



## Predicting the diet of coastal fishes at a continental scale based on taxonomy and body size



German A. Soler<sup>a,\*</sup>, Graham J. Edgar<sup>a</sup>, Rick D. Stuart-Smith<sup>a</sup>, Anthony D.M. Smith<sup>b</sup>, Russell J. Thomson<sup>a</sup>

<sup>a</sup> Institute for Marine and Antarctic Studies, University of Tasmania, Private Bag 49, Hobart, TAS 7001, Australia

<sup>b</sup> CSIRO Oceans and Atmosphere, GPO Box 1538, Hobart, TAS 7001, Australia

### ARTICLE INFO

#### Article history:

Received 2 December 2015

Received in revised form 26 March 2016

Accepted 26 March 2016

Available online 6 April 2016

#### Keywords:

Temperate Australia

Trophic model

Community consumption

Mean prey size

Seagrass

### ABSTRACT

Predicting diet of animals in ecological communities is necessary for a better understanding of trophic links and piecing together food webs to inform ecosystem-based management. A dietary model, *Consume*, was recently developed to predict detailed dietary information for fishes on the basis of fish identity and size. This model was field-tested over a continental scale, predicting community-level consumption for other temperate marine fish communities that differed in species composition and size structure. Using local stomach contents data to field-test predictions, accurate performance of the model was found across 14 locations around southern Australia. Prey type and mean prey size were predicted for fishes at new locations with high accuracy (mean percentage overlap between predicted and actual prey types = 73%;  $r^2$  between predicted and observed mean prey size = 89%) when trained with stomach contents data from subsets of sampled fishes at all locations. Model accuracy dropped, but was still respectable, when using training data only from one location (prey type accuracy = 67%; mean prey size  $r^2$  = 56%). Prey type was more accurately predicted on the basis of consumer body size than species identity, while consumer family identity and size were needed for accurate prediction of mean prey size. The most important factors were evaluated by leaving out predictors (species, genus and family identity; size of consumer; habitat, location, ecoregion and biogeographic province). Exclusion of geographical location information resulted in little loss in accuracy. Our results highlight the need for consideration of consumer body size in trophic models, rather than binning species into functional groups solely on the basis of taxonomy. Application of *Consume* to situations where no dietary information exists, but at least fish family identity and size structure are known, will provide a novel mechanism for testing important ecological hypotheses and assessing trophic consequences of anthropogenically-induced changes in community structure.

© 2016 Elsevier B.V. All rights reserved.

### 1. Introduction

Predation is an important process determining the structure of marine communities (Russ 1980), with critical information on the nature and magnitude of this captured in the diet of predators in the system (Edgar and Shaw 1995a, 1995b). Fish represent major predators in shallow marine systems, and the ability to predict the diet of fishes at a given location should allow improved understanding of the ecological dynamics of the community (Edgar and Shaw 1995b). Making predictions is often necessary because dietary data do not typically exist, or at least not for the majority of species present in a fish community. Dietary predictions for individual animals will allow more accurate models of community consumption and food webs with which to answer key questions about the ecology of marine systems in relation

to human impacts and management interventions, in the context of the effects of environmental variables. Nevertheless, the challenge with testing ecological predictions at large spatial scales is considerable due to logistical constraints and the great spatial variability in the natural environment (Edgar and Shaw 1995b; Peters 1991). Community-level calculations are useful, such as when assessing impacts of fishing and other threats on food web processes. At present, it is near-impossible to identify community-level patterns of consumption from integration of limited dietary data available for particular fish species, given the diversity of fishes of different body sizes at any single location, let alone across multiple locations.

A predictive diet model, *Consume*, was developed using dietary data from shallow water marine fishes sampled in Western Port, Victoria (Soler et al. 2016). Some of the findings of the study in which this model is described included: (a) prey type and mean prey size were accurately predicted for consumer fish of known species and size (77% accurate for prey type and 93% for mean prey size; prey type accuracy was calculated as the mean percentage overlap between predicted and actual prey types, and mean prey size accuracy as the correlation ( $r^2$ ) between predicted and observed mean prey size); (b) when only

\* Corresponding author.

