Systematic bias in studies of consumer functional responses

Mark Novak[1] & Daniel B. Stouffer[2†]

[1]Department of Integrative Biology, Oregon State University,
Corvallis, Oregon, 97331 USA
[2]Centre for Integrative Ecology, School of Biological Sciences,
University of Canterbury, Christchurch 8140, New Zealand

[†] Both authors contributed equally to this work.

RUNNING TITLE: Biased functional-response inferences

Article type	Letter
Words in abstract	159
Words in manuscript	4897
Number of references	98
Number of figures	4
Number of tables	2
Number of boxes	2
Corresponding author	Mark Novak
Phone	(541) 737-3610
Fax	(541) 737-3120
Email	mark.novak@oregonstate.edu

Abstract

Functional responses are a cornerstone to our understanding of consumer-resource interactions, so how to best describe them using models has been actively debated. Here we focus on the consumer dependence of functional responses to evidence systematic bias in the statistical comparison of functional-response models and the estimation of their parameters. Both forms of bias are universal to nonlinear models (irrespective of consumer dependence) and are rooted in a lack of sufficient replication. Using a large compilation of published datasets, we show that – due to the prevalence of low sample size studies – neither the overall frequency by which alternative models achieve top rank nor the frequency distribution of parameter point estimates should be treated as providing insight into the general form or central tendency of consumer interference. We call for renewed clarity in the varied purposes that motivate the study of functional responses, purposes that can compete with each other in dictating the design, analysis and interpretation of functional-response experiments.

Keywords: *nonlinear species interactions, mutual predator effects, predator-dependence, prediction, Fisher information, model flexibility, parameter estimation, model comparison*

Introduction

Functional responses describe how a consumer's feeding rate responds to its prey's abundance as determined by behaviour and the composition and conditions of their community and abiotic environment (Abrams & Ginzburg, 2000; Solomon, 1949; Uiterwaal & DeLong, 2020). Applications of the functional-response concept are found in pest control, invasive species management, and conservation biology, and are fundamental to theory in population, community and evolutionary ecology. Much effort has therefore been devoted to experiments designed to measure and statistically compare the many mathematical models that have been developed to characterize consumer functional responses, with theory evidencing often dramatic differences in the biological inferences and predictions that alternative model forms and parameter values provide (Aldebert & Stouffer, 2018; Arditi & Ginzburg, 2012; Coblentz & DeLong, 2020; Fussmann & Blasius, 2005).

A dominant focus of the literature on functional responses has been the discussion of so-called prey-, ratio-, and consumer-dependent models (Abrams, 1994; 2015; Abrams & Ginzburg, 2000; Arditi & Ginzburg, 2012; Barraquand, 2014; Tyutyunov & Titova, 2020). These respectively describe feeding rates as responding to prey abundance alone, to the ratio of prey and consumer abundances, or to both prey and consumer abundances more generally. Knowing which of these models best characterize consumer feeding rates can be important because, for example, predictions of population and food-web dynamics are qualitatively altered by either the structural replacement of the classically-assumed prey-dependent Type II model of Holling (1959) for the corresponding ratio-dependent model of Arditi & Ginzburg (1989), or by a parametric change to the strength of mutual consumer effects encapsulated in

the consumer-dependent models of Arditi & Akçakaya (1990), Beddington (1975) and DeAngelis *et al.* (1975), and Crowley & Martin (1989).

Although a tremendous amount remains to be learned about consumer functional responses (e.g., Baudrot *et al.*, 2016; Koen-Alonso, 2007; Novak *et al.*, 2017; Uiterwaal & DeLong, 2020), the long-standing debate regarding the relative merits of prey- versus ratio-dependent models in particular is now largely considered by many to be pragmatically resolved: the most recent synthetic assessments of published experiments conclude that most consumer feeding rates respond to both prey and consumer densities, and that they do so in a more general manner than described by exact ratio dependence (DeLong & Vasseur, 2011; Osenberg *et al.*, 1999; Skalski & Gilliam, 2001). More specifically, interference among consumers is considered common, variable in strength, and tending to magnitudes that are "intermediate" to exact prey- and ratio-dependence (Arditi & Ginzburg, 2014; DeLong & Vasseur, 2011; Osenberg *et al.*, 1999). Neither prey- nor ratio-dependence is therefore considered an adequate descriptor of consumer feeding rates (but see Abrams, 2015; Arditi & Ginzburg, 2012; 2014; Tyutyunov & Titova, 2020).

Here we use the perceived resolution of the debate over the consumer-dependent nature of functional responses to bring to light more general issues inherent to the statistical merger of theory and data obtained from functional-response experiments. Our focus is on the use and interpretation of information-theoretic model comparisons (Box 1) and model-specific parameter estimates (Box 2), rather than mathematical analyses of plausibility or the philosophical underpinnings of alternative deterministic model formulations (c.f. Abrams, 2015; Arditi & Ginzburg, 2012; Barraquand, 2014). To assess the extent and severity of these expected statistical issues, we followed prior syntheses of consumer dependence

by Hassell & Varley (1969), Hassell (1971), Arditi & Akçakaya (1990), Skalski & Gilliam (2001), and DeLong & Vasseur (2011) and compiled a total of 77 functional-response datasets, incorporating many studies not considered by or available to earlier compilations. Because prior syntheses either used potentially suboptimal statistical methods or included published parameter estimates from across studies using different statistical methods (see summary in Supplementary Materials S2), we re-analyzed all datasets by fitting to them the suite of commonly considered functional-response models and comparing those fits with a now standard information-theoretic approach.

Our analyses reveal systematic bias regarding the inferred density-dependent nature and strength of consumer interference. Bias is driven by variation in experimental sample sizes: studies with smaller sample sizes are biased towards inferences of stronger interference and of particular functional-response model forms. In highlighting the bias for functional-response models of consumer interference, our analyses bring to light the universal but unappreciated bias that small sample sizes introduce to information-theoretic model comparisons and maximum-likelihood parameter estimates for nonlinear functional-response models in general (Boxes 1 & 2). Since all models are wrong (Box, 1976), our primary objective is not to reignite debate regarding prey-, ratio-, and consumer dependence (though we acknowledge this may happen). Rather, we call for clarity and deeper discussion of the varied purposes that motivate the empirical study of consumer functional responses, purposes that can compete with each other in dictating the analysis and interpretation of functional-response experiments.

Box 1. Bias in information-theoretic model comparisons. Information-theoretic model comparisons have become a mainstay in ecology, as well as in the context of judging the relative performance of alternative functional-response models. By now the most frequently used information criterion is the Akaike information criterion, *AIC* (Akaike, 1973; 1974). *AIC* estimates the Kullback-Leibler divergence of a focal model (whose performance in fitting a set of data is being judged) from the "true" model responsible for generating the data. Asymptotically, it minimizes out-of-sample prediction error. A model's *AIC* reflects the balance between its ability to describe the data as per its likelihood and its complexity as judged by its number of parameters k, thereby justifying comparisons of performance for models that differ in complexity. By parsimony, models with smaller *AIC* are deemed to outperform models with larger *AIC*.

However, *AIC* is only an approximately unbiased estimator of the Kullback-Leibler divergence when k is small relative to the sample size n of the data, and will be negatively biased towards smaller values when n is small (Akaike, 1974; Burnham & Anderson, 2002). Sugiura (1978) and Hurvich & Tsai (1989) introduced the corrected Akaike information criterion, $AIC_c = AIC + 2(k^2 + k)/(n - k - 1)$, to correct for this bias by imposing a greater penalty on a model's parametric complexity. Because AIC_c converges on *AIC* as sample size increases, it is generally the preferred criterion to use. However, the standard equation for AIC_c was derived in a context of linear regression models with Gaussian errors and remains a biased estimator for nonlinear models and non-Gaussian models, especially when sample sizes are small (Burnham & Anderson, 2002). Even AIC_c may therefore be biased towards more complex functional-response models when sample sizes are small.

Furthermore, a model's complexity is not just a function of its number of parameters but also of its functional form. Two models with an equal number of parameters can combine these in different ways, leading to differing degrees of flexibility (Gutenkunst *et al.*, 2007; Ly *et al.*, 2017). This form of complexity (a.k.a. geometric complexity) may be considered problematic in the context of assessing model performance because a model with greater flexibility will have larger estimation uncertainty associated with each of its parameters' point estimates that together "conspire to cancel each other out for the special purpose of fitting the observed data" (Myung, 2000). Criteria such as *AIC* and AIC_c that use only the maximum-likelihood point estimates ignore this flexibility, the effect of which diminishes with increasing sample size for alternative criteria that do consider it (Myung, 2000; Myung *et al.*, 2006). Multiplicative models models like the Arditi-Akçakaya model are generally expected to exhibit greater flexibility than additive models like the Beddington-DeAngelis model (Fig. 1, see Supplementary Materials S3), hence are expected to be preferentially selected by *AIC* and AIC_c when sample sizes are small.

Figure 1 here.

Box 2. Bias in maximum-likelihood parameter estimates. The sample size dependent bias of maximum-likelihood parameters estimates for nonlinear models differs from the model-comparison bias discussed in Box 1. We illustrate the general problem using a simple example co-opted from Box (1971), then show its relevance to the estimation of functional-response parameters.

