

## STATISTICAL REPORT

# Diet analysis using generalized linear models derived from foraging processes using R package *mvtweedie*

James T. Thorson<sup>1</sup>  | Mayumi L. Arimitsu<sup>2</sup>  | Taal Levi<sup>3</sup>  |  
 Gretchen H. Roffler<sup>4</sup> 

<sup>1</sup>AFSC, NMFS, NOAA, Seattle, Washington, USA

<sup>2</sup>U.S. Geological Survey Alaska Science Center, Juneau, Alaska, USA

<sup>3</sup>Department of Forest Ecosystems and Society, Oregon State University, Corvallis, Oregon, USA

<sup>4</sup>Alaska Department of Fish and Game, Douglas, Alaska, USA

## Correspondence

James T. Thorson

Email: [james.thorson@noaa.gov](mailto:james.thorson@noaa.gov)

**Handling Editor:** Andrew O. Finley

## Abstract

Diet analysis integrates a wide variety of visual, chemical, and biological identification of prey. Samples are often treated as compositional data, where each prey is analyzed as a continuous percentage of the total. However, analyzing compositional data results in analytical challenges, for example, highly parameterized models or prior transformation of data. Here, we present a novel approximation involving a Tweedie generalized linear model (GLM). We first review how this approximation emerges from considering predator foraging as a thinned and marked point process (with marks representing prey species and individual prey size). This derivation can motivate future theoretical and applied developments. We then provide a practical tutorial for the Tweedie GLM using new package *mvtweedie* that extends capabilities of widely used packages in R (*mgcv* and *ggplot2*) by transforming output to calculate prey compositions. We demonstrate this approach and software using two examples. Tufted Puffins (*Fratercula cirrhata*) provisioning their chicks on a colony in the northern Gulf of Alaska show decadal prey switching among sand lance and prowlfish (1980–2000) and then Pacific herring and capelin (2000–2020), while wolves (*Canis lupus ligoni*) in southeast Alaska forage on mountain goats and marmots in northern uplands and marine mammals in seaward island coastlines.

## KEYWORDS

diet, foraging theory, generalized additive model, generalized linear model, metabarcoding, point process, Poisson process, predation, Tweedie distribution

## INTRODUCTION

Ecologists often collect data regarding the proportion of a total belonging to different categories (called “compositional data”). The multivariate structure of compositional data is informative regarding trophic relationships of consumers and producers, competition dynamics, and the

flow of energy in food webs (Bonin et al., 2020; Shi et al., 2021). Many methods are used to measure the diet proportions of consumers including morphological identification of prey in stomach contents or feces, ratios of stable isotopes or fatty acids, and molecular approaches including DNA metabarcoding (Roffler et al., 2021; Shi et al., 2021). Potential applications of compositional diet

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Ecology* published by Wiley Periodicals LLC on behalf of The Ecological Society of America. This article has been contributed to by US Government employees and their work is in the public domain in the USA.

data range from defining the foraging niche of individuals, populations, or species to ecosystem-based management informed by trophic models.

Compositional data are often analyzed using models that are difficult to parameterize and interpret. For example, a Dirichlet distribution (the multivariate generalization of the beta distribution for proportions) can be fitted to proportions (ranging continuously from 0 to 1) derived from biomass samples within stomach contents (Ainsworth et al., 2010; Maier, 2021). However, this approach has several limitations, including:

1. The Dirichlet distribution does not support a value of 0 or 1, so the distribution cannot account for instances when the prey is either absent or the only species within a given sample. To rectify this, prey responses must be changed from 0/1 to for example, 0.0001/0.9999 (or some other arbitrary bounds) prior to analysis.
2. It is not clear how to incorporate covariates in an ecologically interpretable manner within the Dirichlet distribution. Analysts have incorporated covariates in a log-linked linear predictor for the concentration parameter for each prey (Maier, 2021). However, there is no explicit link between this concentration parameter and variables that are interpreted in foraging theory, for example, prey densities, predator energetic condition, or functional-response parameters.

In response to the first limitation, Moriarty et al. (2016) developed an alternative method that separately models the probabilities that 0% or 100% of a given stomach belong to each species (also see Douma & Weedon, 2019). This mixture model eliminates the need to pre-specify a lower/upper bound on encounters for any individual prey species. However, this solution in turn introduces many additional parameters, representing the probability that a given prey is 0% or 100% of stomach contents, in addition to parameters governing the expected density for intermediate values.

For these reasons, we instead seek to derive a model for compositional analysis that uses raw sampling data (without transformation prior to analysis), while also being parsimonious and derived directly from variables used in foraging theory. In particular, we build upon efforts to recast ecological models as thinned and marked point processes (Illian et al., 2008). We also seek to provide a simple code example for two contrasting cases using widely used software that is familiar to ecologists. We therefore develop a new R package *mvtweedie*, which provides a new *predict* S3 class that transforms output

from widely used R packages (*mgcv* for GAMs or *glmmTMB* for GLMMs) to predict prey compositions for any new set of covariates. These covariates can represent, for example, predator energetic condition, spatiotemporal variation in habitat predicting prey densities, as well as correlated residual variation. We then show how to explore this output using high-level graphical tools (*ggplot2*).