E-mail addresses: [german.soler@utas.edu.au](mailto:german.soler@utas.edu.au) (G.A. Soler), [G.Edgar@utas.edu.au](mailto:G.Edgar@utas.edu.au) (G.J. Edgar), [Rick.StuartSmith@utas.edu.au](mailto:Rick.StuartSmith@utas.edu.au) (R.D. Stuart-Smith), [Tony.D.Smith@csiro.au](mailto:Tony.D.Smith@csiro.au) (A.D.M. Smith), [Russell.Thomson@utas.edu.au](mailto:Russell.Thomson@utas.edu.au) (R.J. Thomson).

the family of the consumer, rather than the species identity, was included, the loss in accuracy of the models was small (~1%); (c) the most important predictor for prey type was the size of the consumer; and (d) the most important predictor for mean prey size was the taxonomic identity of the consumer.

In this study, the accuracy of the *Consume* model was tested across a larger spatial and taxonomic domain in order to determine the generality and broader utility of this novel means to quantify food web links in the absence of detailed dietary information. The fish stomach content dataset described by Edgar and Shaw (1995a, 1995b) was used for model training and field-testing, and encompassed 14 locations in southern Australia, plus five locations in Western Port, Victoria (Fig. 1). These locations extend over 3000 km of coast and six marine ecoregions of the world and two provinces (Spalding et al. 2007).

The most important predictors of fish diet were tested to determine if they could be generalized over large spatial scales, or whether important location-specific factors and local community composition result in idiosyncratic patterns of prey consumption which may prohibit accurate larger-scale size-based food web modelling. To examine this, the loss in accuracy was tested when predicting fish diet for the 14 locations in southern Australia using models trained only on the Western Port dataset, and therefore naïve to locally-collected dietary data, in comparison to model accuracy when trained with local data from a subset of sampled fishes. Finally, the loss in accuracy was estimated for prey type and mean prey size predictions when information on taxonomy, consumer fish size or locality (habitat, location, province and/or ecoregion) is lacking.

## 2. Methodology

A total of 4336 fish were sampled using seine and gill nets from 19 shallow marine locations in southern Australia, extending from Rottneest Island in Western Australia to Jervis Bay in New South Wales, including five in Western Port, Victoria (Fig. 1, Table 1S; Edgar and Shaw, 1995b). For each location, fish were caught, measured, weighed, and the stomach contents studied using consistent methods (see Edgar and Shaw (1995b)). Stomach contents were identified to the lowest taxonomic level possible, measured using a microscope graticule or Vernier callipers, and binned into 19 log-scale size-classes ranging from 0.125 mm to 64 mm (Edgar and Shaw 1995a).

For dietary predictions, the *Consume* model developed in R-Studio (R-Core-Team, 2014) in a previous study (Soler et al. 2016) was used. *Consume* has two steps, the first step involved predicting the percentage of different prey types for each individual fish. For the second step, the mean prey size was predicted for a given prey type of an individual fish. The prediction of prey type percentage for a given fish was made using a set of fish with known diets, that are most similar to the given fish. This method is akin to a *k*-nearest neighbour procedure (Barber 2011; Conway and White 2012). The prediction of mean prey size (step 2) was made using linear regression models, where the assumptions of normality were met by applying a log transformation to the mean prey size. The importance of predictors was evaluated via cross validation for both steps of the *Consume* model.

In order to estimate the effect of location in the diet predictions, province and ecoregion (Spalding et al. 2007) were included in the *Consume* model. The Western Port and southern Australia dataset was

