

Contributions of body size, habitat and taxonomy to predictions of temperate Australian fish diets

German A. Soler^{1,*}, Russell J. Thomson¹, Rick D. Stuart-Smith¹,
Anthony D. M. Smith², Graham J. Edgar¹

¹Institute for Marine and Antarctic Studies, University of Tasmania, Private Bag 49, Hobart, TAS 7001, Australia

²CSIRO Oceans and Atmosphere Flagship, CSIRO, GPO Box 1538, Hobart, TAS 7001, Australia

ABSTRACT: Using k -nearest neighbour procedures to predict prey type and linear models to predict mean prey size, we developed a 2-step dietary model based on the stomach contents of fish of known species, size and location from Western Port, Victoria (Australia). The model, nicknamed 'Consume', was used to assess the relative extent to which fish diet varied with body size, species identity, season, and location. Both prey type (mean overlap between predicted and actual prey types = 77 %) and mean prey size (r^2 between predicted and observed mean prey size = 93 %) were predicted with reasonable accuracy when species identity and length of consumer fish were known. The most important predictor for prey type was the size of the individual consumer, while the most important predictor for mean prey size was the consumer's taxonomic identity. Predictors were individually removed from both k -nearest neighbour and linear models to assess their relative contributions to the model. Little loss of accuracy (1 %) was evident when family rather than species identity was used for both prey type and mean prey size. Environmental information associated with the time and location of fish sampling (habitat, site and season) contributed only marginally to predictions of prey type. Use of the Consume model will allow for an improved understanding of community-level trophic pathways through the integration of prey type and size predictions for consumer fishes.

KEY WORDS: Predator–prey relations · Diet prediction · Australia · Consume model · Tropho-dynamics · Allometry

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

Studies that predict and provide insight about trophic pathways represent important tools for ecology (Edgar & Shaw 1995c). Furthermore, studies that consider the size of consumers as well as that of their prey can increase our understanding of food-web structure and function, as well as providing new approaches in the integration of community ecology with energetic and metabolic theory (Brown et al. 2004, Belgrano 2005, Trebilco et al. 2013). Using a comprehensive dataset for consumers and their prey collected from Western Port, Australia (Edgar & Shaw 1995b), we present a new predictive diet

model based on the combination of k -nearest neighbour (kNN) procedures to predict prey type, and a linear model to predict prey size. The kNN algorithm is an instance-based learning classification technique that categorizes new cases using similarity measures between the new case and a training dataset. Here, the new cases are fish, and the similarity measure is calculated from the predictor variables. The kNN method has an advantage over previous methods for predicting fish diets in that it is non-parametric and relatively simple, and thus makes fewer assumptions about the data.

The classical approach in community ecology considers species as the central unit within an ecosys-

tem, with similar roles amongst individuals. Nevertheless, individuals of many species experience profound changes in diet throughout their life histories, and their body size and that of their prey can both increase several orders of magnitude (Cushing 1975, Rudolf & Rasmussen 2013). Moreover, per-unit-biomass consumption rates and population densities decrease with animal size. Hence, higher metabolic rates and higher per-biomass consumption rates are typical of small individuals (Rudolf & Rasmussen 2013).

The trophic role of species in aquatic ecosystems should therefore be regarded as flexible, changing with body size—a fundamental reason to develop size-based analyses of food webs (Dickie et al. 1987, Jennings & Warr 2003, Blanchard et al. 2011, Zhang et al. 2014). Furthermore, studies from the North Sea and the Western Arabian Sea found that while body size was only a weak predictor of trophic position within species, there was a strong community-scale relationship between trophic position and body size (Jennings et al. 2001, Al-Habsi et al. 2008). Amongst the few studies that have generated predictive models for fish diets, Link (2004) developed a rank proportion algorithm (RPA) model to predict prey preference from first principles of predation. Link's model incorporated ambient prey concentrations, prey selectivity, consumption rates, predator biomass and prey biomass to predict diet composition. This RPA model ranked each component of the interaction between predator and prey as an individual event. Using this RPA, the preferences for a particular prey type can be assessed for a given predator (Link 2004). In a subsequent study, (Pinnegar et al. 2014) applied the RPA model to predict the effect of an invasive species on a Corsican inshore food-web using predicted diets as an input to the modelling framework, Ecopath.

Another approach to modelling diet consumption used prey concentration (as a proxy of encounter rate) and prey size to predict the diet of planktivorous bluegill (Werner & Hall 1974, O'Brien et al. 1976). Wright & O'Brien (1984) studied feeding selectivity of planktivorous white crappie by assuming that the predation process could be subdivided into different steps—prey location, pursuit, attack and retention. Each of these steps was assigned a probability, and the predicted prey consumption was calculated as the sum of the individual probabilities. Other studies have assessed predator–prey size/mass ratios (Hahm & Langton 1984, Scharf et al. 2000), with prey abundance and type as important factors in determining optimum foraging patterns. A recent study investi-

gated the morphological and behavioural traits of prey that make individuals more vulnerable to predation by lionfish (Green & Côté 2014). Based on these traits, the authors predicted which prey was more susceptible to predation. They cross-validated their results based on *in situ* observations and stomach contents of captured lionfish, obtaining a high correlation. Stable isotope analyses have also been used as a tool to predict predator–prey ratios and trophic transfer efficiencies (Jennings et al. 2002).

Early quantitative studies on dietary pathways linking benthic invertebrates to fish communities in Western Port, Victoria, focused on seagrass and unvegetated habitats (Edgar & Shaw 1995a,b,c). The fish community predominantly consisted of small-sized (<10 g) animals, and was significantly more abundant and productive in seagrass than in unvegetated habitats. Crustacean production was highly correlated with fish production, and crustaceans >1 mm were the most important dietary component (Edgar & Shaw 1995b). On average, prey length was 7.5% of consumer fish length. Furthermore, major ontogenetic changes in diet were evident in Western Port, with the diets of adult and juvenile fish of the same species differing as much as that between different species.