## METHODS

### Thinned and marked Poisson processes

We first briefly introduce the theory of a thinned marked point process (Illian et al., 2008). These processes provide a useful mathematical representation for animal foraging, but have received surprisingly little attention from ecologists for this or other uses (Illian & Burslem, 2017). Modifying notation from Illian et al. (2008), a point process has an intensity function  $\Lambda(s)$  for every location  $s$  with units of numbers per area, and the number of individuals within a subset  $A$  of the spatial domain is a random variable  $N(A)$  with expectation drawn from the intensity function,  $\mathbb{E}(N(A)) = \lambda \equiv \int_A \Lambda(s) ds$ .

Our presentation combines the following four elaborations (see Appendix S1 for details) of this basic point process to make it relevant for estimating predator feeding habits:

1. Prey encounters are distributed randomly at local scales (a Poisson process);
2. Prey densities are governed by spatial processes and environmental covariates that can be expressed as a log-linked linear model (a log-Gaussian-Cox process);
3. Prey animals have “marks” representing species and size (a marked point process);
4. Predators attack and consume a subset of local prey encounters (a thinned point process).

This thinned and marked Poisson process includes several well-known cases as submodels. In particular, we note the Multinomial-Poisson transformation as a widely used submodel (Appendix S2) before introducing the Tweedie generalized linear model (GLM) that is our primary focus.

### Tweedie distribution for thinned and marked Poisson process

The thinned and double-marked Poisson process results in a convenient approximation for biomass responses

when individual prey size is assumed to follow a Gamma distribution. Given that we have multiple prey categories, we have a Poisson process for counts  $N_{ic}$  of each prey type  $c$  in each sample  $i$  of predator food habits,  $N_{ic} \sim \text{Poisson}(\lambda_{ic})$ , and a gamma distribution for the mass  $B_{icn}$  of the  $n$ th of these  $N_{ic}$  animals,  $B_{icn} \sim \text{Gamma}(k_c, \theta_c)$ , using the shape-scale parameterization where  $k_c = v_c^{-2}$  and  $\theta_c = w_c v_c^2$  given that  $w_c$  is the average biomass and  $v_c$  is the coefficient of variation in individual biomass for each prey  $c$ .

Total biomass  $Y_{ic}$  for each prey  $c$  in sample  $i$  is then the sum of individual biomass  $B_{icn}$  for each of these  $N_{ic}$  animals of prey  $c$  in that sample. This is called a compound Poisson-gamma distribution for  $Y_{ic} = \sum_{n=1}^{N_{ic}} B_{icn}$ . Usefully, this compound Poisson-gamma process can be exactly expressed as a Tweedie distribution (Foster & Bravington, 2013),

$$Y_{ic} \sim \text{Tweedie}(\mu_{ic}, \phi_{ic}, p_c) \quad (1)$$

where the Tweedie mean parameter  $\mu_{ic} = \lambda_{ic} w_c = \mathbb{E}(Y_{ic})$  represents the expected thinned biomass for each prey, that is, the product of thinned local abundance  $\lambda_{ic}$  and average weight  $w_c$  for prey  $c$  (see Appendix S3: Figure S1). We call  $\phi_{ic}$  the scale parameter and  $p_c$  the power parameter, noting that  $\text{Var}(Y_{ic}) = \phi_{ic} \mu_{ic}^{p_c}$  (Foster & Bravington, 2013). These three Tweedie parameters can be calculated from thinned local prey abundance and size  $\mu_{ic} = \lambda_{ic} k_c \theta_c$ ,  $\phi_{ic} = \theta_c (k_c + 1) \mu_{ic}^{\frac{1}{k_c+1}}$ , and  $p_c = \frac{k_c+2}{k_c+1}$ , and all three vary among samples and/or prey.

### Estimation using a Tweedie GLM

We next show how to fit this Tweedie distribution (representing a thinned and marked Poisson process) as a GLM. This involves specifying that power and scale parameters are constant across all categories and samples ( $\phi_{ic} = \phi$  and  $p_c = p$ ). The resulting model can then be conveniently fitted to data as

$$\text{vec}(Y_{ic}) \sim \text{Tweedie}(\text{vec}(\mu_{ic}), \phi, p) \quad (2)$$

where operator  $\text{vec}(\cdot)$  vectorizes a matrix such that for example,  $\text{vec}(Y_{ic})$  is the vector formed by stacking columns of matrix  $\mathbf{Y}$ , and the same ordering is used for  $\text{vec}(\mu_{ic})$ . This can be fitted using standard software for a Tweedie GLM with prey as a category, for example, using R packages *mgcv* (Wood, 2006) or *glmmTMB* (Brooks et al., 2017).

Estimation using a Tweedie GLM allows us to rapidly explore covariates predicting thinned biomass density for each prey:

$$\log(\mu_{ic}) = \alpha_i + \sum_{k=1}^{n_k} \beta_{kc} x_{ik} \quad (3)$$

where  $\alpha_i$  represents the product of area-swept and thinning rate that underlies sample  $i$  (see Appendices S1 and S2 for more details). In some designs, it is reasonable to assume that all samples arise from the same area swept and have equivalent thinning rates, such that  $\alpha_i = \alpha$ . In other cases,  $\alpha_i$  could itself be modeled as a function of predator body size or other covariates, to again control for unknown differences in area swept and thinning rate associated with each sampled predator.