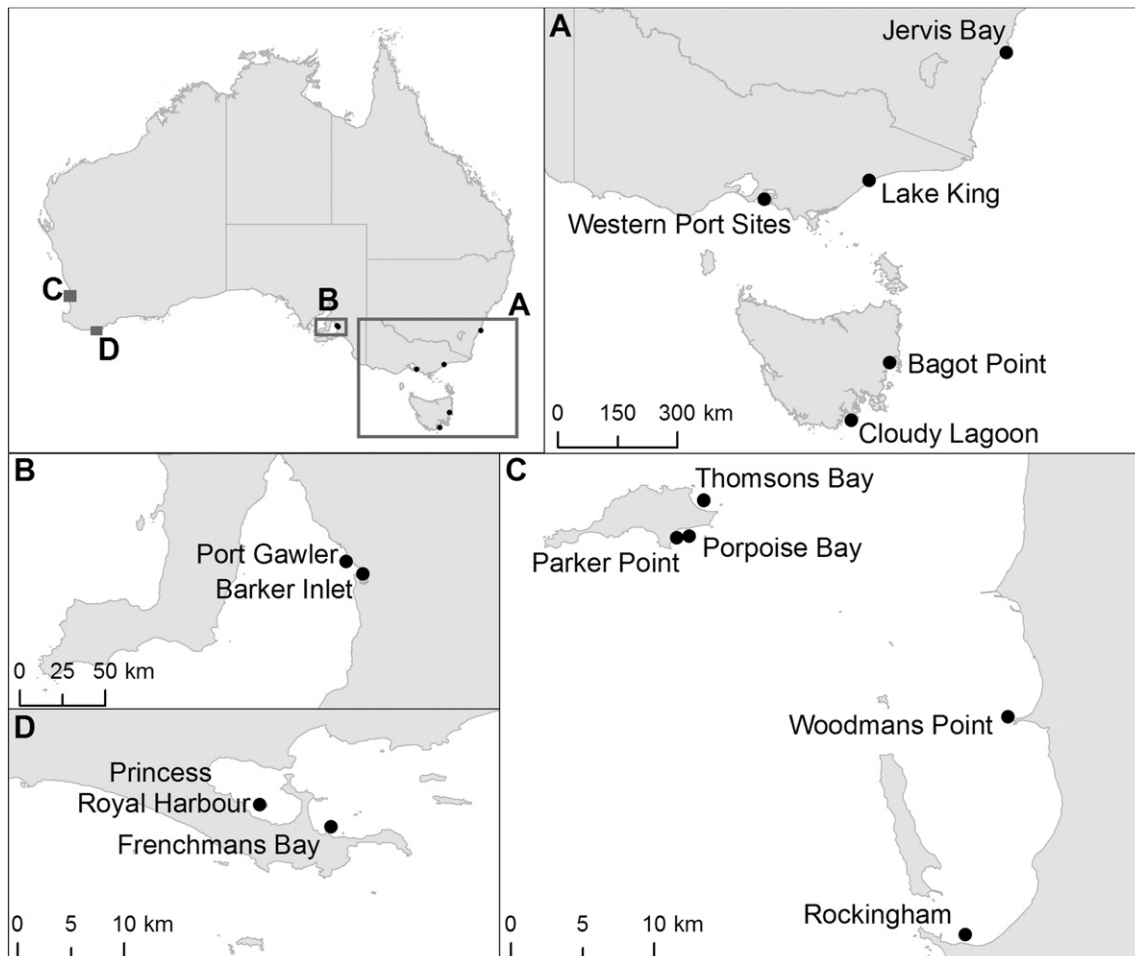


Fig. 1. Locations sampled in southern Australia.

used to predict the prey type and mean prey size for fish at all locations. To measure the accuracy of the prey type predictions, the overlap in percentages of different dietary items between predicted and observed values was calculated, as described by Soler et al. (2016). The correlation ( $r^2$ ) between predicted and observed mean prey size was used as a measure of accuracy of the model (Soler et al. 2016). Different combinations of predictors were omitted to assess their effects on the accuracy of the model.

The accuracy of mean prey size predictions was tested as the weighted correlation ( $r^2$ ) between the predicted mean prey size and the observed mean prey size. The weighting of the correlation was based on the diet percentage predictions. This means that each of the predicted prey types was not given equal weight, but rather weighted the calculation by the proportion of each of the predicted prey types. If a weighted correlation had not been used and, for example, the model predicted 1% prey type A and 99% prey type B but size was not predicted well for prey type A, this would have resulted in a low correlation despite the fact that prey type A was rare. Little difference between the weighted correlation and the non-weighted correlation was found, suggesting that the mean prey size was predicted with reasonable consistency across the different prey types. To graph the relationship between the predicted and the observed mean prey size, the locally-weighted smoothing LOWESS line was used. The LOWESS line allows for a curvilinear relationship whereas the regression forces the curve to be linear.