In this study, we used the Western Port dataset, consisting of individual items counted and measured from fish stomach contents (Edgar & Shaw 1995a, b,c), to address the following questions:

- Through application of kNN theory, how accurately can the prey types of a fish of known species and size be predicted?
- Using the kNN model, what is the loss in accuracy of prey type prediction if species-level identity is unknown (e.g. only genus or family are known) or if size information for the consumer fish is lacking?
- Can the mean size of each prey type consumed by a fish of known size and species be estimated accurately using a linear model?
- What is the loss in accuracy in the linear model for mean prey size predictions without complete taxonomic or size information?
- Can the accuracy of the kNN predictions for prey type and the linear model predictions for mean prey size be improved by including environmental factors such as site, habitat and season?

MATERIALS AND METHODS

This study was based on data collected by Edgar & Shaw (1995a,b,c) from Western Port, Victoria, in

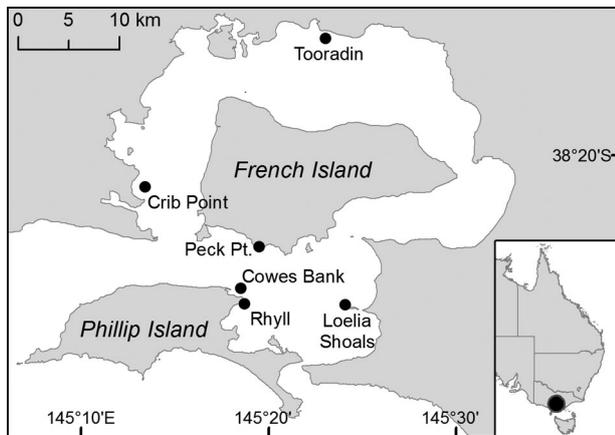


Fig. 1. Western Port, Australia, showing locations of the sampling sites

southeastern Australia. The 2974 individual fishes used in that study were collected using seine and gill nets from 3 different habitat types (seagrass, unvegetated intertidal flats, unvegetated channel) at 6 locations (Fig. 1). Of the 6 locations sampled, 3 (Peck Point, Rhyll and Tooradin) contained all 3 habitat types. Cowes Bank and Crib Point were composed of 2 habitat types (seagrass, unvegetated intertidal flats) while Loelia Shoals was comprised of one (unvegetated channel). In this paper, the term 'site' refers to these 6 locations.

Fish caught at these sites were identified to species level, weighed, and their length to caudal fork measured. The calculations used to estimate fish densities (m^{-2}) were described by Edgar & Shaw (1995a). Individual prey items in the stomachs of captured fishes were identified and their body lengths measured using a microscope graticule or Vernier calipers, and binned into 19 different log size-classes ranging from 0.125 to 64 mm (Edgar & Shaw 1995b). A percentage of total gut contents was allocated in the case of algae, sponges, bryozoans, ascidians and hydroids, or in the case of prey that were unidentifiable because of advanced states of digestion. The percentage of these prey items was estimated as a proportion of the biomass of prey in the stomach of the captured fish (for details see Edgar & Shaw 1995b). Prey items were grouped by taxon into 11 major prey types: (1) algae, (2) sponges, (3) crustacean epifauna, (4) crustacean infauna, (5) mol-

lusc epifauna, (6) mollusc infauna, (7) polychaete epifauna, (8) polychaete infauna, (9) other epifauna, (10) other infauna and (11) fish. Using this information, a matrix was constructed showing environmental characteristics (habitat, site and month), taxonomic information (family, genus and species), morphometric measurements (length and wet weight of fish) and prey information (prey items, taxon and type).

The Consume model was constructed in 2 steps using R (R Development Core Team 2014) (see Supplements 1, 2 and 3 at www.int-res.com/articles/suppl/m545p239_supp/), with results combined in an output matrix showing prey type, prey type percentage and mean prey size predictions within each prey type for each individual fish (Fig. 2). The first model step used the premise of the kNN algorithm (Barber 2011, Conway & White 2012) to predict prey type in percentages for each fish. The second step used linear models to predict the size of prey for each particular prey type consumed by a fish. The training dataset used for this algorithm was described by Edgar & Shaw (1995b). As predictors for the prey type of a particular fish and size of its prey, we used environmental, taxonomic, morphometric and prey information. We used the same predictors for prey size as for prey type, and included the estimates of prey type as another predictor. We opted for this approach because it brought together information associated with prey type, established with input from a dietary database, into the predictions of prey size. Furthermore,

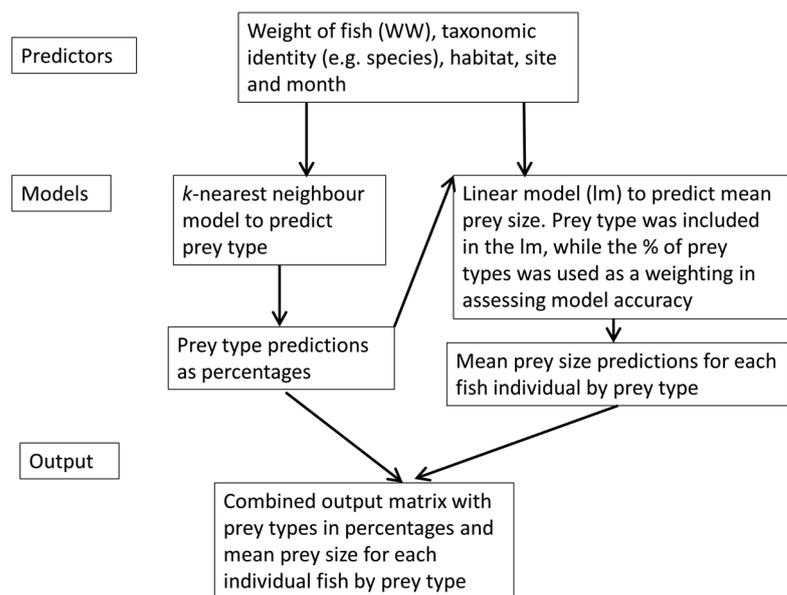


Fig. 2. Predictors used in the 'Consume' 2-step model, and how they combined to produce the output matrix with prey type, prey type percentages and mean prey size for a consumer fish

prey type proved to be an important predictor for the prey size model, as observed by the variation explained in the sum of squares calculations. We were unable to develop a model that simultaneously estimated prey type, prey type percentage and prey size; however, the 2-step approach avoided undue model complexity and achieved good results in the comparisons between predicted and observed values.