Box (1971) considered the issue of estimating the rate parameter θ for a model of exponential decay,

$$y_i = e^{-\theta t_i} + \epsilon_i, \tag{1}$$

where all $i = 1, 2, \ldots, n$ observations of response variable y_i are made at the same time ($t^* = t_1 = t_2 = ... = t_n$) and the ϵ_i residuals are assumed to be independent and described by a Gaussian distribution with a mean of zero and a variance of σ^2. The value of θ that maximizes the likelihood of the data is

$$\hat{\theta} = \frac{-\ln(\bar{y})}{t^*} = \frac{\ln\left(\frac{1}{\bar{y}}\right)}{t^*}, \tag{2}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the arithmetic mean of all observations. Since all observations were taken at the same time, $\bar{y}$ will follow a Gaussian distribution whose mean of $\exp(-\theta t^*)$ will have a standard error of $\sigma_{\bar{y}} = \sigma/\sqrt{n}$. Because the numerator of eqn. 2 is a concave-up function of $\bar{y}$, Jensen's inequality dictates that $\hat{\theta}$ will be biased upward with respect to the true value of θ unless $\sigma_{\bar{y}} = 0$. Estimation bias is therefore an intrinsic consequence of observation noise or variation in the assumed decay process, but becomes smaller as sample sizes increase.

Equivalent variation-induced bias occurs for the parameter estimates of all nonlinear functional response models (see Supplementary Materials S6 for details and derivations of qualitative bias in representative models). For example, for the Arditi-Akçakaya model on which most prior inference of consumer interference has focused, the maximum-likelihood estimator for the interference strength m when eaten prey are replaced is

$$\hat{m} = \frac{\ln\left(\alpha N\left(\frac{PT}{\bar{y}} - h\right)\right)}{\ln P}, \tag{3}$$

where $\bar{y}$ is the arithmetic mean number of eaten prey and all other parameters are defined in Table 1. Because $\bar{y}$ cannot exceed the possible number of prey eaten, PT/h, the numerator of eqn. 3 is positive and is also a concave-up function of $\bar{y}$. This makes $\hat{m}$ biased upward with respect to the true value of m unless there is no residual variation among experimental replicates or sample sizes are large, regardless of whether the numbers of prey eaten in the course of an experiment are assumed to be samples from a Poisson or Gaussian distribution. The estimator for m is similarly biased when the proportions of prey eaten are assumed to be drawn from a binomial distribution for non-replacement studies. In fact, for non-replacement studies, even the estimator for the attack rate of linear functional response models is biased. Since maximum-likelihood estimators derived assuming a Gaussian likelihood are equivalent to the estimators of least squares regression, bias will occur when applying the latter as well. Further, since maximum-likelihood estimators derived under a log-normal likelihood are a function of the geometric mean y rather than the arithmetic mean, and the geometric mean will always be less than the arithmetic mean, estimation bias for parameters like m will be compounded when using regression on log-transformed counts of prey eaten.

Methods

Data compilation and functional-response models

We considered all published functional-response datasets of predator and parasitoid consumers having variation in both prey (host) and consumer abundances, the majority of which represented manipulative experiments (Table S1). When possible we obtained the original data from the authors. Otherwise we extracted data points, or means and associated uncertainty estimates, from the publication, either from provided tables or using GraphClick (2010). A few published studies were not included in our analyses because, for example, prey experienced significant mortality not attributable to predation (e.g., Schenk *et al.*, 2005; Stier *et al.*, 2013; Stier & White, 2014) or because observations were not all independent (e.g., Blowes *et al.*, 2017; Cresswell, 1998), which necessitated tailored models beyond the scope of our analyses. Similarly, a few studies in which prey abundances or the numbers of prey eaten were quantified as non-integer values (i.e. densities or feeding rates, e.g., Hebblewhite 2013) were not included since these necessitated alternative statistical models, though for most such previously analyzed studies these data could be converted to counts (i.e. number of prey available and eaten in a given amount of time, e.g., Vucetich *et al.* (2002)). Unlike some prior syntheses (Arditi, *pers. comm.*; see also Abrams, 2015), we did not exclude datasets in which significant prey-depletion occurred (but see Fig. S1). Datasets entailing consumers feeding on multiple prey species were included and treated as separate prey-specific datasets only when this was done by the original authors (e.g., Chan *et al.*, 2017).

We considered a total of eight commonly considered deterministic functional-response models (Table 1): the prey-dependent Holling Type I and Type II forms (Holling, 1959), the corresponding ratio-

dependent forms (Arditi & Ginzburg, 1989; Sutherland, 1983), and the more general consumer-dependent forms of Beddington (1975) and DeAngelis *et al.* (1975), Crowley & Martin (1989), Hassell & Varley (1969), and Arditi & Akçakaya (1990). Of the two Holling-like consumer-dependent models, the Beddington-DeAngelis model assumes interference occurs only when consumers are searching for prey, while the Crowley-Martin model assumes it occurs equally during both searching and handling. Of the two ratio-like consumer-dependent models, the Hassell-Varley model assumes that handling times are inconsequential and the Arditi-Akçakaya model does not; both assume that interference among consumers occurs through a generic multiplicative process. We did not consider any of the many additional nor less-phenomenological functional-response models that have been proposed (e.g., Abrams, 2010; Jeschke *et al.*, 2002), including parasitoid-specific models (e.g., Rogers, 1972). Hence, for some parasitoid datasets and models, attack rate estimates are better interpreted as reflecting encounter rates (Arditi, 1983).

Model-fitting and performance

We fit all functional-response models to each dataset by first obtaining maximum-likelihood parameter estimates using the *sbplx* optimization function of the *nloptr* R package (Johnson, 2020), then passing these to the *mle2* function of the *bbmle* package (Bolker, 2020) using the default optimizer to estimate parameter uncertainty. For each dataset, we assumed one of two statistical models: a Poisson likelihood for studies in which eaten prey were continually replaced, and a binomial likelihood for studies in which they were not. For non-replacement studies, we generated predictions for the number of eaten prey using the Lambert W function (Lehtonen, 2016) after confirming that numerical integration using *odeintr:rk4*

Table 1: The eight deterministic functional-response models we considered for describing the rate at which prey (hosts) are eaten (parasitized). Variables N and P refer to the abundances of prey (hosts) and predators (parasitoids). Parameters a and α respectively refer to the attack rates of the Holling- and ratio-like models; their dimensions differ (Arditi & Ginzburg, 2012). Parameters c and m respectively represent interference strengths for Holling- and ratio-like models; we use separate symbols because their dimensions also differ. Note that m is constrained to positive values in the formulations used here; much previous work uses equivalent formulations in which m is negative or is made positive by explicit negation. When fitting the BD and CM models, we replaced P with $P - 1$ for datasets in which P reflected a count rather than density. The number of parameters, k, refers to the number of free parameters in each model because only these determine the mean and variance of the Poisson and binomial statistical models we employed. For replacement studies to which the Poisson was applied, the predicted count of prey eaten corresponds to the functional response multiplied by P and the time period of the experiment or observation window, T. For non-replacement studies to which the binomial was applied, the predicted proportion of prey eaten by P consumers after time period T was solved for analytically (see Supplementary Materials S6).

Name	Abbreviation	Model	Parameters (k)
Holling Type I	H1	aN	1
Holling Type II	H2	$\frac{aN}{1+ahN}$	2
Beddington-DeAngelis	BD	$\frac{aN}{1+ahN+cP}$	3
Crowley-Martin	CM	$\frac{aN}{(1+ahN)(1+cP)}$	3
Linear Ratio-dependent	LR	$\frac{\alpha N}{P}$	1
Arditi-Ginzburg	AG	$\frac{\alpha N}{P+\alpha hN}$	2
Hassell-Varley	HV	$\frac{\alpha N}{P^m}$	2
Arditi-Akçakaya	AA	$\frac{\alpha N}{P^m+\alpha hN}$	3

(Keitt, 2017) gave equivalent results. All parameters were constrained to be positive by exponentiation (Bolker, 2008). For observational field studies, we followed the original author's assumed prey replacement scenario (Table S1). Parasitoids were classified as either discriminatory or non-discriminatory, with studies of non-discriminatory parasitoids being treated as replacement studies (Table S1). Data available only in the form of means and uncertainties were analyzed by a bootstrapping procedure in which 250 new datasets were created assuming either a treatment-specific binomial or Poisson process as dictated by the study's replacement of prey. More specifically, for a treatment level of N prey having n replicates and a mean μ and standard error s prey eaten, we drew n random values x from a Gaussian distribution,

$\mathcal{N}(\mu, s\sqrt{n})$, to parameterize a Poisson distribution (setting $x < 0$ to $x = 0$) or binomial distribution (setting $x < 0$ to $x = 0$ and $x > N$ to $x = N$) to generate integer counts of prey eaten. We considered the median of the resulting distribution of parameter estimates as the point estimate. In one case where no estimates of uncertainty were available (Katz, 1985), we treated the available means as raw data, as done in prior syntheses.

We assessed the relative performance of the fitted models for each dataset by ranking them in two ways: by their AIC_c and by their mean absolute deviation (*MAD*). Despite its assumptions (Box 1), AIC_c is typically the preferred information criterion when sample sizes are small or vary across datasets; it converges on the most widely-used information criterion, *AIC*, as sample size increases. As with *AIC*, models having AIC_c values within 2 units of the best-performing model are considered to have equivalent support (Burnham & Anderson, 2002). In theory, *AIC* is asymptotically optimal in the sense of minimizing a model's out-of-sample mean squared error (*MSE*). *MAD* is similar to *MSE* but describes the within-sample differences between observations and predictions, and is more easily interpreted by not differentialy re-scaling deviations of different magnitude the way that squaring deviations does. For the additional purpose of judging model performance in a relative goodness-of-fit sense, we also evaluated each model's *MAD* as a percentage of the *MAD* of the best-performing model having the lowest *MAD*, arbitrarily considering a 1% difference as reflecting a negligible difference in model performance. We then looked for systematic variation across all datasets in their rankings of alternative models by ordering datasets by their sample size. Sample size was defined as the total number of experimental replicates (within and across treatments) or observations (e.g., time points) as defined by the original authors.

