Should We *Transform* Our Understanding of Linearity in Generalized Linear Models?

David Heit[1], Mairi Poisson[1], Andrew Butler[1], Waldemar Ortiz-Calo[2], Robert Montgomery[3], Remington Moll[1]

1 Wildlife Modeling and Management Laboratory, University of New Hampshire, Durham, NH, USA **2** Wildlife Biology Program, W.A. Franke College of Forestry, University of Montana, Missoula, MT, USA
3 Department of Biology, University of Oxford, Oxford, UK

The problem with linearity

- Linear models, and by extension **generalized linear models (GLMs)** are an incredibly widespread and useful tool in wildlife ecology.
- Linear models inherently assume that the relationship between a predictor and a response is linear, which may not perfectly capture the **complex nature of ecological relationships**.
- Generalized linear models (GLMs) expand the applicability of linear models beyond just positive continuous data using link funcitons.
- However link functions do not remove the linearity assumption, the assumption shifts to the link scale.
- While the effects of non-linearity are more intuitive for simple linear regression, they become less straigt-forward in GLMs.

Link functions

IDENTITY

$y_i = \beta_0 + \beta_1 * x_i$

$y_i = \beta_0 + \beta_1 * \log(x_i)$

LOGIT

$\log(\text{odds ratio}_{y_i}) = \beta_0 + \beta_1 * x_i$

$\log(\text{odds ratio}_{y_i}) = \beta_0 + \beta_1 * x_i^2$

LOG

$\log(y_i) = \beta_0 + \beta_1 * x_i$

$\log(y_i) = \beta_0 + \beta_1 * x_i + \beta_2 * x_i^2$

linear on link scale
non-linear on link scale

CASE STUDY

Songbird Occupancy Modeling

POINT COUNT SURVEYS

- ***Study area:*** 63 sites across UNH owned woodlands and public areas in Durham and Dover, NH.
- ***Field Season:*** Summer breeding season 2021 & 2022
- ***Protocol:*** All surveys completed between 30 mins. before sunrise and 10:00am. Surveys last 8 minutes. Each site is surveyed 3 times per season.

	No. Obs.	Effect (linear)	Effect (logged)	DIC (linear)	DIC (logged)
Gray Catbird	52	0.178	-0.247	312.7	309.2
Nothern cardinal	17	1.131	0.6259	141.2	165.65
Wood Thrush	11	-0.48	-0.928	130.2	110.3
Black & White Warbler	15	-1.24	-1.5	134.99	108.7
House Sparrow	14	1.51	1.45	150.58	120.21

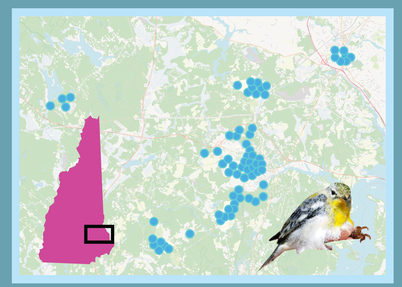


Figure 1. Survey sites for bird point count sampling in southeastern New Hampshire, USA.

MODELING METHODS

- Retained birds with >20 observations (42 spp.)
- Used bayesian occupancy-detection models to conduct Box-Tidwell linearity tests (Box & Tidwell 1962)

BOX - TIDWELL TEST (Box & Tidwell 1962)