The algorithm for the first step in the Consume model used a combination of the importance or weighting of the predictors as well as an optimum number of neighbouring fish, k , where k refers to the most similar fish in size and taxonomy used as a predictor. The optimum weighting of the predictors and k were chosen using a leave-1-out cross validation technique. This was done by predicting the prey type for each fish based on the rest of the dataset for various weights and values of k . The weights that gave the greatest accuracy of the predictions were subsequently used.

Nearest neighbours were chosen based on the pairs of fish with the smallest dissimilarity, where dissimilarity was calculated as the weighted sum of the absolute difference between the values of predictors for each pair of fish,

$$D_{ij} = \sum_{k=1}^{n_p} w_k \text{abs}(\beta_{ik} - \beta_{jk}) \quad (1)$$

and D_{ij} is the dissimilarity between fish i and j , $i \neq j$ and $k = 1, \dots, n_p$, with n_p being the total number of predictor variables; w_k is the given weight or importance of predictor k ; $\text{abs}(\beta_{ik} - \beta_{jk})$ is the absolute difference between fish i and fish j , for predictor variable k . When the predictor variable is a factor, this equals 0 if the predictors are the same, or 1 otherwise. Wet weight of the consumer fish was \log_e transformed before calculating the absolute difference.

A number of dissimilarity cut-offs were trialled through several runs of the model, with the accuracy of predictions used to determine the most appropriate cut-off. Decreasing the cut-off to a maximum dissimilarity between individuals of 0.01 maximised model accuracy (75.8%). This reduced the number of individuals used to predict prey type to an average of 1.1 (see Table S1 in Supplement 4 at www.int-res.com/articles/suppl/m545p239_supp/). If all dissimilarities exceeded the cut-off, the individual with the minimum dissimilarity was used to predict prey type.

The overlap in percentage between predicted and observed values was used to assess the accuracy of the model for prey type predictions. For example, if a

fish was predicted to have a diet content of 20% crustacean epifauna and 80% mollusc epifauna, and the observed data revealed 10% crustacean epifauna, 10% algae, 70% mollusc epifauna and 10% mollusc infauna, then the model was estimated to have an accuracy of 80%. In this case the overlap in the proportion that was correct was 80% (10% crustacean epifauna + 70% mollusc epifauna). In computing the accuracy of the model, the predicted diet content of the fish was never based on a model that included the fish in question (i.e. $i \neq j$ in Eq. 1).

The algorithm for the second step of the Consume model, to predict the size of the prey consumed by each fish, was based on the linear model described below:

$$y_i = \mu + \beta_1 \text{Species}_i + \beta_2 \text{WW}_i + \beta_3 \text{Prey type}_i + \beta_4 \text{Habitat}_i + \beta_5 \text{Site}_i + \beta_6 \text{Month}_i + \varepsilon_i \quad (2)$$

where y_i = the natural log of the mean prey size prediction for the i^{th} fish, given the effects of species, wet weight (WW), prey type, habitat, site and month; μ = overall mean; ε_i = residual error. A natural log transformation made the data more normally distributed. The mean prey size within each prey type for a specific fish was predicted using these factors. We also tested a linear mixed effect model with random nested effects of family/genus/species. However, the results of the linear model and the linear mixed effects model were almost the same in predicting mean prey size. We chose the linear model because it allowed us to more easily test the decrease in accuracy when removing each predictor, one at a time. We tried different combinations of omitted predictors to assess their individual effects on the accuracy of the linear model. We measured the accuracy of the linear model as the square of the Pearson's weighted correlation between the predicted and observed values, without including the information for the fish the model was predicting (i.e. leave-1-out cross validation). The weighting of the correlation was based on the predicted percentages of prey type.

We also developed a third model to predict the standard deviation of mean prey size. The importance of this measure was to better understand the circumstances that affect whether individual fish focus on prey within a narrow size band, or are largely unselective, capturing prey in a variety of sizes. This third model used the same predictors as the mean prey size predictions, and included the mean prey size estimates as another predictor. We used a generalized linear model (described below) to predict the standard deviation.

$$y_i = \mu + \beta_1 \text{Species}_i + \beta_2 \text{WW}_i + \beta_3 \text{Prey type}_i \\ + \beta_4 \text{Habitat}_i + \beta_5 \text{Site}_i + \beta_6 \text{Month}_i \\ + \beta_7 \text{Mean Prey size}_i + \epsilon_i \quad (3)$$

where y_i = square root of the standard deviation of the predictions of the \log_e mean prey size for the i^{th} fish. $\beta_{1,2,3,4,5,6}$ are the same predictors as described above for mean prey size predictions; β_7 is the mean prey size prediction used in the model. In this way, our model predicted the standard deviation of the predicted mean prey size by testing if these predictions were comparable to the standard deviations of the observed prey size. We used a square root transformation to make the data more normally distributed. A similar process to that used in the second model was followed for the third model; we tested the effect on the weighted correlation between the observed and predicted standard deviations by removing the different predictors.

In order to standardize the contribution of different species to the model predictions, we examined 2 aspects of data selection. First, to account for differences in the number of individuals of each species, we used a random sample of 12 ind. species⁻¹ in a wide range of sizes. Sampling was achieved with a random number selector based on each individual fish, which were numbered beforehand. Species were also dropped if the inter-quartile range of the \log_e of the wet weight was <0.25. Of the 79 species in the full dataset, 37 were dropped using this criterion. Basically, this excluded species with only a narrow sampled size range, and included species with a wide range in sizes. This also excluded species for which less than 12 individual fish were sampled. Trials

indicated that the greatest accuracy for prey type and mean prey size was achieved when 12 ind. species⁻¹ were used. Second, we tested if using different numbers of fish for each species greatly affected prey type predictions, and found that changes were negligible when at least 5 fish were considered (Table S2 in Supplement 4 at www.int-res.com/articles/suppl/m545p239_supp/).

RESULTS

The kNN model algorithm predicted the prey types of consumer fish with an average accuracy of 76% when using the whole dataset (Table 1). We standardized the contribution of each species by randomly selecting only 12 ind. species⁻¹ in a wide range of sizes. By doing so, we only used 42 species and a total of 504 fish from the original dataset that comprised 79 species and 2974 fish (Table S2 in Supplement 4). With this subset of the data, our prey type predictions increased marginally to $77 \pm 1\%$. This variation was due to the fact that the algorithms randomly chose a different set of 12 fish per species in each run of the model.