Parameter inference

Past synthetic assessments of consumer dependence fall into two categories: those that evaluated relative support for alternative model forms (as above, e.g., Skalski & Gilliam, 2001), and those that investigated magnitudes of interference as estimated using the consumer-dependent Arditi-Akçakaya model (e.g., Arditi & Ginzburg, 2012; DeLong & Vasseur, 2011). The Arditi-Akçakaya model combines the saturating form of the Holling Type 2 and Arditi-Ginzburg models with the multiplicative interference effects of the Hassell-Varley model. Its multiplicative nature makes the interference strength parameter, m, dimensionless (unlike all other parameters of this and all other models). This permits direct comparisons across studies, with estimates of $m = 0$ evidencing perfect prey dependence and estimates of $m = 1$ evidencing perfect ratio dependence. Therefore, and although sample size bias may occur for the parameters of all nonlinear models (Box 2, Supplementary Materials S6), we focused our assessment of bias on m, plotting estimates as a function of each study's sample size and other covariates (i.e. predator or parasitoid, replacement or non-replacement, no depletion or depletion, raw or bootstrapped data). To assess the specificity of inferences to our chosen statistical approach, we also repeated our estimation of m by employing the method used by Arditi & Akçakaya (1990) for all datasets where this method could be applied (see Supplementary Materials S4).

Past assessments of interference strengths have relied on point estimates alone, just as is implicitly done when model performance is assessed using information criteria such as AIC_c (Box 1). However, assessing the uncertainty of parameter estimates is equally important when drawing inferences (Aldebert & Stouffer, 2018; Osenberg *et al.*, 1999). We determined standard errors in one of three ways: For non-

bootstrapped datasets, we first attempted the *confint* profiling method of *bbmle* (Bolker, 2020) assisted by the *SEfromHessian* function of the *HelpersMG* package (Girondot, 2020). On occasion, this failed because the optimization converged but the likelihood surface was nearly flat around the optimum. In these cases, we resorted to the quadratic approximation of *mle2* (Bolker, 2020). For bootstrapped datasets, we used the 16th and 84th quantiles of the distribution of estimates (Efron & Tibshirani, 1994).

Results

Model performance

As judged by AIC_c, the Crowley-Martin, Arditi-Akçakaya and Arditi-Ginzburg models were respectively each ranked first by 23 (29.9%), 17 (22.1%) and 17 (22.1%) of the 77 datasets (Fig. 2A, Table 2). This was followed by the Beddington-DeAngelis model which ranked first for 11 (14.3%) datasets. No datasets ranked the linear ratio-dependent model as first. The Beddington-DeAngelis and Arditi-Akçakaya models were the most frequent models to be ranked second at a respective 35 (45.5%) and 20 (26.0%) times. In total, 46 (59.7%) datasets had a single model outperform all others by more than two AIC_c units. For the 23 datasets where the Crowley-Martin model was ranked first, the Beddington-DeAngelis, Arditi-Akçakaya, and Arditi-Ginzburg models were each within two AIC_c units for one dataset. For the 17 datasets where the Arditi-Akçakaya model was ranked first, at least one of three models – the Beddington-DeAngelis, the Crowley-Martin, and the Arditi-Ginzburg model – were within two AIC_c units a total of 6 (35.3%) times.

Sample sizes varied greatly among the datasets and were skewed towards smaller studies (range = 10–528, median = 80, mean = 115.1, Fig. 2). The smaller their sample size, the greater was the proportion

Table 2: The number of datasets for which each functional response model achieved a given rank relative to all other models as judged by AIC_c or by MAD. Models having equivalent support to the top-ranked model are indicated in Fig. effig:Ranks.

Rank	H1	H2	BD	CM	LR	HV	AG	AA
				AIC_c				
1	2	2	10	23	0	5	17	18
2	0	1	35	12	1	1	8	19
3	0	2	18	16	2	4	8	27
4	1	13	9	20	0	1	22	11
5	3	28	2	3	4	24	12	1
6	9	4	3	3	18	37	2	1
7	34	19	0	0	12	5	7	0
8	28	8	0	0	40	0	1	0
				MAD				
1	1	3	18	26	0	3	1	25
2	0	3	34	13	0	2	7	18
3	0	4	18	12	0	1	17	25
4	1	11	4	19	3	3	28	8
5	4	28	1	4	5	22	13	0
6	12	8	2	3	12	35	4	1
7	31	17	0	0	11	11	7	0
8	28	3	0	0	46	0	0	0

of datasets that ranked the Arditi-Ginzburg, Hassell-Varley, or Holling Type II models as first (Fig. 3A). Thus, for example, for the 39 datasets having sample sizes equal to the median sample size or less, the Arditi-Ginzburg model was ranked first 15 (38.5%) times while the Crowley-Martin, Arditi-Akçakaya and Beddington-DeAngelis models were ranked first a respective 12 (30.8%), 5 (12.8%) and 1 (2.6%) times, and were ranked second a respective 3 (7.7%), 13 (33.3%) and 15 (38.5%) times. Correspondingly, the more that datasets with larger sample sizes were considered, the greater was the proportion that ranked either the Crowley-Martin, Arditi-Akçakaya or Beddington-DeAngelis models as first (Fig. 3A), and the Beddington-DeAngelis model as second (Fig. 3B). Datasets with larger sample sizes were somewhat better able to discriminate among models than were smaller datasets, with 19 (48.7%) of the 39 smaller datasets and 27 (71.1%) of the 38 larger datasets having a single model outperform all others by more than two AIC_c units (Fig. 2).

As judged by *MAD*, the Arditi-Akçakaya model ranked first for 28 (36.4%) of all considered datasets (Fig. 2B, Table 2). This was followed by the Crowley-Martin and Beddington-DeAngelis models which respectively ranked first for 24 (31.2%) and 18 (23.2%) datasets. The Beddington-DeAngelis and Arditi-Akçakaya models were the most frequent models to be ranked second at a respective 34 (44.2%) and 18 (23.4%) times. According to *MAD*, no datasets ranked the linear ratio-dependent model as first and one ranked the Arditi-Ginzburg model as first. Forty-six (59.7%) datasets had at least one model producing a *MAD* within 1% of the first-ranked model. Of the 28 datasets for which the Arditi-Akçakaya model was ranked first, the Beddington-DeAngelis, Crowley-Martin, and Holling Type II models produced *MAD* values within 1% a respective 17 (60.7%), 8 (28.6%), and 8 (28.6%) times. The Arditi-Ginzburg,

Hassell-Varley, and linear ratio-dependent models did so a respective 7 (25.0%), 2 (7.1%), and 1 (3.6%) times. The *MAD* of more complex models was not always less than that of their corresponding simpler models, in contrast to their likelihood (underlying AIC_c) which must improve with model complexity.

For *MAD*, the smaller their sample size, the greater was the proportion of datasets that ranked the Arditi-Akçakay or Hassell-Varley models as first (Fig. 3C), and the Arditi-Akçakay model as second (Fig. 3D). Correspondingly, the more that datasets with larger sample sizes were considered, the greater was the proportion of datasets that ranked either the Crowley-Martin or Beddington-DeAngelis models as first, and the Beddington-DeAngelis model as second. Datasets with larger sample sizes were no better in their ability to discriminate among models than were smaller datasets, with 16 (41.0%) of the 39 smaller datasets and 15 (39.5%) of the 38 larger datasets having a single model outperform all others by the 1% criterion (Fig. 2B).

Parameter inference

Magnitudes of consumer interference as estimated assuming the Arditi-Akçakaya model ranged from $m \approx 0$ to 3.8 with an overall mean of 0.94 ($\pm$ 0.62 std. dev.). However, studies with smaller sample sizes generated larger interference estimates (Fig. 4) such that, for example, the mean estimate of studies with a sample size exceeding the median sample size was 0.71 ($\pm$ 0.44 std. dev.) with a range of $m \approx 0$ to 1.93. No other covariate was similarly associated with the estimated strength of consumer interference (Fig. S1). Although uncertainty around point estimates also tended to increase the lower the sample size, this increase occurred at a slower rate than did that of the point estimate magnitudes; most large-magnitude estimates associated with small sample sizes were therefore well constrained. Use of the

Arditi & Akçakaya (1990) method of interference strength estimation did not alter inferences of sample size bias for the subset of datasets to which it could be applied (Figs. S2-S3).

Discussion

Past appeals for a data-driven resolution to the debate over consumer dependence have been answered by a doubling of the number of published datasets since the clarion call of Abrams & Ginzburg (2000). Barring an exceptionally strong publication bias (Abrams, 2015; Osenberg *et al.*, 1999), our analyses of the accumulated data substantiate the pragmatic resolution of the debate's primary focus: consumer dependence is common, and neither exact prey- nor ratio-dependent models reflect statistically-parsimonious descriptors (as judged by AIC_c and *MAD*) of the density dependence exhibited by most consumer feeding rates. Our analyses also confirm prior syntheses of prey-only experiments, showing that consumer feeding rates are almost invariably a saturating function of prey densities (Jeschke *et al.*, 2004).

Our primary new insight, however, is the existence of strong statistical bias in the ability of the available data to help characterize and quantify consumer dependence. First, inadequate replication biases the standard information-theoretic comparison of model performance towards the models of Hassell & Varley (1969), Arditi & Ginzburg (1989), and, by extension, their generalized combination by Arditi & Akçakaya (1990). Our results on AIC_c, together with those on mean absolute deviation, are consistent with this bias being due to a combination of the (mis-)estimated parametric and geometric complexity that these models entail (Box 1). This bias added to the poor identifiability of m at low sample sizes which caused AIC_c to (correctly) give less support to the Arditi & Akçakaya (1990) model than the model of Arditi & Ginzburg (1989) for which $m = 1$. Second, the estimation of interference strengths

by application of the long-established method of maximum likelihood to the model of Arditi & Akçakaya (1990) is biased by inadequate replication towards larger-than-true values of interference strength. Although the magnitude of this estimation bias will depend on the details of each dataset, its qualitative direction is to have been anticipated (Box 2).