For computing the accuracy of models, predictions were based on a training dataset that excluded data from the consumer fish whose diets were being predicted (Soler et al. 2016). In order to assess if the model based on a localised dataset was able to predict across larger scales, the Western Port data was used to inform predictions of prey type and mean prey size for fishes across the 14 other locations. Results were validated using information on observed fish diets from these locations.

To avoid a few species with large sample sizes biasing predictions, the same algorithm as used in Soler et al. (2016) to randomly select 12 individual fish was applied, encompassing a wide range of sizes, to represent each species, location and habitat. Initial trials indicated that 12 individuals per species gave the maximum accuracy for prey type and mean prey size predictions (Soler et al. 2016). Different subsets of 12 individuals per species were used as predictors for fish diet in different runs of the model, and variability associated with the selection of subsets was assessed.

The model used for predictions of mean prey size was (Soler et al. 2016):

$$y_i = \mu + \beta_1 \text{ species}_i + \beta_2 \text{ WW}_i + \beta_3 \text{ prey type}_i + \beta_4 \text{ habitat}_i + \beta_5 \text{ location}_i + \beta_6 \text{ province}_i + \beta_7 \text{ ecoregion}_i + \varepsilon_i \quad (1)$$

where  $y_i$  = the  $\log_e$  mean prey size prediction for the  $i$ th consumer fish, given the effects of species, wet weight (WW), prey type, habitat, location, province and ecoregion;  $\mu$  = overall mean;  $\varepsilon_i$  = residual error.

When predicting for the southern Australian locations using only Western Port data, the model prediction progressed sequentially from species to genus to family, and finally to wet weight, whenever the exact match was not found. This allowed the model to generate predictions for consumer fish that lacked information on that species, genus or family in the Western Port dataset. Furthermore, when predicting for southern Australian locations using the Western Port dataset, the habitat, location, ecoregion or province were not included in the models. Habitat was excluded because southern Australian locations included an additional habitat (*Posidonia*) that was not present in Western Port. Moreover, both ecoregion and province comprised extra levels not present in Western Port.

From preliminary analyses (Soler et al. 2016), it was identified that the most important predictors were taxonomy, size of the consumer fish (wet weight) and prey type. Hence, the model used for predicting the diet of fish for the southern Australian locations using the Western Port data was as follows (using notation as in Eq. (1)):

$$y_i = \mu + \beta_1 \text{ Species}_i + \beta_2 \text{ WW}_i + \beta_3 \text{ Prey type}_i + \varepsilon_i \quad (2)$$

Prey types were categorised into similar trophic and functional groups as applied in the initial Western Port model (Soler et al. 2016) with eleven prey types in total: algae, sponges, epifaunal crustaceans, infaunal crustaceans, epifaunal molluscs, infaunal molluscs, epifaunal polychaetes, infaunal polychaetes, other epifauna, other infauna, and fishes (as prey).

In order to assess breadth in mean prey size within diets, the standard deviation of mean prey size in guts was predicted for the 14 locations in southern Australia using the Western Port dataset, and the same model described by Soler et al. (2016). This model also included a progression mechanism from species to genus to family and finally to wet weight whenever the exact taxonomic match was not found, as described above for mean prey size predictions.

### 3. Results

When using the *Consume* model (Soler et al. 2016) to predict the diet of fish at a continental scale, prey type was predicted with an accuracy of  $73 \pm 1\%$ , where  $\pm 1\%$  indicates the variation associated with different sets of 12 individual fish per species randomly selected in each run of the model. These predictions were based on the complete dataset including Western Port and other southern Australian locations. The accuracy in prey type predictions for southern Australia (Western Port excluded), only dropped marginally, to  $67 \pm 1\%$ , when only Western Port data were used to train the model (and when all predictors were used).

The loss in the accuracy in the prey type predictions when predictors were sequentially dropped is presented for all locations in Table 1, and with Western Port locations excluded in Table 2. In both cases, the most important predictor for prey type was the wet weight of the consumer fish, while taxonomic identity contributed little. When the

**Table 1**

Accuracy (%) for predicted prey type of fishes at all sampled locations in southern Australia, including Western Port (WW =  $\log_e$  wet weight of consumer fish).