We assessed loss in accuracy in prey type predictions by step-wise removal of predictors from the model (Table 1), and determined that the most important predictor for prey type was wet weight of the consumer fish. By removing this predictor, the model lost over 16% of its accuracy. Taxonomic predictors proved to be of little relative importance, with a loss in accuracy of only 3%. The least important predictors were habitat, site and month, which decreased

Table 1. Percentage correct between predicted and observed prey types for the full dataset, and for a random subset of 12 individuals from each species encompassing a wide range of sizes for fish collected from Western Port, Australia. Also shown is the percentage correct of the predicted prey for all predictors and the subsequent loss in accuracy when removing factors from the model. WW: \log_e wet weight of consumer fish. These predictions of prey type were based on Step 1 of the 'Consume' model using the k -nearest neighbour algorithm. mo = month

	Predictors	% correct with all fish included	% correct using a subset of 12 ind. species ⁻¹
With all predictors	Species, WW, habitat, site, mo	76	77
Without species	Genus, WW, habitat, site, mo	76	77
Without species and genus	Family, WW, habitat, site, mo	75	76
Without species, genus and family	WW, habitat, site, mo	70	74
Without WW	Species, habitat, site, mo	62	61
Without habitat	Species, WW, site, mo	74	76
Without site	Species, WW, habitat, mo	75	76
Without month	Species, WW, habitat, site	74	76
Without habitat, site, month	Species, WW	72	75
Without all predictors but site	Site	23	18
Without all predictors but habitat	Habitat	23	18

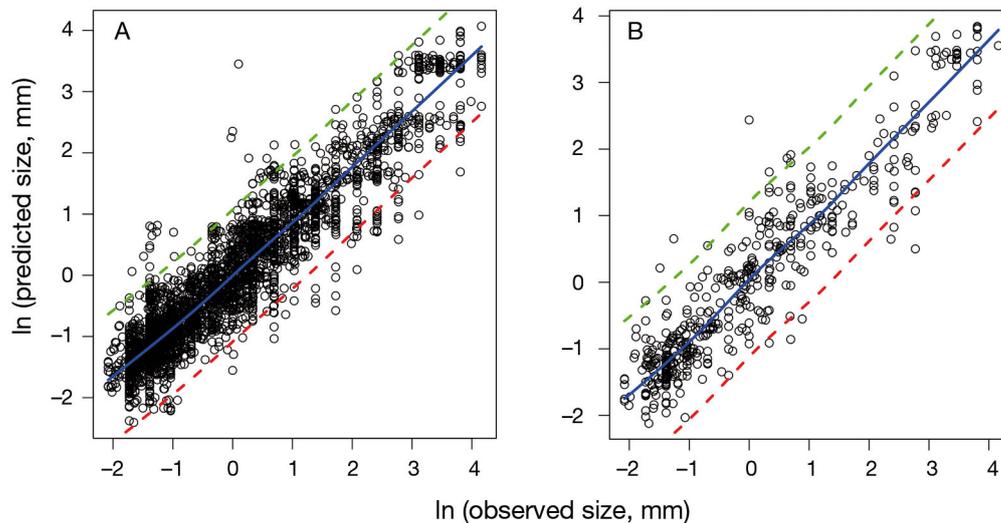


Fig. 3. Relationship between predicted versus observed mean prey size for consumer fish, using a linear model with all predictors (species, genus, family, wet weight [WW], prey type, site, habitat and month; weighted correlation), with (A) all fish included ($R^2 = 0.90$, 79 species) and (B) a subset of 12 individuals from each species ($R^2 = 0.93 \pm 0.01$, 42 species, where \pm indicates SE of estimates generated by different sets of individuals randomly selected to represent each species). Green and red lines: upper and lower 95% prediction intervals, respectively; blue line: locally weighted scatterplot smoothing curve (LOWESS)

the accuracy of the model by 1% when removed. If all predictors were removed except site or habitat, the accuracy of the model dropped to 18%.

The accuracy of the linear model for estimating prey size, measured as the square of the weighted correlation between predicted and observed mean prey size, was $R^2 = 0.90$ (Fig. 3A) when all predictors and all fish in the dataset were used. With the subset of 12 ind. species⁻¹, this weighted correlation increased to $R^2 = 0.93 \pm 0.01$ (Fig. 3B). This model used the prey type prediction from Step 1 as part of the predictors and used the predictions for prey type for weighting of the correlation. The most important predictors were taxonomic group, size of the consumer fish and prey type (Tables S3 & S4 in Supplement 4).

Removing the species and genus information of the consumer had little effect on the accuracy of the model, provided that the family designation was retained along with the other predictors ($R^2 = 0.92 \pm 0.01$; Table 2). Consumer taxonomic information proved to be the most important predictor in mean prey size prediction (Table 2). In order to determine the relative importance of each factor, we removed the different predictors one at the time and estimated the loss in accuracy of the model with each removal (Table 2). When WW was excluded, the correlation dropped to $R^2 = 0.86$ (Table 2). Consequently, the loss in accuracy when WW was not used was ~ 0.07 . When taxonomic information was removed and all other

predictors included, the correlation dropped to $R^2 = 0.80$. This corresponded to a loss in accuracy for mean prey size predictions of 0.13 when not including taxonomic information. When using the whole dataset, the loss in accuracy when not using taxonomic information was 0.16. This indicates that taxonomic information is the most important predictor of mean prey size. Conversely, the loss in accuracy was extremely small (<0.005) when habitat, site or month were removed (Table 2). Moreover, when these 3 predictors were all removed from the model, the accuracy was only affected by 0.01 (Table 2).

We also tested if some species affected our predictions more than others. The diet of the most abundant species, the yellow-eyed mullet *Aldrichetta forsteri*, had previously been found to consist of anomalously small prey (Edgar & Shaw 1995b). By including the yellow-eyed mullet in the model, the weighted correlation between WW of the consumer fish and the observed prey size was $R^2 = 0.54$ for all fish species, and $R^2 = 0.68 \pm 0.01$ for the subset of 12 ind. species⁻¹ (Fig. 4A, 4B). When the yellow-eyed mullet was removed, the weighted correlation increased to $R^2 = 0.71$ (all data included) and $R^2 = 0.70 \pm 0.01$ (for the subset of 12 ind. species⁻¹) (Fig. 4C). Most of the yellow-eyed mullet were around 270 mm (approximate exponential of 5.6), which is where the largest variation of prey size was observed (Fig. 4A). Other species could also potentially influence the variation in prey size, but were not tested independently.