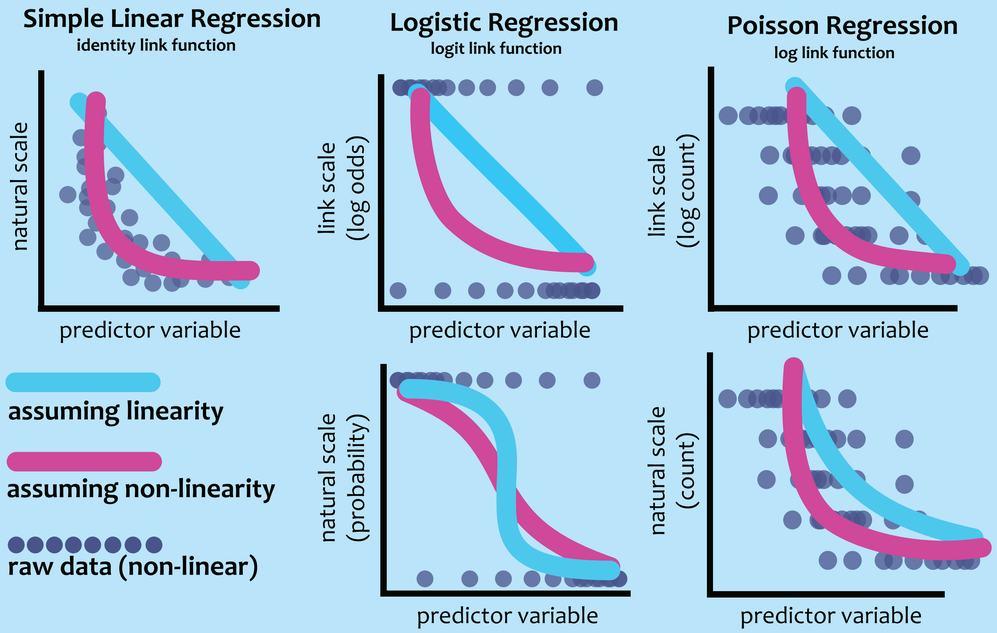
$\log(\text{odds ratio}_{y_i}) = \beta_0 + \beta_1 * x_i + \beta_2 * x_i * \log(x_i)$

RESULTS

- **12% (5/42) of species failed Box-Tidwell test**
- **Of those 5 species, 4 had improved DIC with log-transformed covariate.**
- **One species (Northern Cardinal) failed Box-Tidwell test, but the model was not improved by logging, indicating some other non-linear relationship**

DATA SIMULATION

Non-linearity in GLMs



FROM THE LITERATURE

How often are ecologists thinking about linearity?

LITERATURE REVIEW

273 ecology papers published in the last 5 years
Results shown here are preliminary (50 papers)

RESULTS

Tested for linearity:	**14%**
Transformed covariates:	**20%**
Fit non-linear model:	**22%**

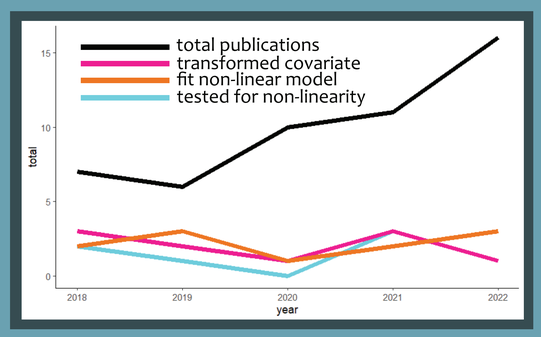


Figure 2. Results from preliminary literature review showing the relative frequency of incorporating efforts to address linearity.

Putting it all together

- Based on preliminary results of literature review, **testing for linearity is not common practice.**
- It is more common to fit a non-linear model than to test for linearity.
- Relationships between covariates and link-scale responses can be non-linear.
- Non-linearity can also **vary by species** within a community.
- Non-linear and linear relationships can vary in direction and magnitude within the same species.
- Log-linear transformation may not be sufficient to address non-linearity.
- Future best practices may include testing for linearity in communities, species, populations, and individuals.

Future Directions

- Expand literature review
- Add two more case studies to explore the dynamics of linearity in Poisson regression and generalized additive models (GAMs).
- How do species with non-linear effects affect community-level models?

FUNDING

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation

REFERENCES

Box G., Tidwell P. 1962. Transformation of the independent variables. *Technometrics*.