The Tweedie GLM then allows proportions to be back-calculated as

$$\pi_{ic} = \frac{\exp(\sum_{k=1}^{n_k} \beta_{kc} x_{ik})}{\sum_{c^*=1}^{n_c} \exp(\sum_{k=1}^{n_k} \beta_{kc^*} x_{ik})} \quad (4)$$

This shows that proportions  $\pi_i$  in the thinned and marked Poisson process can be fitted to raw samples (e.g., prey biomass in stomach contents) using a standard GLM framework, as an alternative to converting samples to proportions prior to analysis. Covariance in  $\mathbf{Y}_i$  arising from satiation or multi-species functional responses could be approximated via random effects for  $\pi_i$  relative to prey densities at each location, and we leave this as a topic for future exploration. We provide an alternative interpretation for the joint distribution of samples  $\mathbf{Y}_i$  for all prey  $c$  as arising from a ‘‘multivariate Tweedie’’ distribution in Appendix S3.

### Simulation example

We visualize this thinned and double-marked Poisson process using a simplified simulation experiment involving three prey and three foraging habitats, where each foraging habitat has distinct patch size  $A$ . We use this experiment in part to visualize the magnitude of error that may arise when approximating a thinned and double-marked Poisson process with a Tweedie distribution.

To do so, we first simulate locations for 1000 individuals of each prey from density function  $\Lambda_c(s)$ , where  $\Lambda_c(s)$  is specified as a bivariate normal distribution centered at different locations for each prey, and each prey has a gamma distribution for individual mass parameterized with mean  $w_c = 1.2$  kg and coefficient of variation  $v_c = 1$ . We visualize the spatial distribution and associated body size for each prey in a single realization of this experiment, and record the biomass for each prey within each of the three foraging areas.

We then repeat this simulation 1000 times to generate the expected distribution for prey biomass within each foraging area. We calculate the Tweedie parameters ( $\mu_c(s)$ ,  $\phi_c(s)$  and  $p_c(s)$ ) that result from expected prey densities  $\Lambda_c(s)$  at the centroid of each foraging habitat and visualize the predicted distribution of prey densities resulting from the Tweedie distribution. This Tweedie distribution will not exactly match the true sampling distribution because it does not exactly account for spatial heterogeneity within the foraging area (i.e., the Tweedie is calculated from prey densities at the centroid of each area, while the true sampling distribution integrates across this area); the magnitude of this mismatch will decrease as habitats become more homogenous within each foraging area. Finally, we also calculate the optimal value for scale ( $\phi$ ) and power ( $p$ ) when assuming that those parameters are constant for all sites and prey species and visualize the predicted distribution of prey densities resulting from that Tweedie GLM approximation. The latter approximation is what can be fitted conveniently using widely available software, so this comparison allows us to visualize the magnitude of errors arising from ignoring fine-scale spatial heterogeneity and/or specifying a constant value for scale and power parameters.

## Case study demonstrations

We next demonstrate how this approach can be used to solve the inverse problem, that is, provide inference about diet proportions using real-world data. We use two contrasting case studies:

1. Bill-load samples from tufted puffins nesting on Middleton Island in the Gulf of Alaska (Hatch & Sanger, 1992). We bin prey into seven major prey taxa, and apply a Generalized Additive Model (GAM) including a log-linear effect for sea surface temperature 1982–2018 as well as a Gaussian-process smoother for year for each prey.
2. DNA metabarcoding of prey in scat samples from wolves across an archipelagic landscape in southeast Alaska. We bin prey into eight prey taxa, and fit a model with main effects for prey, a Duchon spline for latitude and longitude (representing spatial variation in area swept) as well as a Duchon spline for latitude and longitude estimated for each prey (Roffler et al., 2021).

See Appendices S4 and S5 for further details regarding data and model structure.