Table 2. Correlation between observed and predicted mean prey size for the full dataset and for a random subset of 12 individuals from each species encompassing a wide range of sizes for fish collected in Western Port, Australia. Also shown are the correlations for all predictors available and the subsequent loss in accuracy when predictors were removed. Weighting of the correlation was based on the predicted percentages of prey type. WW = \log_e wet weight of consumer fish. The difference in the prediction intervals indicates the fit with the different combinations of predictors. mo = month

	Predictors	R ² with all fish included	R ² with subset of 12 ind. species ⁻¹	Mean difference between upper and lower prediction interval
With all predictors	Species, WW, prey type, habitat, site, mo	0.90	0.93	2.02
Without species	Genus, WW, prey type, habitat, site, mo	0.90	0.93	2.01
Without species and genus	Family, WW, prey type, habitat, site, mo	0.89	0.92	2.09
Without species, genus and family	WW, prey type, habitat, site, mo	0.74	0.80	2.83
Without WW	Species, prey type, habitat, site, mo	0.85	0.86	2.59
Without habitat	Species, WW, prey type, site, mo	0.90	0.93	2.22
Without site	Species, WW, prey type, habitat, mo	0.90	0.93	2.25
Without month	Species, WW, prey type, habitat, site	0.90	0.93	2.20
Without habitat, site, month	Species, WW, prey type	0.89	0.93	2.28
Without prey type	Species, WW, habitat, site, mo	0.89	0.92	2.31
Without all predictors but site	Site	0.11	0.10	5.52
Without all predictors but habitat	Habitat	0.06	0.05	5.61

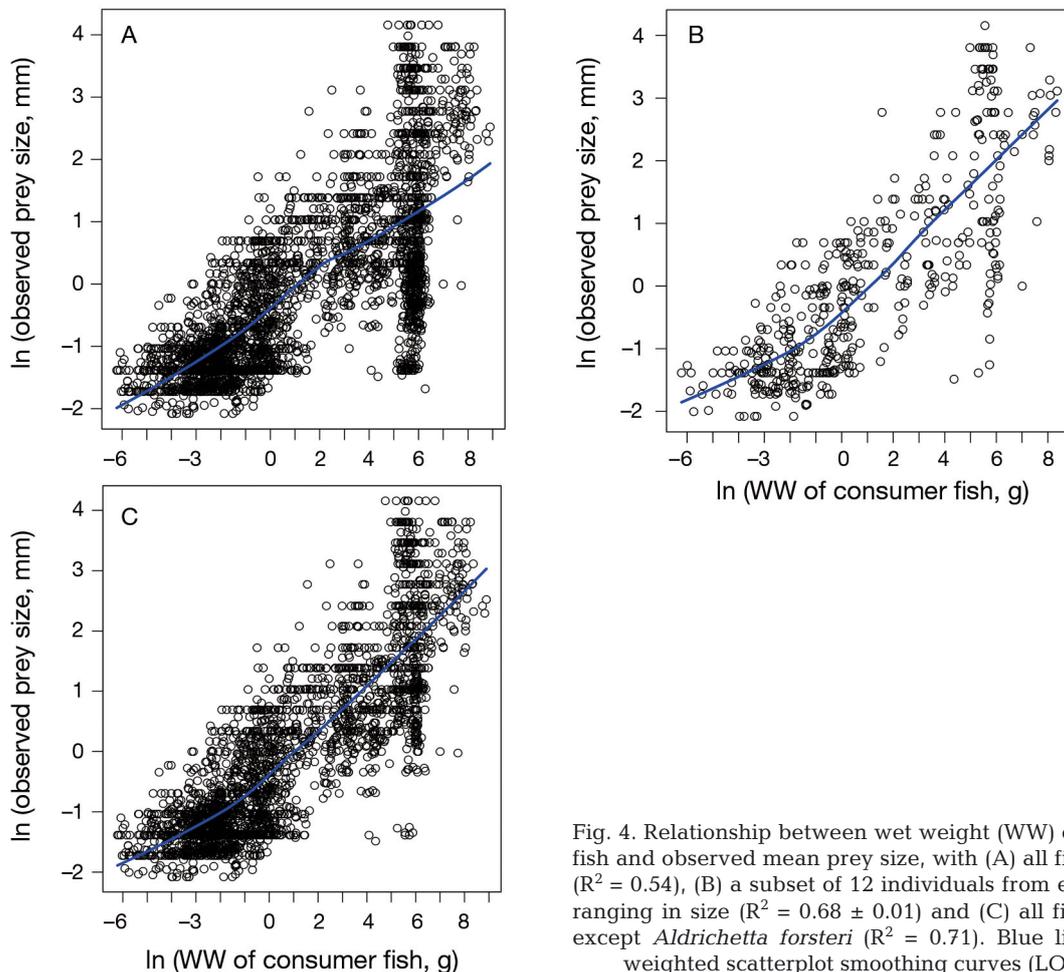


Fig. 4. Relationship between wet weight (WW) of consumer fish and observed mean prey size, with (A) all fish included ($R^2 = 0.54$), (B) a subset of 12 individuals from each species ranging in size ($R^2 = 0.68 \pm 0.01$) and (C) all fish included except *Aldrichetta forsteri* ($R^2 = 0.71$). Blue lines: locally weighted scatterplot smoothing curves (LOWESS)