	Predictors	Accuracy (%) with all locations included
With all predictors	Species, WW, habitat, location, province, ecoregion	73
Without species	Genus, WW, habitat, location, province, ecoregion	73
Without species and genus	Family, WW, habitat, location, province, ecoregion	73
Without species, genus and family	WW, habitat, location, province, ecoregion	70
Without WW	Species, habitat, location, province, ecoregion	62
Without habitat	Species, WW, location, province, ecoregion	73
Without location	Species, WW, habitat, province, ecoregion	72
Without ecoregion	Species, WW, habitat, province	71
Without province	Species, WW, habitat	71
Without all predictors but species and WW	Species, WW	71
Without all predictors but location	Location	24
Without all predictors but province	Province	20

**Table 2**

Accuracy (%) for predicted prey type of fishes in southern Australian locations except Western Port, using model trained only with Western Port data (WW =  $\log_e$  wet weight of consumer fish).

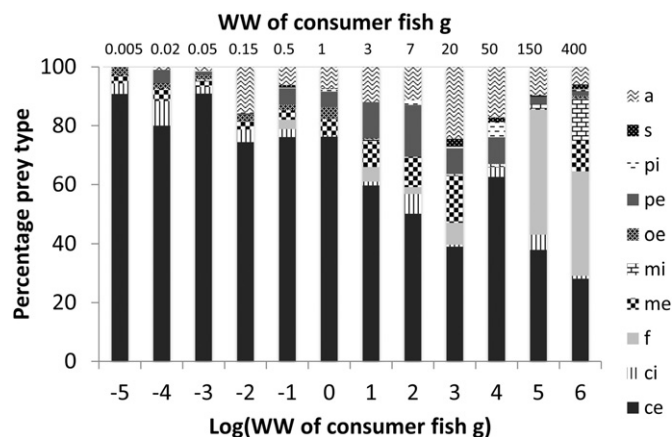
	Predictors	Accuracy (%)
With all predictors	Species, WW	67
Without species	Genus, WW	67
Without species and genus	Family, WW	67
Without species, genus and family	WW	66
Without WW	Species	40

weight of the consumer fish was removed from the model, the accuracy dropped to  $62 \pm 1\%$  (11% loss in accuracy) for the complete dataset and to  $40 \pm 1\%$  (27% loss in accuracy) using only Western Port data to train the model. The loss in accuracy when taxonomic information was excluded was  $\leq 3\%$  (Tables 1 and 2).

When the predicted prey type percentages were related to the natural log of the wet weight of the consumer fish, the most common prey types were found to change with consumer size (Fig. 2). For this figure (Fig. 2), the predicted prey type was estimated without the taxonomic information. The prey types that changed the most with consumer size were crustacean epifauna (ce), small fish (f), and mollusc infauna (mi). In the small size classes, crustacean epifauna (ce) was the preferred prey, dropping with increasing size of the consumer. Small fish (f) prey type increased with increases in the wet weight of the consumer, whereas mollusc infauna (mi) was a relatively important prey type for the larger fishes (Fig. 2). Algae (a) also became a relatively important food item for the middle-sized fishes.

The correlation ( $r^2$ ) between observed and predicted mean prey size (Table 3; Fig. 3) for the complete dataset was high ( $r^2 = 0.89 \pm 0.01$ ). Most of the predictors included had a significant contribution in the calculations (Table 2S). Predictions for the mean prey size (Table 4; Fig. 4) in southern Australia using only training data from Western Port were lower, but still respectable ( $r^2 = 0.56 \pm 0.01$ ).

The most important predictor for the mean prey size was the taxonomic information to at least family level; nevertheless, this was only marginally more important than the weight of the consumer fish, both for the whole dataset and when using only Western Port data as the training dataset (Tables 3 and 4). Furthermore, a high correlation was evident between the wet weight of the consumer fish and observed mean prey size consumed (Fig. 5;  $r^2 = 0.65$  for full dataset). The correlation between the predicted mean prey size, using all predictors except taxonomic information, and the weight of the consumer fish, was high (Fig. 5;  $r^2 = 0.94$ ). The loss in accuracy of predictions on



**Fig. 2.** Predicted prey type percentages in relation to  $\log_e$  wet weight (WW) of the consumer fish. Predictions were based on all predictors other than taxonomic information for the consumer fish. Prey types: algae (a), sponges (s), epifaunal polychaetes (pe), infaunal polychaetes (pi), other epifauna (oe), infaunal mollusc (mi), epifaunal mollusc (me), fish (f), infaunal crustaceans (ci) and epifaunal crustaceans (ce).