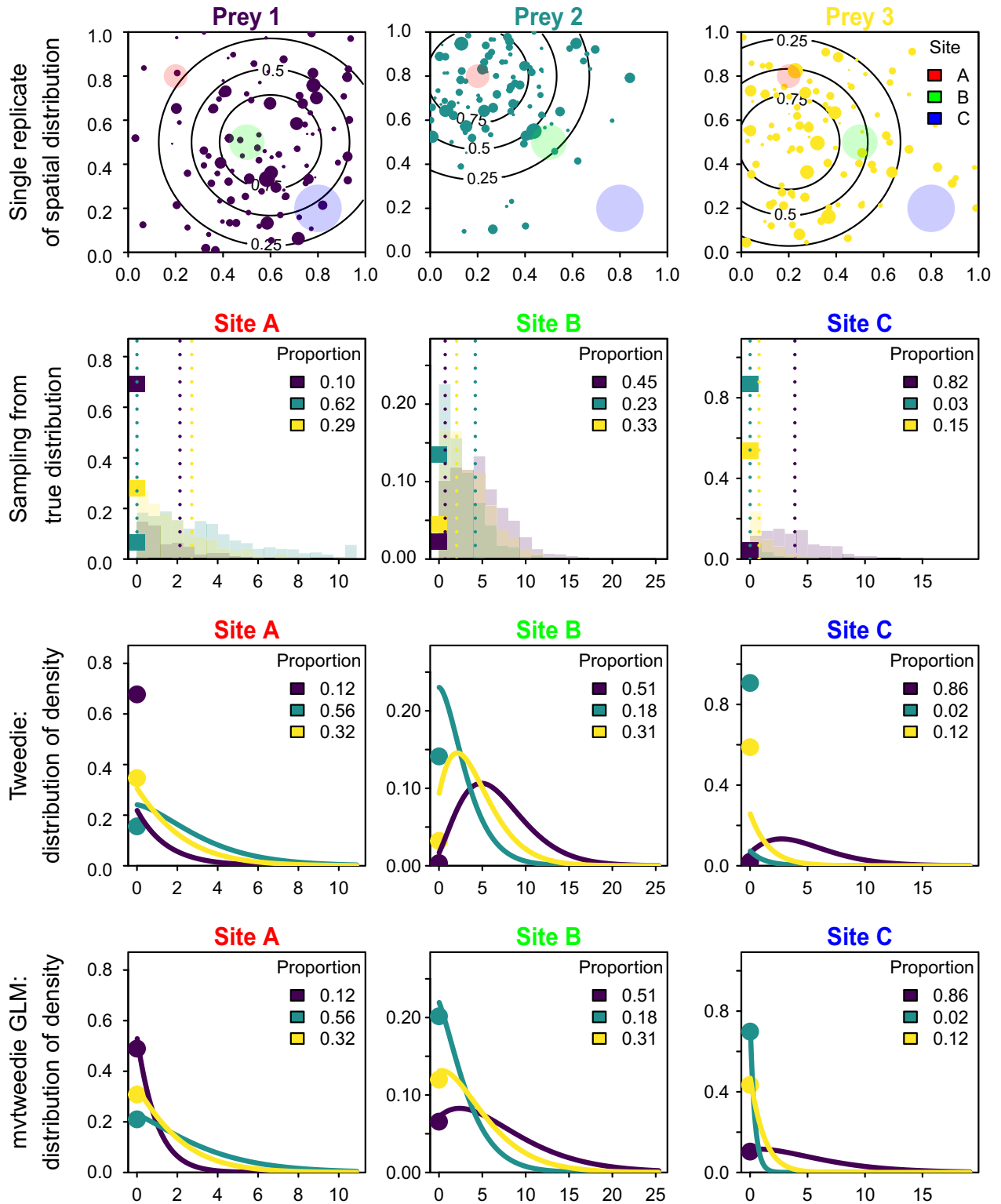
## MODEL FITTING

For all models fitted, we apply GAM fitted using package *mgcv* (Wood, 2006) in the R statistical environment (R Core Team, 2017). We develop a new R package *mvtweedie* that is publicly available (see <https://github.com/James-Thorson-NOAA/mvtweedie>) and that defines an *mvtweedie* S3 class and associated *predict.mvtweedie* function to calculate proportions  $\pi_c^*$  and resulting standard errors given specified values for covariates. We then visualize fits using this predict function and the standard R package *ggplot2* (Wickham, 2016). We provide a simple R script to replicate these analyses as a documented and user-friendly tutorial that can be adopted for future uses (see Appendix S5 for example code and Data S1 and S2 for demo).