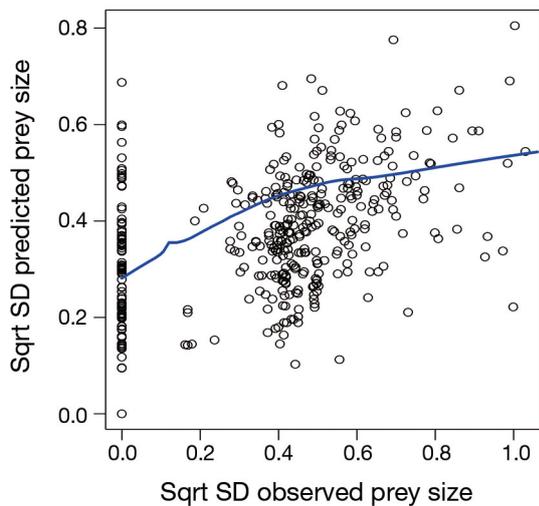


Fig. 5. Relationship between predicted and observed square root of the standard deviation (SD) of mean prey size for fishes in Western Port, Australia ($R^2 = 0.36$). Blue line: locally weighted scatterplot smoothing curves (LOWESS)

We estimated the standard deviation of the predicted mean prey size using a third model that included predicted mean prey size and all other predictors. We found a significant correlation between the predicted and observed standard deviations ($R^2 = 0.36$) (Fig. 5). As observed for the correlation between predicted and observed mean prey size, removing predictors indicated that the most important predictor was taxonomic information (Table 3). Furthermore, regression of the 90% quantile between the \log_e of the WW of the consumer fish and the \log_e of the standard deviation of the observed prey size exhibited a positive trend (slope = 1.24, 95% CI = [0.04, 2.13]), suggesting that the diet of larger fish encompassed a wider size range (Fig. 6).

The relationship between the observed prey types and the \log_e WW of the consumer fish revealed that certain prey types were significantly related to the size of the consumer (Fig. 7).

DISCUSSION

This study quantitatively assessed the importance of body size and taxonomic identity in predicting fish diet composition. We found that for diet estimates, family identity was nearly as important as species identity, and identity was approximately equivalent to body size. Thus, both body size and taxonomic information are required to properly predict the diet of a fish. Moreover, species identity was more important than body size in predicting the size of each prey

type consumed, whereas consumer size was more important than species identity in determining which prey categories were consumed. These results are the inverse of expectations (i.e. that consumer size would be the best predictor of prey size, and fish species identity the best predictor of prey type). If these results are found to be generally applicable, it could have important implications for the development and application of size-based models used to assess the impacts of fishing on community structure and function (Shin & Cury 2001, 2004, Hall et al. 2006, Andersen & Pedersen 2010).

The majority of marine trophic models use taxonomic identity and functional groups to draw their conclusions (Polovina 1984, Pauly et al. 2000), despite the fact that species change their ecological roles by consuming different prey of progressively larger sizes at different stages of their ontogenetic development (Cushing 1975, Maury et al. 2007a,b, Rudolf & Rasmussen 2013). Only a few studies have considered body size as the most important factor in the ecological system (e.g. Shin & Cury 2001, Brown & Gillooly 2003, West & Brown 2005). Our 2-step model integrates both the taxonomic identity of the fish as well as its size in order to achieve accurate predictions of diet. This predictive diet model, developed for the fish community of Western Port, should theoretically be applicable to other locations provided that the necessary information of the consumer fishes and their prey is available. Further research is underway to test such extension.

Our Consume model predicted the percentages of different prey types with reasonable accuracy, as well as the size of dietary items consumed for a particular fish of known species and size. The loss in accuracy was assessed for both prey type and mean prey size predictions if taxonomic, size or environmental data were lacking. The most important predictor for prey type was determined to be the size (wet weight) of the consumer fish.

That consumer size was the most important predictor for prey type is likely related to the ontogenetic changes of the different species, consuming different diets at different stages in their life, which implies changing ecological roles within their habitat (Cushing 1975, Dickie et al. 1987, Jennings & Warr 2003, Rudolf & Rasmussen 2013). In addition, certain prey types of larger size would only be available to larger fish (e.g. fish preying on other fish). The relationships we found between prey type and consumer size (Fig. 7) explain why the prey types were more accurately predicted when fish size was considered rather than taxonomic or other predictors. It is interesting

Table 3. Correlation between the predicted and observed standard deviations of the predicted mean prey size for fish collected in Western Port, Australia. Also shown are the correlations for all the predictors available and the subsequent loss in accuracy when some of these predictors were removed. The weighting of the correlation was based on the predicted percentages of prey type. WW= \log_e wet weight of consumer fish. mo = month

	Predictors	R ² with subset of 12 ind. species ⁻¹
With all predictors	Species, WW, prey type, mean prey size, habitat, site, mo	0.36
Without species	Genus, WW, prey type, mean prey size, habitat, site, mo	0.32
Without species and genus	Family, WW, prey type, mean prey size, habitat, site, mo	0.3
Without species, genus and family	WW, prey type, habitat, mean prey size, site, mo	0.19
Without WW	Species, prey type, mean prey size, habitat, site, mo	0.34
Without habitat	Species, WW, prey type, mean prey size, site, mo	0.34
Without site	Species, WW, prey type, mean prey size, habitat, mo	0.32
Without month	Species, WW, prey type, mean prey size, habitat, site	0.33
Without habitat, site, month	Species, WW, prey type, mean prey size	0.3
Without prey type	Species, WW, mean prey size, habitat, site, mo	0.29
Without mean prey size	Species, WW, prey type, habitat, site, mo	0.34
Without all predictors but site	Site	0.01
Without all predictors but habitat	Habitat	0.03

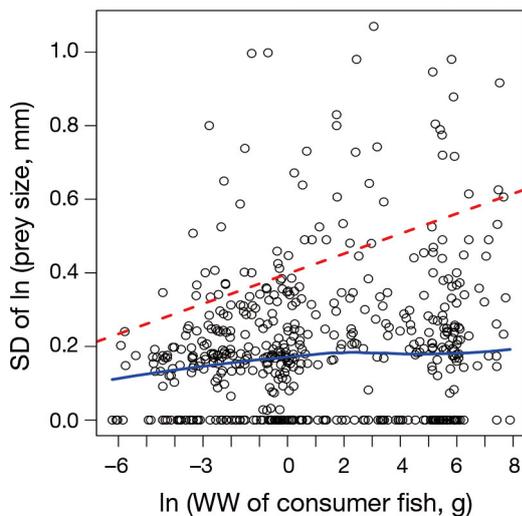


Fig. 6. Observed standard deviation of the \log_e prey size versus the \log_e of the wet weight (WW) of consumer fish. $R^2 = 0.02$. Red line: 90% quantile regression; blue line: locally weighted scatterplot smoothing curves (LOWESS)

that mollusc infauna were consumed mainly (but not exclusively) by elephant sharks *Callorhynchus mili* of relatively large size, which possess specialized mouth parts that help them dig these molluscs from the sediment (Table S5 in Supplement 4 at www.int-res.com/articles/suppl/m545p239_supp/). Changes in prey type related to size have also been documented in a general review of the literature (Werner & Gilliam 1984). Our results show that this trend relates not only to the individual species level, but also at the level of the entire fish community of Western Port.