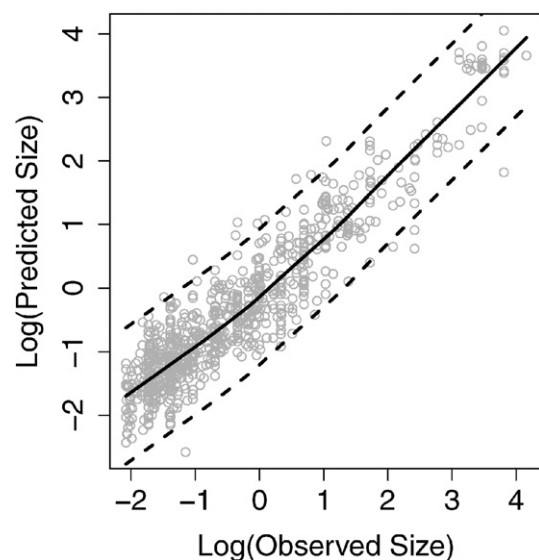
**Table 3**

Correlations between observed and predicted mean prey size of fishes in southern Australia including Western Port, and the change in accuracy when predictors are removed (WW =  $\log_e$  wet weight of consumer fish).

	Predictors	$r^2$
With all predictors	Species, WW, prey type, habitat, location, province, ecoregion	0.89
Without species	Genus, WW, prey type, habitat, location, province, ecoregion	0.89
Without species and genus	Family, WW, prey type, habitat, location, province, ecoregion	0.87
Without species, genus and family	WW, prey type, habitat, location, province, ecoregion	0.79
Without WW	Species, prey type, habitat, location, province, ecoregion	0.81
Without habitat	Species, WW, prey type, location, province, ecoregion	0.87
Without location	Species, WW, prey type, habitat, province, ecoregion	0.89
Without ecoregion	Species, WW, prey type, habitat, province	0.88
Without province	Species, WW, prey type, habitat	0.88
Without prey type	Species, WW, habitat, location, province, ecoregion	0.86
Without all predictors but species, WW and prey type	Species, WW, prey type	0.87
Without all predictors but species, WW and prey type	Species, WW	0.87
Without species, genus, family and WW	Prey type, habitat, location, province, ecoregion	0.42
Without species, genus, family, WW and prey type	Habitat, location, province, ecoregion	0.23
Without all predictors but location	Location	0.22
Without all predictors but province	Province	0.03

mean prey size and prey type was small when geographical factors (location, ecoregion and province) were removed from the model (Tables 1 and 3).

The correlation between the standard deviation of the natural log of the predicted mean prey size and the wet weight of the consumer fish shows only a slight positive relation ( $r^2 = 0.03$ ) (Fig. 6). The loss in accuracy between the predicted and the observed standard deviation (SD) of the natural log of mean prey size indicates that taxonomic



**Fig. 3.** Scatterplot relating predicted and observed  $\log_e$  mean prey size for southern Australia including Western Port ( $r^2 = 0.89$ ). The solid black line is the locally-weighted scatterplot smoothing curve (LOWESS). The dash black lines outline the upper prediction interval and the lower prediction interval (95%); both lines were estimated using LOWESS.



**Table 4**

Correlations between observed and predicted mean prey size of fishes in southern Australia (Western Port excluded), using only Western Port dataset for model training, and the change in accuracy when predictors are removed (WW =  $\log_e$  wet weight of consumer fish).