Both forms of bias are *not* specific to any of the functional-response models we considered, including the Arditi-Akçakaya model and its parameters. Indeed, both forms of bias may occur for all nonlinear functional-response models – and even for linear models when eaten prey are not replaced (Box 2) – regardless of whether they are consumer dependent or even only resource dependent. Unfortunately, their universality and prevalence have not been recognized in the functional-response literature, nor in the larger literatures on the characterization and estimation of species interaction strengths more generally (Wootton & Emmerson, 2005). Therefore, in what follows, we first discuss the general nature of sample size bias and the difficulty of accounting for it. Then, focusing on the results of the largest of datasets, we outline several inferences with which to guide the future development of functional-response theory and experiments. The issues we highlight emphasize the need for a more explicit consideration of the varied purposes that motivate the empirical study of consumer feeding rates and their functional forms.

Bias is general, and simple methods for correction do not exist or apply

Although the sample size dependent bias of information criteria like AIC_c, and of maximum-likelihood parameter estimators, were recognized for nonlinear models long ago (Box 1 and Box 2), existing methods seeking to adjust for bias have not been adopted in the functional-response literature. The likely reason for this is that simple, off-the-shelf methods do not exist. Bias quantification remains an active

area of statistical research even for linear models (Kosmidis, 2014), with the most general methods, such as bootstrapping and cross-validation (Efron & Tibshirani, 1994; Myung, 2000), requiring sample sizes greater than those used in many functional-response experiments.

The root cause of the difficulty is that bias is dependent not only upon sample size alone, but also on the combination of a given model's functional form and the variation inherent within each given dataset. More specifically, both model-comparison bias (Box 1) and parameter-estimation bias (Box 2) are a function of a focal model's parametric and geometric complexity, how well these are estimated, and for consumer functional responses, how these interact with the feeding rate variation within and among the treatment levels of an experiment. Sources of variation will include experimental design (e.g., how many and how large a range of species abundance levels are employed; Sarnelle & Wilson (2008); Uszko *et al.* (2020)), observer error in estimating species densities and the number of eaten prey (Jost & Arditi, 2000), and the behavioural variation that exists among consumer and prey individuals (Abrams, 2010; Chesson, 1984). Many such sources of variation will be present even in the well-controlled laboratory studies that dominate the literature, and could even be magnified in such studies given the small numbers of individual animals that are typically involved at low-abundance treatment levels (Coblentz, 2020; Novak *et al.*, 2017).

Both forms of bias therefore pertain to functional-response models in general, including those that do not consider consumer dependence. For example, although we focused our assessment on the interference parameter m of the Arditi-Akçakaya model because it is dimensionless and thus more easily compared across studies, exponents like it are expected to be particularly sensitive to estimation bias be-

cause their maximum-likelihood estimators involve logarithmic functions of the data (see Box 2 and the Supplementary Materials S6). Bias is similarly probable and strong among published estimates of body mass scaling exponents and the Hill exponent in the respective contexts of allometric scaling theory and the density-dependent per capita attack rates of Type-III functional responses (e.g., Baudrot *et al.*, 2016; Hossie & Murray, 2016; Rall *et al.*, 2012). Few datasets in our compilation evidenced over-prediction at low prey densities by the models we considered (which could indicate density-dependent attack rates) and too few included information on consumer and prey body masses, hence the extent of the issue in these contexts deserves further investigation.

Importantly, parameter estimation bias will inflict all other nonlinear functional-response parameters as well. This will occur directly via the bias inherent in their own maximum-likelihood estimators (Box 2), and indirectly through their functional relationships and data-dependent covariances with the estimates of other biased parameters. For example, bias is also expected for the attack rate parameter with which the interference parameter is functionally intertwined in the Arditi-Akçakaya model (see also Gutenkunst *et al.*, 2007; Uszko *et al.*, 2020), or for any deterministic functional-response parameter when the assumed likelihood model entails "nuisance" parameters whose estimates are biased (e.g., the standard deviation of the Gaussian likelihood; see Supplementary Materials S5). Even for the simple Holling Type II model, estimates of the attack rate and handling time parameters can exhibit significant covariation (Papanikolaou *et al.*, 2016; Uszko *et al.*, 2020). The effect of covariation on bias will be especially large when feeding rate variation is large and heteroskedastic, or when the range over which species abundances are varied is insufficient to constrain estimates. The lower the sample size, the more

challenging such issues of parameter identifiability become. This issue is similar to the bias caused by presuming the absence of (co)variation among parameters, as can occur when one or more parameters is fixed to a specific value (Hossie & Murray, 2016).

Fitting models to data serves varied purposes

What can be said about consumer dependence from the datasets in our compilation that do have large sample sizes? We draw three main conclusions that are applicable to the functional-response literature more generally.

First, as judged by AIC_c, each of the three most complex models that we considered had statistically clear support for being the sole best-performing model for several datasets, all other models having $\Delta AIC_c > 2$. Therefore, consumer dependence is pervasive but there is no universally best functional form with which to characterize it (Skalski & Gilliam, 2001). Second, the strength of consumer interference as judged by fitting the Arditi-Akçakaya model varied from being very weak to very strong across consumer species and even for different life stages of the same species (e.g., Elliott, 2005). Focusing on the general tendency of m after establishing that consumer dependence is prevalent is therefore asking the wrong question, with future work more fruitfully considering the potential predictors of variation in mutual predator effects (DeLong, 2014; Novak *et al.*, 2017). Together, these two conclusions indicate ample room for more complex models to provide useful biological insight and to statistically outperform the models considered here (Abrams, 2015; Evans *et al.*, 2013; Stouffer & Novak, unpubl. ms.). Rigorous evidence of consumers for which conspecific density dependence is entirely absent would also be useful, with much work remaining to justify the relevance of species-isolating laboratory experiments to

the functional responses exhibited by consumers in nature (Novak *et al.*, 2017; Preston *et al.*, 2018).

These first two conclusions do not, however, imply that more complex functional-response models will necessarily be more useful than simpler models. Many have advocated for simple models on the basis of their analytical tractability, elegance, and utility in management contexts (e.g., Adkison, 2009; Arditi & Ginzburg, 2012; Odenbaugh, 2005). However, the third conclusion we draw from our analyses of large sample size datasets is that multiple models of varying complexity perform near equally well when performance is judged by the high but otherwise arbitrary standard of describing the data with a $MAD < 1\%$ of the top-performing model. Although our assessment did not evaluate the goodness-of-fit of the models in an absolute sense, for a number of datasets these "equally well-describing" models were not among the AIC_c-judged top-performing models.

The ostensible contrast between our third and first two conclusions points to the varied purposes that motivate the fitting of functional-response models to functional-response data. Most functional-response studies have been justified by a generic desire to better describe and understand the functional form of consumer feeding rates to better predict population dynamics. Old and rejuvenated literatures on the varied and overlapping meanings of description, understanding, and prediction in ecology showcase the many ways in which this expressed desire is too simplistic (Doak *et al.*, 2008; Evans *et al.*, 2013; Levins, 1966; Maris *et al.*, 2018; Odenbaugh, 2005; Pennekamp *et al.*, 2019; Shmueli, 2010). Although a synthesis of these terms for ecology is far from complete, pertinent highlights of the literature include evidence that simple, non-mechanistic population models can often better forecast population dynamics than even the models with which the dynamics were simulated in the first place (Perretti *et al.*, 2013); that

parameter estimates need not all be well-constrained to make accurate predictions in complex systems (Gutenkunst *et al.*, 2007); that consumers may often experience a small enough range in prey population sizes that their functional responses are effectively linear under field conditions (Novak, 2010; Preston *et al.*, 2018); and that functional nonlinearities important to describing variation at some spatial, temporal, or biological scales need not be important at other scales (Chesson, 2009). Others have distinguished among various types of prediction – including explanatory, out-of-sample, extrapolatory, transferable, and forecasting forms (e.g., Maris *et al.*, 2018; Odenbaugh, 2005; Shmueli, 2010) – revealing how judgments of model performance in regards to predictive utility are by their nature dependent on choices of prediction type. Alternative motivations and choices similarly underlie vibrant discussions regarding the uses of alternative information criteria in statistical model selection (Myung, 2000; Shmueli, 2010).

In fact, the varied purposes that motivate the fitting of functional-response models to data are also relevant in regards to whether the statistical bias of low sample size studies is to be considered problematic or not. Efforts to avoid model-comparison bias through the use of complexity-penalizing information-theoretic criteria are typically driven by a desire to avoid over-fit models whose within-sample deviations may be low but whose out-of-sample prediction error is high relative to that of less complex models (Myung, 2000). However, parameterizations of a given model using biased estimators can result in lower prediction error than when unbiased estimators are used (Shmueli, 2010). This is because a model's prediction error reflects both its bias and the variation in the data it cannot explain (i.e. a model's accuracy and precision, respectively; Bolker, 2008), among which there can exist a trade-off. For example, even an unbiased estimator can grossly underestimate 99% of the data as long as it overestimates the remaining

data even more egregiously, outliers notwithstanding. In this sense, the parameter estimation bias that is present among low sample size studies need not be considered problematic if the sole motivation is to be descriptive of a predator's feeding behaviour in a within-sample explanatory sense.