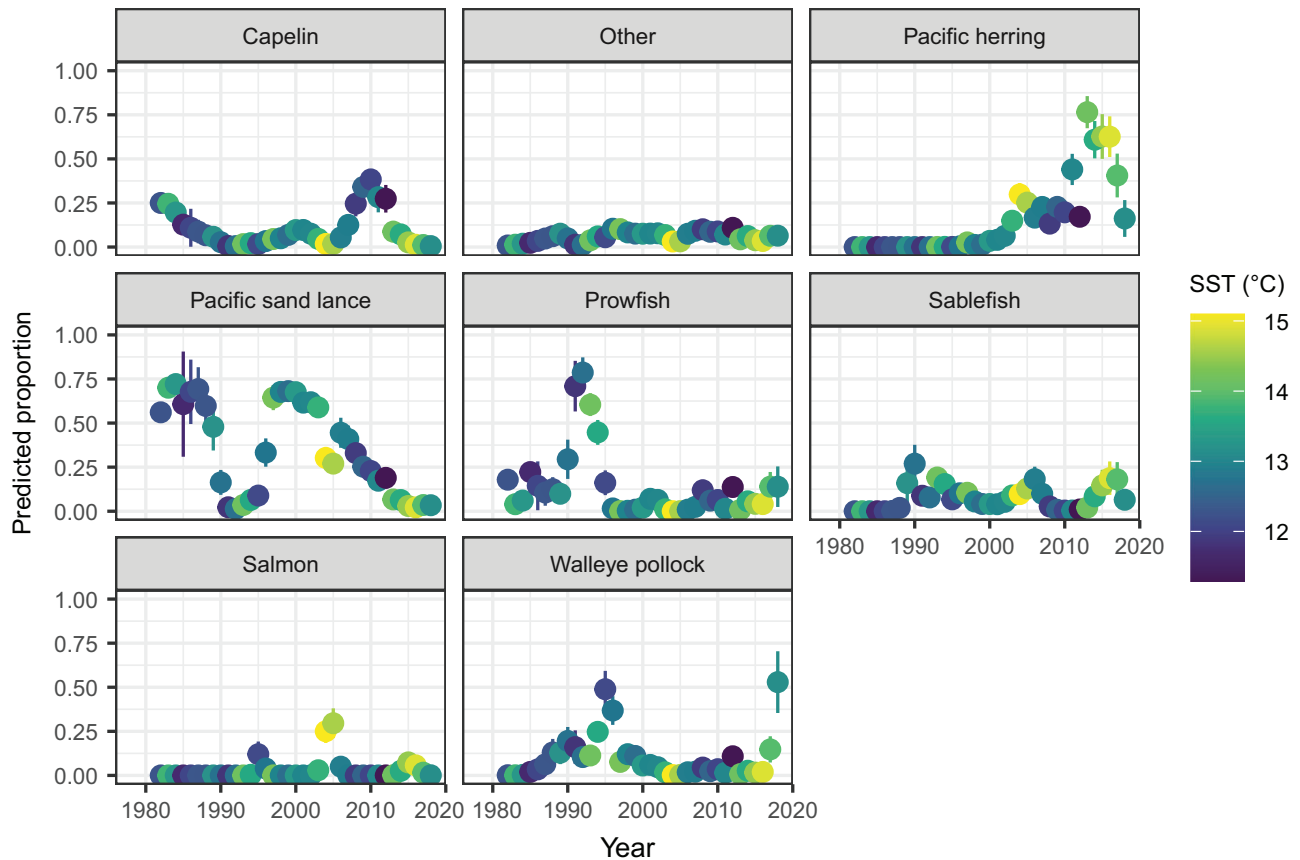
## RESULTS

The simulated example has highest total prey densities at Site A and decreasing densities for Sites B and C (Figure 1, first row). However, Site A also has smallest foraging area, resulting in approximately similar expected feeding rates for all three sites ( $x$ -axis scale of Figure 1, first row). Site A has the highest proportion of Prey 2, while Site B has highest expected proportion for Prey 1, and Site C also has a high proportion for Prey 1. Similarly, Site C has high rates of zero encounters for Prey 2 and 3. Sampling from this prey distribution (Figure 1, second row) shows high rates of encounter for all species in Site B, contrasted with very low encounter rates for Prey 2 and 3 at Site C. A Tweedie distribution ignoring fine-scale spatial heterogeneity (i.e., evaluated using the true prey density at the center of each foraging area) shows similar distributions to the true sampling distribution, but with small differences in the predicted proportion of each prey (i.e., comparing third and second rows of Figure 1). Finally, the Tweedie GLM approximation (using a single value for  $\phi$  and  $p$ ) results in some visible differences relative to the Tweedie distribution using sample-specific parameters, particularly with respect to the predicted proportion of zeros (i.e., comparing height of circles on  $y$ -axis of third vs. fourth rows). However, both approaches result in identical predictions for the proportion for each prey. We, therefore, conclude that approximating the marked Poisson process with a Tweedie distribution while ignoring fine-scale heterogeneity (comparing second and third row) results in a similar magnitude of error as the further approximation using a Tweedie GLM (comparing third and fourth rows).

The resulting model for puffin bill loads shows decadal variability in dominant forage (Figure 2 and Appendix S5:



**FIGURE 1** Simulation of prey distribution (top row) for three prey species (labeled Prey 1, Prey 2, and Prey 3), visualizing the prey density field (contour lines, labeled relative to maximum density at the centroid of each prey’s range) and location for 1000 individuals of each prey (dots with area proportional to individual mass) with the location of three foraging areas overlaid (labeled Site A, Site B, and Site C). Also showing prey densities from 1000 samples from marked Poisson process (squares showing the proportion of zeros, and a histogram for continuous positive values) of prey biomass from food-habits sampling (second row) in those three foraging areas (second row), a Tweedie distribution (showing points at zero for the probability of zero, and a line for continuous positive values) calculated when ignoring fine-scale spatial variation (third row), or a Tweedie distribution when assuming that scale and power parameters are constant for all sites and prey (fourth row). Panels list the true (second row) or predicted (third and fourth rows) proportion from each approximation in the top right of each panel in those rows



**FIGURE 2** Predicted proportion (y-axis) in each year from 1978 to 2018 (x-axis) for each prey (panels) when fitting bill load samples for Tufted Puffins in Middleton Island, showing both prediction (bullets) and  $\pm$ SE (whiskers) given the sea surface temperature (SST) in each year (see color legend on right-hand side)