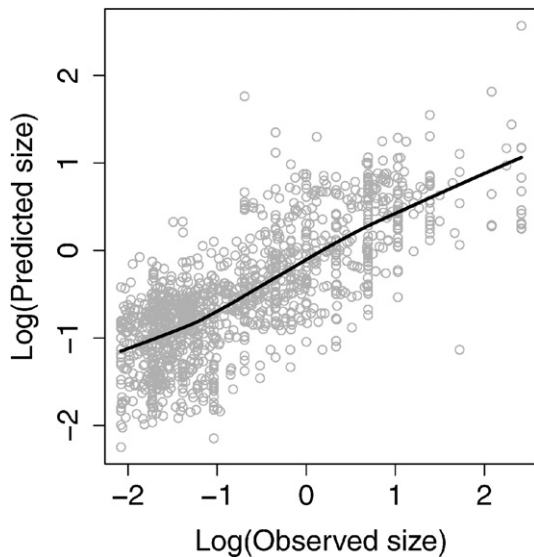
	Predictors	$r^2$
With all predictors	Species, WW, prey type	0.56
Without species	Genus, WW, prey type	0.55
Without species and genus	Family, WW, prey type	0.55
Without species, genus and family	WW, prey type	0.39
Without WW	Species, prey type	0.42
Without all predictors but prey type	Prey type	0.11

information is the most important predictor for SD of the prey (Table 5; Fig. 6).

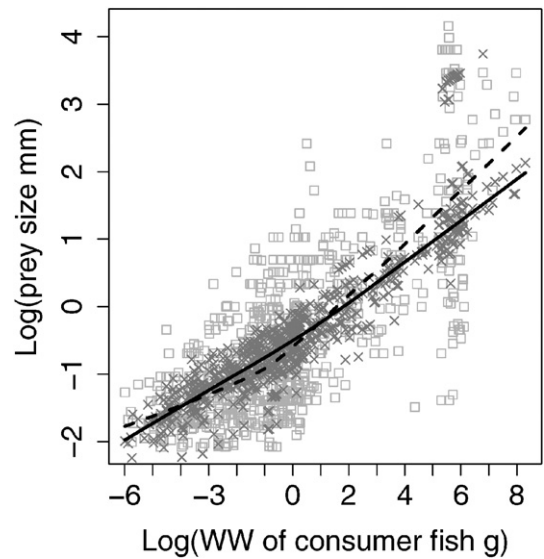
#### 4. Discussion

Diet predictions for fishes based on taxonomy and size provide opportunities to better understand the ecological dynamics of fish communities. A problem with many mechanistically-scaled models is that predictions cannot be tested – predictions are needed because empirical field data are lacking (Rastetter et al. 2003). The present study is unusual in that it includes assessments of the accuracy of predictions for prey type and mean prey size on the basis of observed dietary data. By demonstrating the accuracy of the model in circumstances where little or no dietary data are included, we have shown the utility of this tool for ecological studies in other locations containing different sets of species and for which no dietary data currently exists.

Diet prediction accuracy using the *Consume* model for prey type and mean prey size declined when predictions for southern Australian locations were made only on the basis of dietary data from Western Port; nevertheless, the correlation between predicted and observed values remained relatively high. Furthermore, the initial findings for Western Port were corroborated regarding the importance of taxonomy and size of the consumer fish to accurately predict its diet. Size of the consumer fish was by far the most important predictor for prey type at the larger geographical scale. While the initial findings from Western Port were also validated, showing that taxonomic identity was the most important predictor for prey size, the influence of consumer size was of near equal importance. In both prey type and prey size predictions, taxonomy and body size together achieved the best predictions.

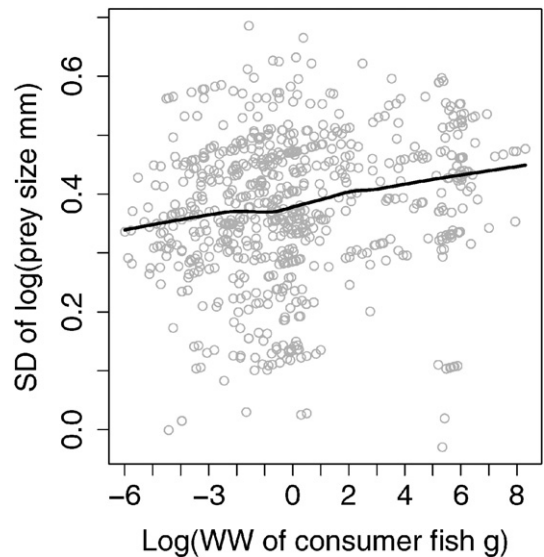


**Fig. 4.** Scatterplot relating predicted and observed  $\log_e$  mean prey size for all consumer fish in the southern Australia locations (Western Port excluded) using training data from Western Port ( $r^2 = 0.56$ ). The solid black line is the locally-weighted scatterplot smoothing curve (LOWESS).