Conclusions

Many forms and sources of bias exist in the functional-response literature. These include recently recognized sources related to an experiment's design and venue (Li *et al.*, 2018; Novak *et al.*, 2017; Preston *et al.*, 2018; Uiterwaal *et al.*, 2018; Uiterwaal & DeLong, 2018), as well as persistent model-associated sources (Barraquand, 2014; Hossie & Murray, 2016; Jost & Arditi, 2000; Marshal & Boutin, 1999; McCoy *et al.*, 2012; Novak, 2010; Novak & Wootton, 2010; Pascual & Kareiva, 1996; Rosenbaum & Rall, 2018; Trexler *et al.*, 1988). Our analyses add to this list of issues, exposing the pervasive effect that low replication has had on inferences of consumer dependence, but they are also not immune to them. For example, we did not address issues related to failing to include the "true" model among the models we considered (Burnham & Anderson, 2002). Hence, for example, our estimates of m assuming the Arditi & Akçakaya (1990) model may have been further biased for datasets where Type III-like density dependence was present (Hossie & Murray, 2016).

For overcoming sample size bias there is but one simple solution: future studies must increase replication. We therefore urge authors, reviewers, and editors to hold studies to a higher standard. That said, we offer no rule of thumb for determining what sample size is enough; as discussed above, the replication necessary to effectively reduce bias will depend on the variance-generating processes of the system and the experimental designs and models with which these processes are studied. Continued developments

in the areas of symbolic regression and optimal experimental design should have much insight to offer in reducing the logistical burden that will be required (Martin *et al.*, 2018; Moffat *et al.*, 2020; Okuyama & Bolker, 2012; Zhang *et al.*, 2018). This notwithstanding, the most fundamental advance that is needed is a conceptual one regarding a clarity of purpose. Using statistics to bridge functional-response models with empirical data requires recognizing that data, theory, and models (both deterministic and statistical) each represent necessarily incomplete and sometimes incompatibly-mismatched characterizations of nature. Hence, for some purposes, failure to fit data need not invalidate a model's utility, while for others a model's superior fit need not substantiate it. Clarity regarding the purpose of functional-response studies will aid in determining the map between biological objectives and methodological approach, and is needed for the functional-response concept to reliably serve as the foundation for work across population, community and evolutionary ecology as it is often considered to be.

Acknowledgments

We thank Stella Uiterwaal and Gregor Kalinkat for providing their compiled bibliographies of functional-response studies. For generously providing data, we thank Roger Arditi, Sven Bacher, Shane Blowes, Kevin Chan, Juang-Horng Chong, Will Cresswell, Malcolm Elliot, Gregor Fussmann, Fatemeh Ganjisaffar, Mark Hebblewhite, Tom Hossie, Pavel Kratina, Birgit Lang, Chris Long, Anders Nilsson, Christina Prokopenko, Timothy Pusack, John Reeve, Thierry Spataro, Adrian Stier, John Vucetich, and Will White, as well as all authors who made their data available in online repositories. We further thank Peter Abrams, Roger Arditi, Lev Ginzburg, Tom Hossie, and the Theoretical Breakfast group at OSU for suggestions that improved the manuscript. DBS acknowledges the support of the Marsden Fund Council, the New

Zealand Government (grant 16-UOC-008), and a University of Canterbury Erskine grant.

Author contributions

Both authors contributed equally to this work. MN wrote the first draft.

Code and data availability

Code for all analyses, as well as most datasets, are available at *https://github.com/stoufferlab/general-functional-responses*. These and additional data sets have also been posted to online repositories per agreement with data contributors (see Table S1), or were obtained from repositories to which they had previously been posted by the original authors.

References

Abrams, P. A. (1994). The fallacies of "ratio-dependent" predation. *Ecology*, 75, 1842–1850.

Abrams, P. A. (2010). Implications of flexible foraging for interspecific interactions: lessons from simple models. *Functional Ecology*, 24, 7–17.

Abrams, P. A. (2015). Why ratio dependence is (still) a bad model of predation. *Biological Reviews*, 90, 794–814.

Abrams, P. A. & Ginzburg, L. R. (2000). The nature of predation: prey dependent, ratio dependent or neither? *Trends in Ecology & Evolution*, 15, 337–341.

Adkison, M. D. (2009). Drawbacks of complex models in frequentist and Bayesian approaches to natural-resource management. *Ecological Applications*, 19, 198–205.

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In: *Second International Symposium on Information Theory* (eds. Petrov, B. & Csaki, F.). Akademiai Kiado, Budapest.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 215–222.

Aldebert, C. & Stouffer, D. B. (2018). Community dynamics and sensitivity to model structure: towards a probabilistic view of process-based model predictions. *Journal of The Royal Society Interface*, 15, 20180741.

Arditi, R. (1983). A unified model of the functional response of predators and parasitoids. *The Journal of Animal Ecology*, 52, 293–303.

Arditi, R. & Akçakaya, H. R. (1990). Underestimation of mutual interference of predators. *Oecologia*, 83, 358–361.

Arditi, R. & Ginzburg, L. R. (1989). Coupling in predator-prey dynamics: Ratio-dependence. *Journal of Theoretical Biology*, 139, 311–326.

Arditi, R. & Ginzburg, L. R. (2012). *How species interact: altering the standard view on trophic ecology*. Oxford University Press.

Arditi, R. & Ginzburg, L. R. (2014). Improving communications between theoretical ecologists, mathe-

matical ecologists, and ecological modelers: response to the critique of our book *How species interact*. *Theoretical Ecology*, 7, 21–22.

Barraquand, F. (2014). Functional responses and predator–prey models: a critique of ratio dependence. *Theoretical Ecology*, 7, 3–20.

Baudrot, V., Perasso, A., Fritsch, C., Giraudoux, P. & Raoul, F. (2016). The adaptation of generalist predators' diet in a multi-prey context: insights from new functional responses. *Ecology*, 97, 1832–1841.

Beddington, J. R. (1975). Mutual interference between parasites or predators and its effect on searching efficiency. *The Journal of Animal Ecology*, 44, 331–340.

Blowes, S. A., Pratchett, M. S. & Connolly, S. R. (2017). Aggression, interference, and the functional response of coral-feeding butterflyfishes. *Oecologia*, 184, 675–684.

Bolker, B. (2020). bbmle: Tools for general maximum likelihood estimation. https://cran.r-project.org/package=bbmle.

Bolker, B. M. (2008). *Ecological models and data in R*. Princeton University Press.

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, 71, 791–799.

Box, M. J. (1971). Bias in nonlinear estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 33, 171–190.

Burnham, K. P. & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. 2nd edn. Springer, New York.

Chan, K., Boutin, S., Hossie, T. J., Krebs, C. J., O'Donoghue, M. & Murray, D. L. (2017). Improving the assessment of predator functional responses by considering alternate prey and predator interactions. *Ecology*, 98, 1787–1796.

Chesson, P. (2009). Scale transition theory with special reference to species coexistence in a variable environment. *Journal of Biological Dynamics*, 3, 149–163.

Chesson, P. L. (1984). Variable predators and switching behavior. *Theoretical Population Biology*, 26, 1–26.

Coblentz, K. E. (2020). Relative prey abundance and predator preference predict individual diet variation in prey-switching experiments. *Ecology*, 101, e02911.

Coblentz, K. E. & DeLong, J. P. (2020). Predator-dependent functional responses alter the coexistence and indirect effects among prey that share a predator. *Oikos*.

Cresswell, W. (1998). Variation in the strength of interference competition with resource density in blackbirds, *Turdus merula*. *Oikos*, 152–160.

Crowley, P. H. & Martin, E. K. (1989). Functional responses and interference within and between year classes of a dragonfly population. *Journal of the North American Benthological Society*, 8, 211–221.

DeAngelis, D. L., Goldstein, R. A. & O'Neill, R. V. (1975). A model for trophic interaction. *Ecology*, 56, 881–892.

DeLong, J. & Vasseur, D. (2011). Mutual interference is common and mostly intermediate in magnitude. *BMC Ecology*, 11, 1.

DeLong, J. P. (2014). The body-size dependence of mutual interference. *Biology Letters*, 10, 20140261.

Doak, D. F., Estes, J. A., Halpern, B. S., Jacob, U., Lindberg, D. R., Lovvorn, J., Monson, D. H., Tinker, M. T., Williams, T. M., Wootton, J. T. *et al.* (2008). Understanding and predicting ecological dynamics: are major surprises inevitable. *Ecology*, 89, 952–961.

Efron, B. & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. CRC press.

Elliott, J. M. (2005). Ontogenetic shifts in the functional response and interference interactions of *Rhyacophila dorsalis* larvae (Trichoptera). *Freshwater Biology*, 50, 2021–2033.

Evans, M. R., Grimm, V., Johst, K., Knuuttila, T., de Langhe, R., Lessells, C. M., Merz, M., O'Malley, M. A., Orzack, S. H., Weisberg, M., Wilkinson, D. J., Wolkenhauer, O. & Benton, T. G. (2013). Do simple models lead to generality in ecology? *Trends in Ecology & Evolution*, 28, 578–583.

Fussmann, G. F. & Blasius, B. (2005). Community response to enrichment is highly sensitive to model structure. *Biology Letters*, 1, 9–12.

Girondot, M. (2020). *HelpersMG: Tools for Environmental Analyses, Ecotoxicology and Various R*

Functions. URL https://CRAN.R-project.org/package=HelpersMG. R package version 4.0.

GraphClick (2010). Graph and movie digitizer for Mac OS X. URL http://www.arizona-software.ch/graphclick/.

Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R. & Sethna, J. P. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLOS Computational Biology*, 3, e189.