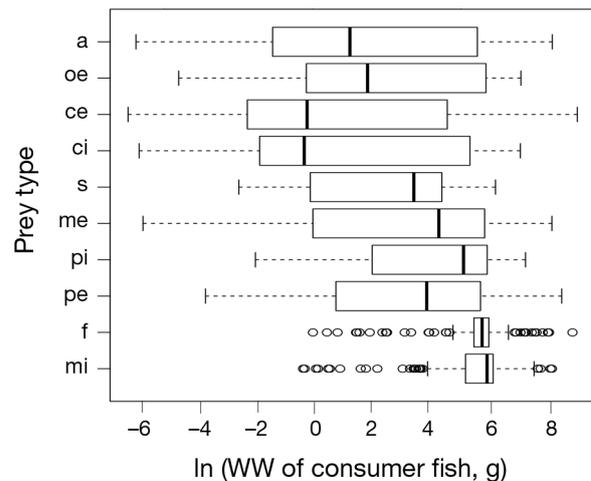


Fig. 7. Relationship between wet weight (WW) of consumer fish and observed prey types (a: algae; oe: other epifauna; ce: crustacean epifauna; ci: crustacean infauna; s: sponges; me: mollusc epifauna; pi: polychaete infauna; pe: polychaete epifauna; f: fish; mi: mollusc infauna). Mid-line: median; box limits: 25th and 75th percentiles; whiskers: 1.5 times the inter-quartile range or the maximum or minimum value, whichever is closer to the median. The points are outliers

The most important predictor for mean prey size was information on the species, genus or family of the consumer fish. Fish size was second in importance as a predictor after taxonomic information, but also predicted the size of the prey with reasonable accuracy. Nevertheless, we expected that the size of the consumer would be the most important predictor for both prey type and prey size based on previous work in southern Australia (Edgar & Shaw 1995c) as

well as other studies that have highlighted the importance of consumer size in prey predictions (Cushing 1975, Rudolf & Rasmussen 2013). However, Edgar & Shaw (1995b) pointed out that the relationship between the size of the consumer and its prey was consistent for most species they encountered in Western Port other than the yellow-eyed mullet *Aldrichetta forsteri*, which consumed prey considerably smaller than predicted based on its size. The importance of taxonomic information as a factor in the model is largely explained by the variation in prey size shown by certain species outside size predictions, as was the case for the yellow-eyed mullet.

In order to balance the contribution of each species to overall predictions, the model randomly selected 12 individuals from each species in a wide range of sizes. This was necessary because some species, most notably *A. forsteri*, contributed a disproportionately large number of individuals to the total sample ($n = 325$ for *A. forsteri*), biasing the results. The loss in accuracy for mean prey size predictions for the balanced contribution was reduced from 16 to 13% when the model used the subset of 12 ind. species⁻¹. In a previous study, Scharf et al. (2000) also found that the distribution of predator size–prey size was related to species identity and had an asymmetric distribution. In addition, they suggested that the larger piscivorous fishes continued to consume prey of small size as well as including larger prey items in their diet, and that the slope in the relationship between the consumer and the maximum prey size showed a significant variation between species, adding support to the importance of taxonomic information for prey size predictions.

The importance of taxonomic information in predicting mean prey size is also explained by inter-specific variation in maximum body size. For example, a greater range of body sizes is possible for larger fish species, and hence the potential for ontogenetic diet shifts will be greater than that of small species that quickly reach maximum size. Furthermore, some prey types will not be available to small fish species. A relationship between prey type and mean prey size suggests this was the case within our dataset, and that maximum body size of fishes likely contributed to the importance of taxonomic identity for prey size predictions.

The standard deviation of the mean prey size confirmed the assumption that larger fish consumed a wider size range of prey. These estimated standard deviations could also be used to feed into community level studies, to generate predictive probability distributions that describe the likelihood of prey of a par-

ticular size being consumed. The sequential removal of predictors shows that taxonomic information is the most important estimator for prey size breadth.

Little loss of information occurred when family rather than species identity was used. Thus, the predictions for a fish species lacking detailed dietary information may have much the same accuracy as the predictions for a fish species belonging to the same family for which this information is available. This has important positive implications for community level analyses where many species have not been previously studied.

In conclusion, the Consume model presented in this study allows accurate prediction of the type and mean size of prey, and prey size distribution for consumer fish of known identity and size. Rather than trophically subdividing fishes using categorical divisions based on species (e.g. higher carnivores, benthic carnivores, herbivores and planktivores), this model allows a more rigorous compartmentalization of prey consumption by describing diets at an individual fish level, including those species that exhibit large ontogenetic changes in diet. Through the aggregation of data on total prey consumption for whole assemblages of individual fish, this novel approach should open new pathways in trophic ecology.

Acknowledgements. This study was supported by the Australian Research Council, a postgraduate scholarship from TGRS (to G.A.S.), and the Marine Biodiversity Hub, a collaborative partnership funded under the Australian Government's National Environmental Research Program. We thank Craig Shaw for assistance in field sampling, Justin Hulls for designing the map of the study area, and Dr. Kilian Stehfest for supporting the development of the model. We also thank Dr. Simon Wotherspoon and Dr. Andre Punt for their input and comments in developing the diet predictive model.

