. <https://doi.org/10.2307/3564710>.
- Tirozzi, Pietro, Orioli, Valerio, Dondina, Olivia, Kataoka, Leila, Bani, Luciano, 2021. Species traits drive long-term population trends of common breeding birds in northern Italy. *Animals* 11 (12), 3426. <https://doi.org/10.3390/ani11123426>.
- Tirozzi, Pietro, Orioli, Valerio, Dondina, Olivia, Kataoka, Leila, Bani, Luciano, 2022. Population trends from count data: handling environmental bias, Overdispersion and excess of zeroes. *Ecol. Inform.* 69 (July), 101629. <https://doi.org/10.1016/j.ecoinf.2022.101629>.
- Tredennick, Andrew T., Hooker, Giles, Ellner, Stephen P., Adler, Peter B., 2021. A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology* 102 (6). <https://doi.org/10.1002/ecy.3336>.
- Vergara, Pablo, M., et al., 2007. Arrival date, age and breeding success in white stork *Ciconia ciconia*. *Journal of Avian Biology* 38, 573–579. <https://doi.org/10.1111/j.1365-0708.8857.03983.x>.
- Verner, Jared, Ritter, Lyman V., 1986. Hourly variation in morning point counts of birds. *Auk* 103 (1), 117–124. <https://doi.org/10.1093/auk/103.1.117>.
- Vitousek, Peter M., Mooney, Harold A., Lubchenco, Jane, Melillo, Jerry M., 1997. *Human Domination of Earth’s Ecosystems*, 277, p. 7.
- Wagner, David L., 2020. Insect declines in the Anthropocene. *Annu. Rev. Entomol.* 65 (1), 457–480. <https://doi.org/10.1146/annurev-ento-0111019-025151>.
- Wakeham-Dawson, Andrew, 1995. Hares and Skylarks as Indicators of Environmentally Sensitive Farming on the South Downs. PhD thesis The Open University. <https://doi.org/10.21954/ou.ro.00000fb7b>.
- White, J. Wilson, Rassweiler, Andrew, Samhuri, Jameal F., Stier, Adrian C., White, Crow, 2014. Ecologists should not use statistical significance tests to interpret simulation model results. *Oikos* 123 (4), 385–388. <https://doi.org/10.1111/j.1600-0706.2013.01073.x>.
- WWF, 2022. *Living Planet Report 2022 - Building a nature-positive society*. WWF, Gland, Switzerland.
- Xekalaki, Evdokia, 2006. Under- and Overdispersion. In: Teugels, Jozef L., Sundt, Björn (Eds.), *Encyclopedia of Actuarial Science*. John Wiley & Sons, Ltd., Chichester, UK. <https://doi.org/10.1002/9780470012505.tau003>.
- Zuur, Alain F., Elena, N. Ieno, Graham, M. Smith, 2007. *Analysing Ecological Data*. Springer New York, NY. <https://doi.org/10.1007/978-0-387-45972-1>.
- Zuur, Alain, Ieno, Elena, Walker, Neil, Saveliev, Anatoly, Smith, Graham, 2009. *GLMM and GAMM. Mixed effects models and extensions in ecology with R*. [https://doi.org/10.1007/978-0-387-87458-6\\_13](https://doi.org/10.1007/978-0-387-87458-6_13).