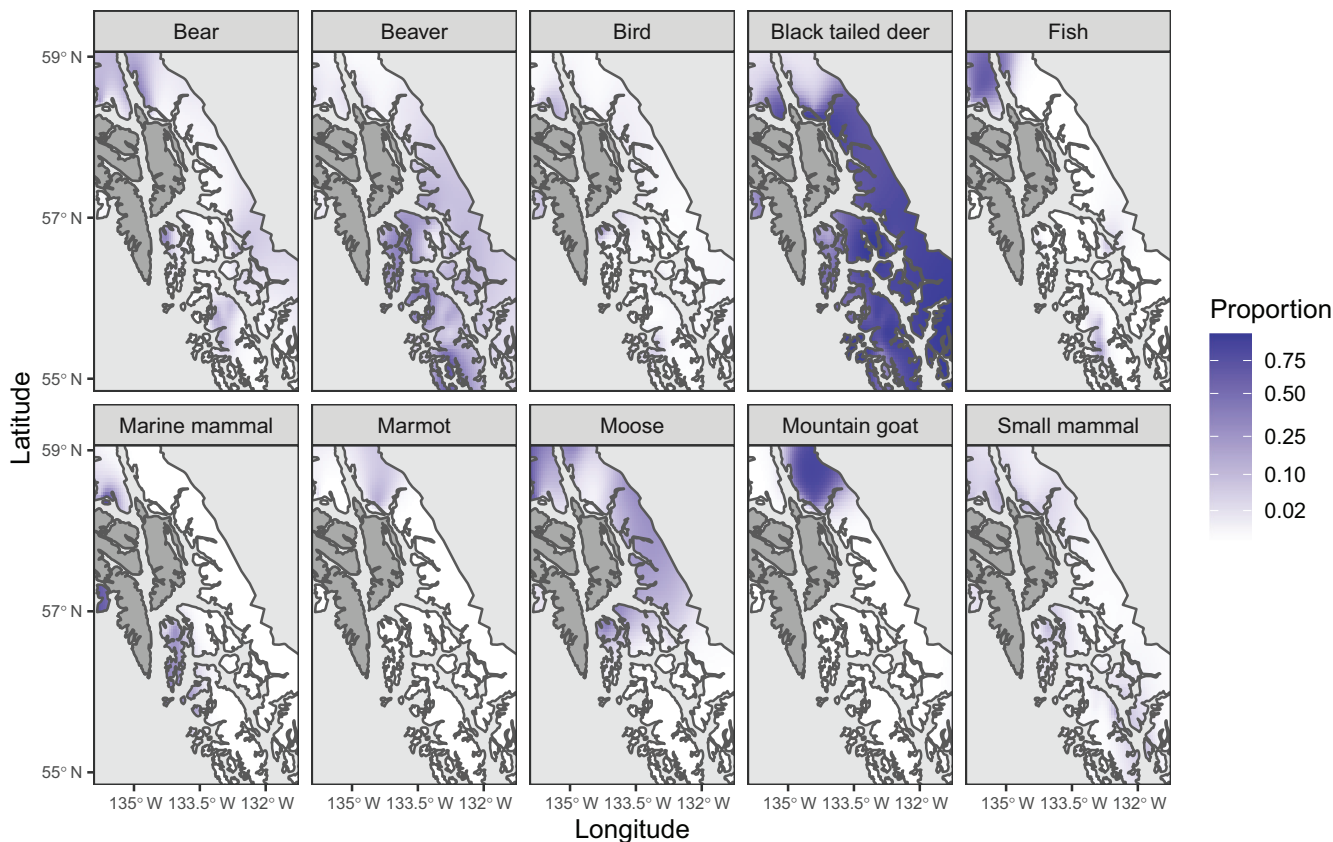
Figure S1). Specifically, Pacific sand lance (*Ammodytes hexapterus*) and prowfish (*Zaprora Silenus*) alternate for highest proportion from 1978–2000, followed by an alternation between Pacific herring (*Clupea pallasii*) and capelin (*Mallotus villosus*) from 2001 to 2018, while salmon (*Oncorhynchus* spp.), pollock (*Gadus chalcogrammus*), and sablefish (*Anoplopoma fimbria*) also show elevated proportions during individual decades. The log-linear effect of sea surface temperature is significantly negative for prowfish and pollock, and significantly positive for Pacific herring (Appendix S5: Table S3 and Figure S2). The positive effect of sea surface temperature on Pacific herring then explains why herring was a higher proportion of puffin diet in recent warm periods (2001–2007 and 2013–2018), while capelin was a larger component in the intervening cool period (2008–2012).

As a contrasting example, wolves in southeast Alaska show strong spatial variation in prey proportions in their diet (Figure 3). Specifically, Sitka black-tailed deer are the dominant prey throughout all areas except the northern edge, where mountain goat and marmot both have higher proportions. Marine mammals show higher proportions along the seaward coastline of northern islands,

while moose are a substantial proportion in the central and northern portions and beavers in the central and southern portions. Collectively, these case studies illustrate that spatial and interannual variation can be easily inferred and visualized using this Tweedie GLM while using common statistical and plotting packages.

## DISCUSSION

In this study, we derived a Tweedie distribution for biomass samples from predator food-habits samples when foraging within a thinned and double-marked Poisson process. We then showed how to parameterize this distribution using either (1) the thinned biomass density for each prey in Equations (2)–(4) or (2) the total forage biomass and proportions for each prey in Appendix S3. Finally, we provided an easy-to-use R package *mvtweedie* that can implement this approach using GAMs (package *mgcv*) or GLMMs (package *glmmTMB*), while visualizing results using high level methods (package *ggplot2*). We provide two contrasting case studies, which highlight decadal variability in prey species for a central-place



**FIGURE 3** Predicted diet proportion (color scale on right-hand side) at each longitude–latitude (x- and y-axes) for each prey (panels) for wolves foraging in southeast Alaska, where predicted proportions are not plotted over water (light gray), Canada (light gray), or islands not inhabited by wolves (Admiralty, Baranof, and Chichagof; dark gray)

forager (Tufted Puffins) and substantial spatial variability for a generalist carnivore (wolves).