Hassell, M. P. (1971). Mutual interference between searching insect parasites. *Journal of Animal Ecology*, 40, 473–486.

Hassell, M. P. & Varley, G. C. (1969). New inductive population model for insect parasites and its bearing on biological control. *Nature*, 223, 1133–1137.

Hebblewhite, M. (2013). Consequences of ratio-dependent predation by wolves for elk population dynamics. *Population Ecology*, 55, 511–522.

Holling, C. S. (1959). Some characteristics of simple types of predation and parasitism. *The Canadian Entomologist*, 91, 385–398.

Hossie, T. J. & Murray, D. L. (2016). Spatial arrangement of prey affects the shape of ratio-dependent functional response in strongly antagonistic predators. *Ecology*, 97, 834–841.

Hurvich, C. M. & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76, 297–307

Jeschke, J. M., Kopp, M. & Tollrian, R. (2002). Predator functional responses: Discriminating between handling and digesting prey. *Ecological Monographs*, 72, 95–112.

Jeschke, J. M., Kopp, M. & Tollrian, R. (2004). Consumer-food systems: why type i functional responses are exclusive to filter feeders. *Biological Reviews*, 79, 337–349.

Johnson, S. G. (2020). The NLopt nonlinear-optimization package. http://ab-initio.mit.edu/nlopt.

Jost, C. & Arditi, R. (2000). Identifying predator–prey processes from time-series. *Theoretical Population Biology*, 57, 325–337.

Katz, C. H. (1985). A nonequilibrium marine predator-prey interaction. *Ecology*, 66, 1426–1438.

Keitt, T. (2017). odeintr: C++ ODE solvers compiled on-demand. https://cran.r-project.org/package=odeintr.

Koen-Alonso, M. (2007). A process-oriented approach to the multispecies functional response. In: *From Energetics to Ecosystems: The Dynamics and Structure of Ecological Systems* (eds. Rooney, N., McCann, K. S. & Noakes, D. L. G.). Springer, Dordrecht.

Kosmidis, I. (2014). Bias in parametric estimation: reduction and useful side-effects. *WIREs Computational Statistics*, 6, 185–196.

Lehtonen, J. (2016). The lambert W function in ecological and evolutionary models. *Methods in Ecology and Evolution*, 7, 1110–1118

Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, 54, 421–431.

Li, Y., Rall, B. C. & Kalinkat, G. (2018). Experimental duration and predator satiation levels systematically affect functional response parameters. *Oikos*, 127, 590–598.

Ly, A., Marsman, M., Verhagen, J., Grasman, R. P. & Wagenmakers, E.-J. (2017). A tutorial on Fisher information. *Journal of Mathematical Psychology*, 80, 40–55.

Maris, V., Huneman, P., Coreau, A., Kéfi, S., Pradel, R. & Devictor, V. (2018). Prediction in ecology: promises, obstacles and clarifications. *Oikos*, 127, 171–183.

Marshal, J. P. & Boutin, S. (1999). Power analysis of wolf-moose functional responses. *The Journal of wildlife management*, 396–402.

Martin, B., Munch, S. & Hein, A. (2018). Reverse-engineering ecological theory from data. *Proceedings of the Royal Society B: Biological Sciences*, 285, 20180422.

McCoy, M. W., Stier, A. C. & Osenberg, C. W. (2012). Emergent effects of multiple predators on prey survival: the importance of depletion and the functional response. *Ecology Letters*, 15, 1449–1456.

Moffat, H., Hainy, M., Papanikolaou, N. E. & Drovandi, C. (2020). Sequential experimental design for predator-prey functional response experiments. *Journal of The Royal Society Interface*, 17, 20200156.

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, 44, 190–204.

Myung, J. I., Navarro, D. J. & Pitt, M. A. (2006). Model selection by normalized maximum likelihood. *Journal of Mathematical Psychology*, 50, 167–179.

Novak, M. (2010). Estimating interaction strengths in nature: experimental support for an observational approach. *Ecology*, 91, 2394–2405.

Novak, M., Wolf, C., Coblentz, K. E. & Shepard, I. D. (2017). Quantifying predator dependence in the functional response of generalist predators. *Ecology Letters*, 20, 761–769.

Novak, M. & Wootton, J. T. (2010). Using experimental indices to quantify the strength of species interactions. *Oikos*, 119, 1057–1063.

Odenbaugh, J. (2005). Idealized, inaccurate but successful: A pragmatic approach to evaluating models in theoretical ecology. *Biology and Philosophy*, 20, 231–255.

Okuyama, T. & Bolker, B. M. (2012). Model-based, response-surface approaches to quantifying indirect interactions. In: *Trait-Mediated Indirect Interactions* (eds. Ohgushi, T., Schmitz, O. J. & Holt, R. D.). Cambridge University Press, pp. 186–204.

Osenberg, C. W., Sarnelle, O., Cooper, S. D. & Holt, R. D. (1999). Resolving ecological questions through meta-analysis: goals, metrics, and models. *Ecology*, 80, 1105–1117.

Papanikolaou, N. E., Williams, H., Demiris, N., Preston, S. P., Milonas, P. G. & Kypraios, T. (2016).

Bayesian inference and model choice for Holling's disc equation: a case study on an insect predator-prey system. *Community Ecology*, 17, 71–78.

Pascual, M. A. & Kareiva, P. (1996). Predicting the outcome of competition using experimental data: Maximum likelihood and Bayesian approaches. *Ecology*, 77, 337–349.

Pennekamp, F., Iles, A. C., Garland, J., Brennan, G., Brose, U., Gaedke, U., Jacob, U., Kratina, P., Matthews, B., Munch, S., Novak, M., Palamara, G. M., Rall, B. C., Rosenbaum, B., Tabi, A., Ward, C., Williams, R., Ye, H. & Petchey, O. L. (2019). The intrinsic predictability of ecological time series and its potential to guide forecasting. *Ecological Monographs*, 89, e01359.

Perretti, C. T., Munch, S. B. & Sugihara, G. (2013). Model-free forecasting outperforms the correct mechanistic model for simulated and experimental data. *Proceedings of the National Academy of Sciences*, 110, 5253–5257.

Preston, D. L., Henderson, J. S., Falke, L. P., Segui, L. M., Layden, T. J. & Novak, M. (2018). What drives interaction strengths in complex food webs? A test with feeding rates of a generalist stream predator. *Ecology*, 99, 1591–1601.

Rall, B. C., Brose, U., Hartvig, M., Kalinkat, G., Schwarzmüller, F., Vucic-Pestic, O. & Petchey, O. L. (2012). Universal temperature and body-mass scaling of feeding rates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 2923–2934.

Rogers, D. (1972). Random search and insect population models. *Journal of Animal Ecology*, 41, 369–383.

Rosenbaum, B. & Rall, B. C. (2018). Fitting functional responses: Direct parameter estimation by simulating differential equations. *Methods in Ecology and Evolution*, 9, 2076–2090.

Sarnelle, O. & Wilson, A. E. (2008). Type III functional response in daphnia. *Ecology*, 89, 1723–1732

Schenk, D., Bersier, L.-F. & Bacher, S. (2005). An experimental test of the nature of predation: neither prey-nor ratio-dependent. *Journal of Animal Ecology*, 74, 86–91.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25, 289–310.

Skalski, G. T. & Gilliam, J. F. (2001). Functional responses with predator interference: Viable alternatives to the Holling Type II model. *Ecology*, 82, 3083–3092.

Solomon, M. (1949). The natural control of animal populations. *The Journal of Animal Ecology*, 1–35.

Stier, A. C., Geange, S. W. & Bolker, B. M. (2013). Predator density and competition modify the benefits of group formation in a shoaling reef fish. *Oikos*, 122, 171–178.

Stier, A. C. & White, J. W. (2014). Predator density and the functional responses of coral reef fish. *Coral Reefs*, 33, 235–240.

Stouffer, D. B. & Novak, M. (unpubl. ms.). Hidden layers of density dependence in consumer feeding rates.

Sugiura, N. (1978). Further analysts of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics - Theory and Methods*, 7, 13–26.

Sutherland, W. J. (1983). Aggregation and the 'ideal free' distribution. *Journal of Animal Ecology*, 52, 821–828.

Trexler, J. C., McCulloch, C. E. & Travis, J. (1988). How can the functional reponse best be determined? *Oecologia*, 76, 206–214.

Tyutyunov, Y. & Titova, L. (2020). From Lotka-Volterra to Arditi-Ginzburg: 90 years of evolving trophic functions. *Biology Bulletin Reviews*, 10, 167–185.

Uiterwaal, S. F., Dell, A. I. & DeLong, J. P. (2018). Arena size modulates functional responses via behavioral mechanisms. *Behavioral Ecology*, 30, 483–489.

Uiterwaal, S. F. & DeLong, J. P. (2018). Multiple factors, including arena size, shape the functional responses of ladybird beetles. *Journal of Applied Ecology*, 55, 2429–2438.

Uiterwaal, S. F. & DeLong, J. P. (2020). Functional responses are maximized at intermediate temperatures. *Ecology*, 101, e02975.

Uszko, W., Diehl, S. & Wickman, J. (2020). Fitting functional response surfaces to data: a best practice guide. *Ecosphere*, 11, e03051.

Vucetich, J. A., Peterson, R. O. & Schaefer, C. L. (2002). The effect of prey and predator densities on wolf predation. *Ecology*, 83, 3003–3013.

Wootton, J. T. & Emmerson, M. (2005). Measurement of interaction strength in nature. *Annual Review of Ecology, Evolution, and Systematics*, 36, 419–444.

Zhang, J. F., Papanikolaou, N. E., Kypraios, T. & Drovandi, C. C. (2018). Optimal experimental design for predator–prey functional response experiments. *Journal of The Royal Society Interface*, 15, 20180186.