LITERATURE CITED

- Al-Habsi SH, Sweeting CJ, Polunin NVC, Graham NAJ (2008) $\delta^{15}\text{N}$ and $\delta^{13}\text{C}$ elucidation of size-structured food webs in a Western Arabian Sea demersal trawl assemblage. *Mar Ecol Prog Ser* 353:55–63
- Andersen KH, Pedersen M (2010) Damped trophic cascades driven by fishing in model marine ecosystems. *Proc R Soc B* 277:795–802
- Barber D (2011) Bayesian reasoning and machine learning. Cambridge University Press, Cambridge
- Belgrano A (2005) Aquatic food webs: an ecosystem approach. Oxford University Press, Oxford
- Blanchard JL, Law R, Castle MD, Jennings S (2011) Coupled energy pathways and the resilience of size-structured food webs. *Theor Ecol* 4:289–300
- Brown JH, Gillooly JF (2003) Ecological food webs: high-quality data facilitate theoretical unification. *Proc Natl Acad Sci USA* 100:1467–1468

- Brown JH, Gillooly JF, Allen AP, Savage VM, West GB (2004) Towards a metabolic theory of ecology. *Ecology* 85:1771–1789
- Conway D, White J (2012) Machine learning for hackers. O'Reilly Media, Sebastopol, CA
- Cushing DH (1975) Marine ecology and fisheries. Cambridge University Press, Cambridge
- Dickie LM, Kerr SR, Boudreau PR (1987) Size-dependent processes underlying regularities in ecosystem structure. *Ecol Monogr* 57:233–250
- Edgar GJ, Shaw C (1995a) The production and trophic ecology of shallow-water fish assemblages in southern Australia. I. Species richness, size-structure and production of fishes in Western Port, Victoria. *J Exp Mar Biol Ecol* 194:53–81
- Edgar GJ, Shaw C (1995b) The production and trophic ecology of shallow-water fish assemblages in southern Australia. II. Diets of fishes and trophic relationships between fishes and benthos at Western Port, Victoria. *J Exp Mar Biol Ecol* 194:83–106
- Edgar GJ, Shaw C (1995c) The production and trophic ecology of shallow-water fish assemblages in southern Australia. III. General relationships between sediments, seagrasses, invertebrates and fishes. *J Exp Mar Biol Ecol* 194:107–131
- Green SJ, Côté IM (2014) Trait-based diet selection: prey behaviour and morphology predict vulnerability to predation in reef fish communities. *J Anim Ecol* 83:1451–1460
- Hahm W, Langton R (1984) Prey selection based on predator/prey weight ratios for some northwest Atlantic fish. *Mar Ecol Prog Ser* 19:1–5
- Hall SJ, Collie JS, Duplisea DE, Jennings S, Bravington M, Link J (2006) A length-based multispecies model for evaluating community responses to fishing. *Can J Fish Aquat Sci* 63:1344–1359
- Jennings S, Warr KJ (2003) Smaller predator-prey body size ratios in longer food chains. *Proc R Soc B* 270:1413–1417
- Jennings S, Pinnegar JK, Polunin NV, Boon TW (2001) Weak cross-species relationships between body size and trophic level belie powerful size-based trophic structuring in fish communities. *J Anim Ecol* 70:934–944
- Jennings S, Warr KJ, Mackinson S (2002) Use of size-based production and stable isotope analyses to predict trophic transfer efficiencies and predator-prey body mass ratios in food webs. *Mar Ecol Prog Ser* 240:11–20
- Link JS (2004) A general model of selectivity for fish feeding: a rank proportion algorithm. *Trans Am Fish Soc* 133:655–673
- Maury O, Faugeras B, Shin YJ, Poggiale JC, Ari TB, Marsac F (2007a) Modeling environmental effects on the size-structured energy flow through marine ecosystems. Part 1: The model. *Prog Oceanogr* 74:479–499
- Maury O, Shin YJ, Faugeras B, Ari TB, Marsac F (2007b) Modeling environmental effects on the size-structured energy flow through marine ecosystems. Part 2: Simulations. *Prog Oceanogr* 74:500–514
- O'Brien WJ, Slade NA, Vinyard GL (1976) Apparent size as the determinant of prey selection by bluegill sunfish (*Lepomis macrochirus*). *Ecology* 57:1304–1310
- Pauly D, Christensen V, Walters C (2000) Ecopath, Ecosim, and Ecospace as tools for evaluating ecosystem impact of fisheries. *ICES J Mar Sci* 57:697–706
- Pinnegar JK, Tomczak MT, Link JS (2014) How to determine the likely indirect food-web consequences of a newly introduced non-native species: a worked example. *Ecol Model* 272:379–387
- Polovina JJ (1984) Model of a coral reef ecosystem. *Coral Reefs* 3:1–11
- R Development Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Rudolf VH, Rasmussen NL (2013) Ontogenetic functional diversity: size structure of a keystone predator drives functioning of a complex ecosystem. *Ecology* 94:1046–1056
- Scharf FS, Juanes F, Rountree RA (2000) Predator size-prey size relationships of marine fish predators: interspecific variation and effects of ontogeny and body size on trophic-niche breadth. *Mar Ecol Prog Ser* 208:229–248
- Shin YJ, Cury P (2001) Exploring fish community dynamics through size-dependent trophic interactions using a spatialized individual-based model. *Aquat Living Resour* 14:65–80
- Shin YJ, Cury P (2004) Using an individual-based model of fish assemblages to study the response of size spectra to changes in fishing. *Can J Fish Aquat Sci* 61:414–431
- Trebilco R, Baum JK, Salomon AK, Dulvy NK (2013) Ecosystem ecology: size-based constraints on the pyramids of life. *Trends Ecol Evol* 28:423–431
- Werner EE, Gilliam JF (1984) The ontogenetic niche and species interactions in size-structured populations. *Annu Rev Ecol Syst* 15:393–425
- Werner EE, Hall DJ (1974) Optimal foraging and the size selection of prey by the bluegill sunfish (*Lepomis macrochirus*). *Ecology* 55:1042–1052
- West GB, Brown JH (2005) The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization. *J Exp Biol* 208:1575–1592
- Wright DI, O'Brien WJ (1984) The development and field test of a tactical model of the planktivorous feeding of white crappie (*Pomoxis annularis*). *Ecol Monogr* 54:65–98
- Zhang L, Hartvig M, Knudsen K, Andersen KH (2014) Size-based predictions of food web patterns. *Theor Ecol* 7:23–33

Editorial responsibility: Konstantinos Stergiou, Thessaloniki, Greece

Submitted: May 26, 2015; Accepted: December 15, 2015
Proofs received from author(s): February 4, 2016