The appropriate interpretation of estimates from the Tweedie GLM will depend upon assumptions about the thinning process occurring during predator foraging. Regardless of thinning processes, the predicted proportions ( $\pi_i$  from Equation 4) represent diet fractions for those sampled individuals, and analysts could then infer population-level properties about diet proportions if the sampled individuals are representative of a wider population of predators (Cressie et al., 2009). Alternatively, analysts may seek to interpret variation in diet proportions as indicating variability in the numerical density of those prey species, but this requires additional assumptions about predator thinning rates. Given the thinned point process that underlies this method, prey observed in diets are best understood as a thinned sample from the actual spatial distribution of each prey species. If all prey species of interest are within the diet breadth of an optimally foraging consumer and the most energetically profitable prey are not too abundant, then thinning rates may be constant across space and/or time (Stephens &

Krebs, 1987), such that prey proportions would be representative of spatial/temporal variation in density for each.

Analysts increasingly use diets of generalist predators to understand changes in prey availability across time and space when such information may be otherwise limited (Clare et al., 2014; Grüss et al., 2020). Developing exploratory models with covariates (as facilitated by using common packages such as *mgcv*) can improve the accuracy of predictions across space and time, and provide greater understanding of factors regulating trophic interactions including numerical and functional responses. As one example, a Tweedie-GLM would allow formal tests of prey switching as predators grow. In this case, including an interaction between predator size and prey species (“*formula = ~ ... + Size:group*”, involving numeric vector “*Size*” and factor “*group*”) would permit such a test. In this case, a greater estimated slope for one prey than another indicates a progressive increase in proportion for the former prey with predator size, and the resulting response curve could be visualized as a partial dependence or marginal effects plot. Alternatively, a joint model of diet fractions and predator densities

supports calculation of predator-expanded diet fractions. A joint model of diet and predator density therefore can be used to measure landscape-level consumption, even when analyzing data from monitoring programs wherein diets are sampled in an unbalanced design (Grüss et al., 2020).

Alternatively, the thinning rate in the thinned and marked Poisson process may vary spatially and over time (e.g., if the predator switches from ignoring suboptimal prey in some places or times) and the Tweedie GLM could be adapted to simultaneously estimate spatiotemporal variation in prey densities and thinning rates. However, variation in thinning rate is confounded with variable prey densities; this is analogous to the confounded effect of sampling intensity and population density in presence-only models (Dorazio, 2014). As in the case of presence-only models, identifying the effect of a covariate on both thinning rates and prey densities would presumably require integrating data from multiple sampling methods (e.g., diets and prey density surveys). One advantage of the Tweedie GLM is to explicitly acknowledge the confounding between prey densities and thinning rates, while pointing to monitoring and modeling that can separately identify these two processes underlying diet data.

Ecologists increasingly seek to develop integrated models for all available data without needing to transform data prior to analysis (O'Hara & Kotze, 2010). We therefore encourage further use and development of the Tweedie GLM (and the associated multivariate Tweedie distribution, see Appendix S3) as a numerically efficient approximation to the thinned and marked point process that underlies predator foraging. More generally, the Tweedie GLM is appropriate to model multivariate proportions that include 0s and 1s, which arise in many ecological contexts beyond diet samples, and we hope that the *mvtweedie* package facilitates exploration in these cases.

## ACKNOWLEDGMENTS

We thank B. Chasco for development of Appendix S3: Figure S1, T. Essington, K. Aydin, A. Grüss, and three anonymous reviewers for comments on an earlier draft, and Scott Hatch for decadal sampling of Middleton Island seabird bill loads. Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.


## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT


Data (Thorson et al., 2021a) and code (Thorson et al., 2021b) are available in Dryad (<https://doi.org/10.5061/dryad.08kpr53h>) and Zenodo (<https://doi.org/10.5281/zenodo.5579713>), respectively. The R package (Thorson, 2021) documented here is available in Zenodo (<https://doi.org/10.5281/zenodo.5579705>).