Figure legends

Figure 1: Model complexity as assessed by the flexibility term of the Fisher Information Approximation applied to each of the functional-response models we considered (Table 1). We assumed an experimental design entailing $N \in \{2, 4, 8, 16, 32\}$ and $P \in \{1, 2, 5\}$ treatment levels with prey replacement, and that the numbers of prey eaten were samples from a Poisson distribution constrained such that no more than an expected 32 prey were eaten in any treatment level in a single unit time period T. Note that flexibility will differ for different experimental designs and may only be compared among models having the same number of parameters; an additional term in the Fisher Information Approximation penalizes models for their number of parameters (see Supplementary Materials S3).

Figure 2: The rank-order performance of functional-response models as judged by (A) AIC_c and (B) *MAD*, with datasets ordered by their sample size. The gray region in (A) demarcates models interpreted as having equivalent support to the top model (i.e. $\Delta AIC_c < 2$). The gray region in (B) demarcates models having a *MAD* value within 1% of that of the top-ranked model. See Table 1 for model abbreviations.

Figure 3: The effect of sample size on the proportion of datasets for which each of the considered functional-response models was ranked as (A) first or (B) second as judged by AIC_c, and as (C) first or (D) second as judged by *MAD*. In each panel, more and more datasets are included in the pool of considered datasets as the maximum considered sample size increases from left to right, starting with all datasets having sample sizes of 20 or less. Tick-marks reflect the distribution of sample sizes greater than 20. Note that a model's rank is not equivalent to the support for that model because lower-ranked models may have equivalent support (i.e. $\Delta AIC_c < 2$). See Table 1 for model abbreviations.

Figure 4: The strength of consumer interference, m, as estimated for each dataset assuming the Arditi-Akçakaya model, with datasets ordered by their sample size and intervals representing one standard error. The large interval for some datasets is due to poor parameter identifiability despite large sample size. See Fig. S1 for an equivalent figure on which co-variate information is overlaid.

Figure 1

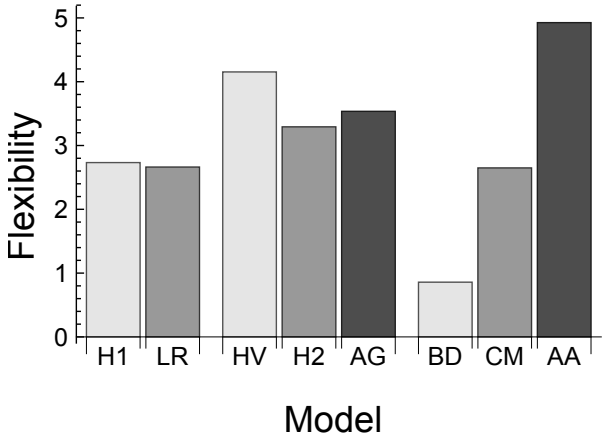


Figure 2

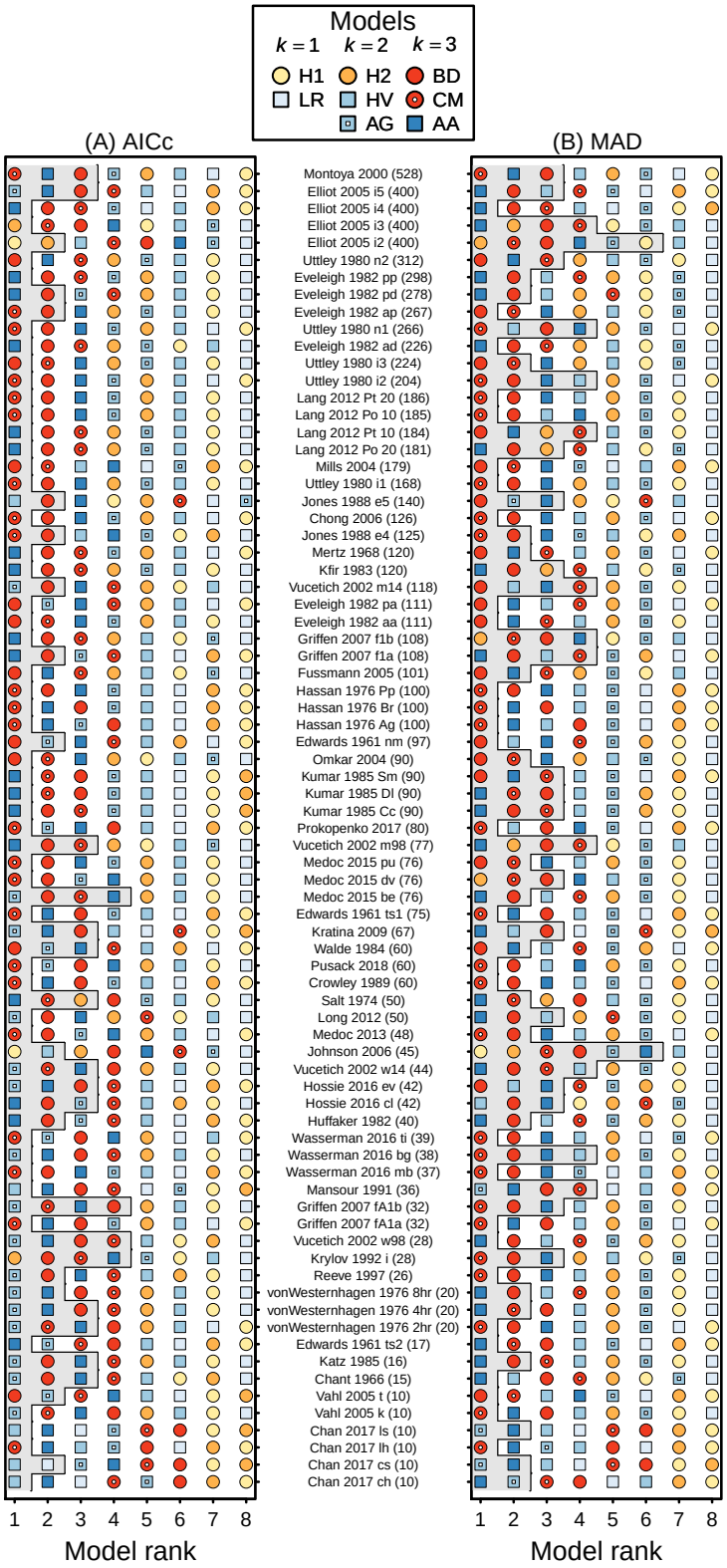

Models
k = 1 k = 2 k = 3
H1 H2 BD
LR HV CM
AG AA
(A) AICc
(B) MAD
Montoya 2000 (528)
Elliot 2005 i5 (400)
Elliot 2005 i4 (400)
Elliot 2005 i3 (400)
Elliot 2005 i2 (400)
Uttley 1980 n2 (312)
Eveleigh 1982 pp (298)
Eveleigh 1982 pd (278)
Eveleigh 1982 ap (267)
Uttley 1980 n1 (266)
Eveleigh 1982 ad (226)
Uttley 1980 i3 (224)
Uttley 1980 i2 (204)
Lang 2012 Pt 20 (186)
Lang 2012 Po 10 (185)
Lang 2012 Pt 10 (184)
Lang 2012 Po 20 (181)
Mills 2004 (179)
Uttley 1980 i1 (168)
Jones 1988 e5 (140)
Chong 2006 (126)
Jones 1988 e4 (125)
Mertz 1968 (120)
Kfir 1983 (120)
Vucetich 2002 m14 (118)
Eveleigh 1982 pa (111)
Eveleigh 1982 aa (111)
Griffen 2007 f1b (108)
Griffen 2007 f1a (108)
Fussmann 2005 (101)
Hassan 1976 Pp (100)
Hassan 1976 Br (100)
Hassan 1976 Ag (100)
Edwards 1961 nm (97)
Omkar 2004 (90)
Kumar 1985 Sm (90)
Kumar 1985 Dl (90)
Kumar 1985 Cc (90)
Prokopenko 2017 (80)
Vucetich 2002 m98 (77)
Medoc 2015 pu (76)
Medoc 2015 dv (76)
Medoc 2015 be (76)
Edwards 1961 ts1 (75)
Kratina 2009 (67)
Walde 1984 (60)
Pusack 2018 (60)
Crowley 1989 (60)
Salt 1974 (50)
Long 2012 (50)
Medoc 2013 (48)
Johnson 2006 (45)
Vucetich 2002 w14 (44)
Hossie 2016 ev (42)
Hossie 2016 cl (42)
Huffaker 1982 (40)
Wasserman 2016 ti (39)
Wasserman 2016 bg (38)
Wasserman 2016 mb (37)
Mansour 1991 (36)
Griffen 2007 fA1b (32)
Griffen 2007 fA1a (32)
Vucetich 2002 w98 (28)
Krylov 1992 i (28)
Reeve 1997 (26)
vonWesternhagen 1976 8hr (20)
vonWesternhagen 1976 4hr (20)
vonWesternhagen 1976 2hr (20)
Edwards 1961 ts2 (17)
Katz 1985 (16)
Chant 1966 (15)
Vahl 2005 t (10)
Vahl 2005 k (10)
Chan 2017 ls (10)
Chan 2017 lh (10)
Chan 2017 cs (10)
Chan 2017 ch (10)
1 2 3 4 5 6 7 8
Model rank
1 2 3 4 5 6 7 8
Model rank

Figure 3

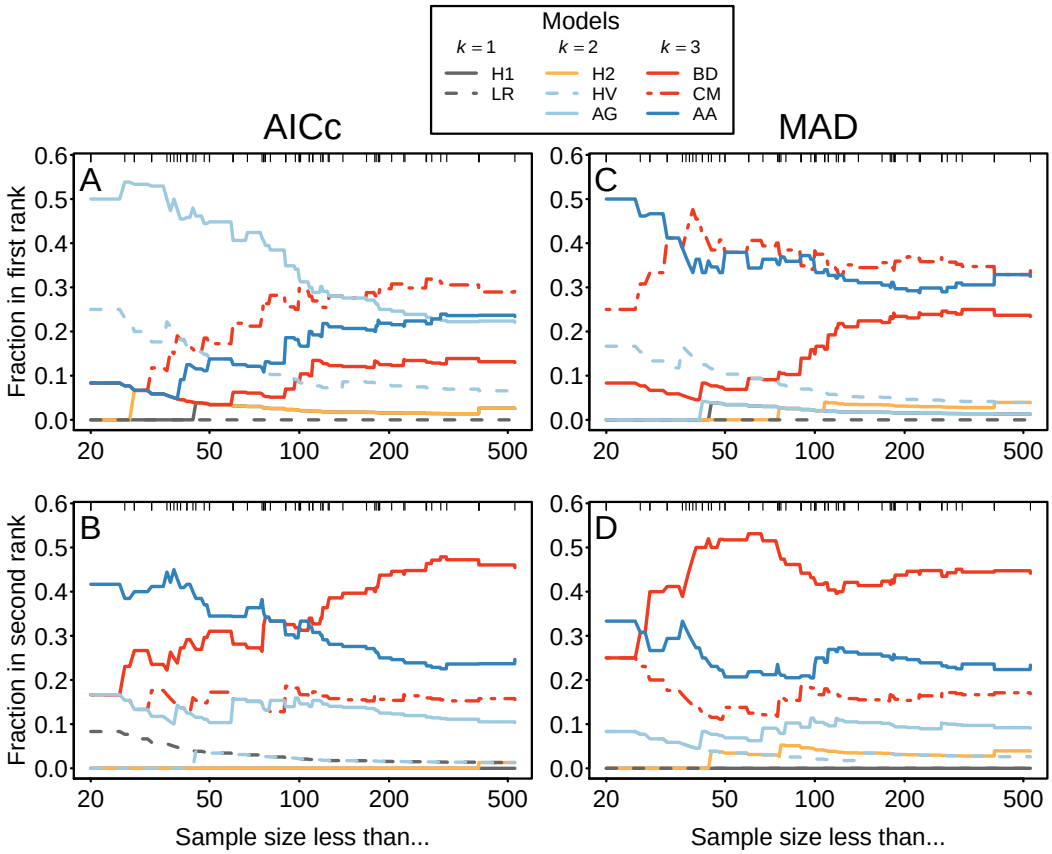
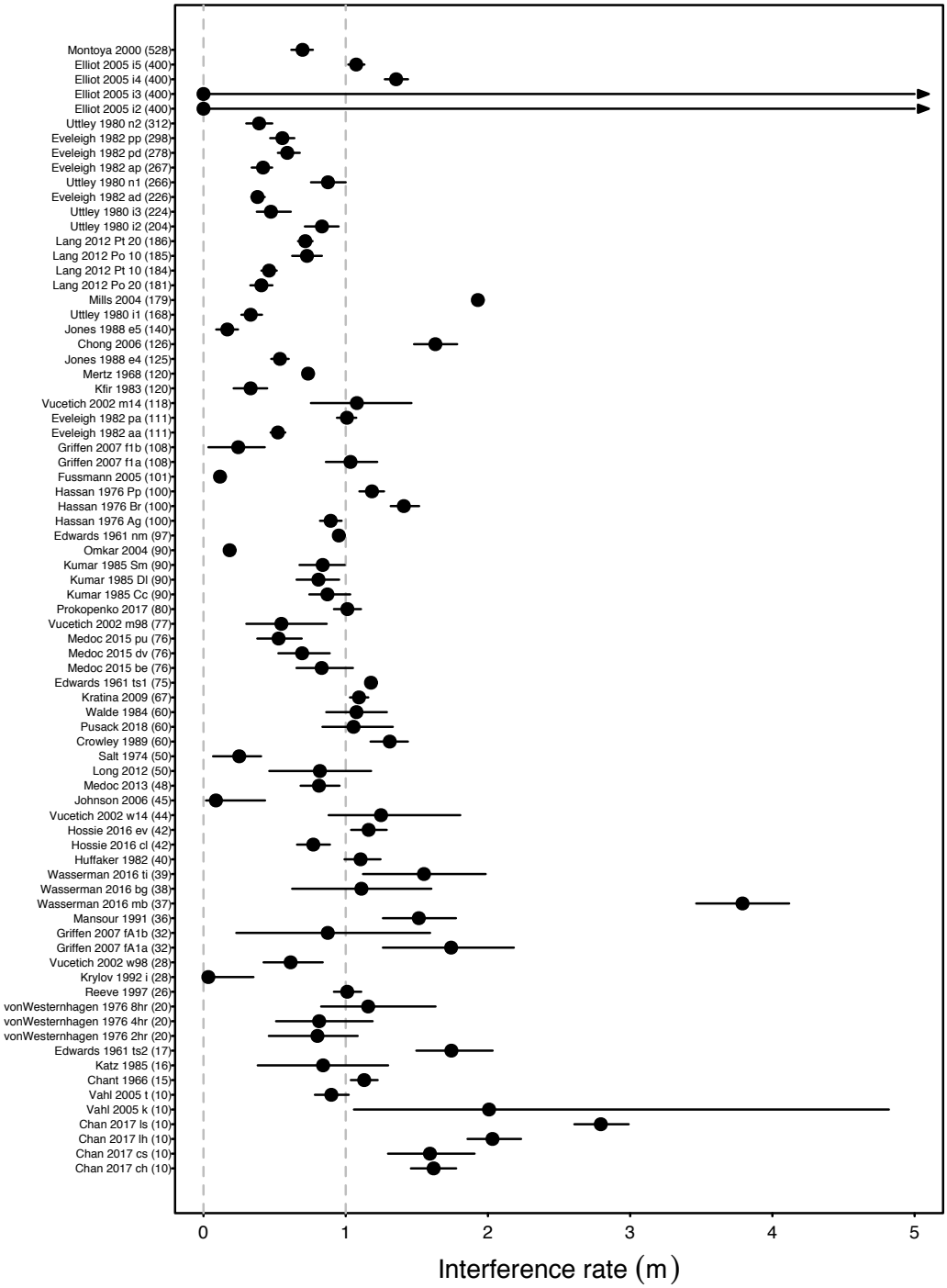

Models
k = 1
k = 2
k = 3
H1
LR
H2
HV
AG
BD
CM
AA
AICc
MAD
A
B
C
D
Fraction in first rank
Fraction in second rank
0.0
0.1
0.2
0.3
0.4
0.5
0.6
20
50
100
200
500
Sample size less than...
Sample size less than...

Figure 4


Montoya 2000 (528)
Elliot 2005 i5 (400)
Elliot 2005 i4 (400)
Elliot 2005 i3 (400)
Elliot 2005 i2 (400)
Uttley 1980 n2 (312)
Eveleigh 1982 pp (298)
Eveleigh 1982 pd (278)
Eveleigh 1982 ap (267)
Uttley 1980 n1 (266)
Eveleigh 1982 ad (226)
Uttley 1980 i3 (224)
Uttley 1980 i2 (204)
Lang 2012 Pt 20 (186)
Lang 2012 Po 10 (185)
Lang 2012 Pt 10 (184)
Lang 2012 Po 20 (181)
Mills 2004 (179)
Uttley 1980 i1 (168)
Jones 1988 e5 (140)
Chong 2006 (126)
Jones 1988 e4 (125)
Mertz 1968 (120)
Kfir 1983 (120)
Vucetich 2002 m14 (118)
Eveleigh 1982 pa (111)
Eveleigh 1982 aa (111)
Griffen 2007 f1b (108)
Griffen 2007 f1a (108)
Fussmann 2005 (101)
Hassan 1976 Pp (100)
Hassan 1976 Br (100)
Hassan 1976 Ag (100)
Edwards 1961 nm (97)
Omkar 2004 (90)
Kumar 1985 Sm (90)
Kumar 1985 Dl (90)
Kumar 1985 Cc (90)
Prokopenko 2017 (80)
Vucetich 2002 m98 (77)
Medoc 2015 pu (76)
Medoc 2015 dv (76)
Medoc 2015 be (76)
Edwards 1961 ts1 (75)
Kratina 2009 (67)
Walde 1984 (60)
Pusack 2018 (60)
Crowley 1989 (60)
Salt 1974 (50)
Long 2012 (50)
Medoc 2013 (48)
Johnson 2006 (45)
Vucetich 2002 w14 (44)
Hossie 2016 ev (42)
Hossie 2016 cl (42)
Huffaker 1982 (40)
Wasserman 2016 ti (39)
Wasserman 2016 bg (38)
Wasserman 2016 mb (37)
Mansour 1991 (36)
Griffen 2007 fA1b (32)
Griffen 2007 fA1a (32)
Vucetich 2002 w98 (28)
Krylov 1992 i (28)
Reeve 1997 (26)
vonWesternhagen 1976 8hr (20)
vonWesternhagen 1976 4hr (20)
vonWesternhagen 1976 2hr (20)
Edwards 1961 ts2 (17)
Katz 1985 (16)
Chant 1966 (15)
Vahl 2005 t (10)
Vahl 2005 k (10)
Chan 2017 ls (10)
Chan 2017 lh (10)
Chan 2017 cs (10)
Chan 2017 ch (10)
0
1
2
3
4
5
Interference rate (m)