## ORCID

James T. Thorson  <https://orcid.org/0000-0001-7415-1010>

Mayumi L. Arimitsu  <https://orcid.org/0000-0001-6982-2238>

Taal Levi  <https://orcid.org/0000-0003-1853-8311>

Gretchen H. Roffler  <https://orcid.org/0000-0002-8534-3664>

## REFERENCES

- Ainsworth, C. H., I. C. Kaplan, P. S. Levin, and M. Mangel. 2010. "A Statistical Approach for Estimating Fish Diet Compositions from Multiple Data Sources: Gulf of California Case Study." *Ecological Applications* 20: 2188–202.
- Bonin, M., C. Dussault, J. Taillon, N. Lecomte, and S. D. Côté. 2020. "Combining Stable Isotopes, Morphological, and Molecular Analyses to Reconstruct the Diet of Free-Ranging Consumers." *Ecology and Evolution* 10: 6664–76.
- Brooks, M. E., K. Kristensen, K. J. van Benthem, A. Magnusson, C. W. Berg, A. Nielsen, H. J. Skaug, M. Maechler, and B. M. Bolker. 2017. "glmmTMB Balances Speed and Flexibility among Packages for Zero-Inflated Generalized Linear Mixed Modeling." *R Journal* 9: 378–400.
- Clare, E. L., W. O. C. Symondson, H. Broders, F. Fabianek, E. E. Fraser, A. MacKenzie, A. Boughen, et al. 2014. "The Diet of *Myotis Lucifugus* across Canada: Assessing Foraging Quality and Diet Variability." *Molecular Ecology* 23: 3618–32.
- Cressie, N., C. A. Calder, J. S. Clark, J. M. V. Hoef, and C. K. Wikle. 2009. "Accounting for Uncertainty in Ecological Analysis: The Strengths and Limitations of Hierarchical Statistical Modeling." *Ecological Applications* 19: 553–70.
- Dorazio, R. M. 2014. "Accounting for Imperfect Detection and Survey Bias in Statistical Analysis of Presence-Only Data." *Global Ecology and Biogeography* 23: 1472–84.
- Douma, J. C., and J. T. Weedon. 2019. "Analysing Continuous Proportions in Ecology and Evolution: A Practical Introduction to Beta and Dirichlet Regression." *Methods in Ecology and Evolution* 10: 1412–30.
- Foster, S. D., and M. V. Bravington. 2013. "A Poisson–Gamma Model for Analysis of Ecological Non-negative Continuous Data." *Environmental and Ecological Statistics* 20: 533–52.
- Grüss, A., J. T. Thorson, G. Carroll, E. L. Ng, K. K. Holsman, K. Aydin, S. Kotwicki, H. N. Morzaria-Luna, C. H. Ainsworth, and K. A. Thompson. 2020. "Spatio-Temporal Analyses of Marine Predator Diets from Data-Rich and Data-Limited Systems." *Fish and Fisheries* 21: 718–39.
- Hatch, S., and G. Sanger. 1992. "Puffins as Samplers of Juvenile Pollock and Other Forage Fish in the Gulf of Alaska." *Marine Ecology Progress Series* 80: 1–14.
- Illian, J. B., and D. F. R. P. Burslem. 2017. "Improving the Usability of Spatial Point Process Methodology: An Interdisciplinary Dialogue between Statistics and Ecology." *ASTA Advances in Statistical Analysis* 101: 495–520.



- Illian, D. J., P. A. Penttinen, D. H. Stoyan, and D. D. Stoyan. 2008. *Statistical Analysis and Modelling of Spatial Point Patterns*. Hoboken, NJ: John Wiley & Sons.
- Maier, M. J. 2021. "DirichletReg: Dirichlet Regression." R package version 0.7-1. <https://github.com/maiermarco/DirichletReg>
- Moriarty, P. E., T. E. Essington, and E. J. Ward. 2016. "A Novel Method to Estimate Prey Contributions to Predator Diets." *Canadian Journal of Fisheries and Aquatic Sciences* 74: 1–10.
- O'Hara, R. B., and D. J. Kotze. 2010. "Do Not Log-Transform Count Data." *Methods in Ecology and Evolution* 1: 118–22.
- R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Roffler, G. H., J. M. Allen, A. Massey, and T. Levi. 2021. "Wolf Dietary Diversity in an Island Archipelago." *Bulletin of the Ecological Society of America* 102: 1–6.
- Shi, Y., Y. Hoareau, E. M. Reese, and S. K. Wasser. 2021. "Prey Partitioning between Sympatric Wild Carnivores Revealed by DNA Metabarcoding: A Case Study on Wolf (*Canis Lupus*) and Coyote (*Canis Latrans*) in Northeastern Washington." *Conservation Genetics* 22: 293–305.
- Stephens, D. W., and J. R. Krebs. 1987. *Foraging Theory*, First ed. Princeton, NJ: Princeton University Press.
- Thorson, J. 2021. "James-Thorson-NOAA/mvtweedie: Adding Zenodo for DOI (1.0.1)." *Zenodo*, software. <https://doi.org/10.5281/zenodo.5579705>
- Thorson, J. T., M. L. Arimitsu, T. Levi, and G. H. Roffler. 2021a. "Data and Code to Replicate: Diet Analysis Using Generalized Linear Models Derived from Foraging Processes Using R Package mvtweedie." *Dryad*, data set. <https://doi.org/10.5061/dryad.08kpr53h>.
- Thorson, J. T., M. L. Arimitsu, T. Levi, and G. H. Roffler. 2021b. "Data and Code to Replicate: Diet Analysis Using Generalized Linear Models Derived from Foraging Processes Using R Package mvtweedie." *Zenodo*, software. <https://doi.org/10.5281/zenodo.5579713>.
- Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*, Second ed. Houston, TX: Springer International Publishing.
- Wood, S. N. 2006. *Generalized Additive Models: An Introduction with R*, First ed. Boca Raton, FL: Chapman and Hall/CRC Press.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Thorson, James T., Mayumi L. Arimitsu, Taal Levi, and Gretchen H. Roffler. 2022. "Diet Analysis Using Generalized Linear Models Derived from Foraging Processes Using R Package *mvtweedie*." *Ecology* 103(5): e3637. <https://doi.org/10.1002/ecy.3637>