**Fig. 5.** Scatterplot relating  $\log_e$  of observed (grey squares) and predicted (darker grey crosses) mean prey size in mm ( $\log_e$ ) against  $\log_e$  wet weight (WW) of the consumer fish in g ( $\log_e$ ) for all locations investigated in southern Australia, including Western Port. The model used a random sample of 12 individuals per species with a wide range of size. The dashed black line ( $r^2 = 0.65$ ) is the observed mean prey size locally-weighted scatterplot smoothing curve (LOWESS). The black solid LOWESS line relates predicted mean prey size ( $r^2 = 0.94$ ) to the size of the consumer fish (WW), with habitat, location and ecoregion included in models but no taxonomic information.

Body size is the most important predictor for prey type, presumably because of the magnitude of changes in the diets of fishes through different life-history stages (Soler et al. 2016). Individuals of different size within a species must therefore play different functional roles within marine ecosystems. Other authors have previously recognised the importance of ontogenetic changes in prey size, as well as variations in the trophic level. For example, Cushing (1975) found that the prey size of herring (*Clupea harengus*) ranged up to 4 orders of magnitude through different life history stages. Body size also relates to the quantity of resources exploited (Werner and Gilliam 1984), and has been suggested as a good measure of trophic energy flow within the fish community, relating the size of the fish and its prey (Dickie et al.



**Fig. 6.** Scatterplot relating predicted standard deviation of the  $\log_e$  mean prey size (mm) and the  $\log_e$  of the wet weight (WW) of the consumer fish (g) for all locations investigated in southern Australia using all predictors ( $r^2 = 0.03$ ). LOWESS curve is shown in solid black.

**Table 5**  
Correlation between the predicted and the observed standard deviations for dietary items in all of southern Australia, Western Port included (WW = log<sub>e</sub> wet weight of consumer fish).

	Predictors	r <sup>2</sup>
With all predictors	Species, WW, prey type, mean prey size, habitat, location, ecoregion	0.28
Without species	Genus, WW, prey type, mean prey size, habitat, location, ecoregion	0.27
Without species and genus	Family, WW, prey type, mean prey size, habitat, location, ecoregion	0.25
Without species, genus and family	WW, prey type, habitat, mean prey size, location, ecoregion	0.13
Without WW	Species, prey type, mean prey size, habitat, location, ecoregion	0.25
Without habitat	Species, WW, prey type, mean prey size, location, ecoregion	0.26
Without location	Species, WW, prey type, mean prey size, habitat, ecoregion	0.26
Without ecoregion	Species, WW, prey type, mean prey size, habitat, location	0.25
Without habitat, location, ecoregion	Species, WW, prey type, mean prey size	0.24
Without prey type	Species, WW, mean prey size, habitat, location, ecoregion	0.23
Without mean prey size	Species, WW, prey type, habitat, location, ecoregion	0.28
Without all predictors but location	Location	0.05
Without all predictors but habitat	Habitat	0.02

1987). The length of food chains is affected by predator–prey mass ratios, highlighting the importance of body size in ecological studies (Jennings and Warr 2003).

The importance of ontogenetic dietary shifts is also well-known from other taxa and ecosystems, including aquatic instar development in freshwater insects like dragonflies, and further illustrates the variety of ways in which ontogenetic changes can affect community structure (Rudolf and Rasmussen 2013). Thresholds are often evident, where certain prey types only become available when an animal reaches a particular size. This has been particularly noted for carnivorous marine fishes (Kulbicki et al. 2005). In the associated Western Port study (Soler et al. 2016), it was found that profound changes in the predicted prey category were directly related to the wet weight of the consumer fish, although in that case, it was due to a shift in feeding strategy with growth in body size, as diet categories were inclusive of the full range of potential prey size.

As highlighted by other authors (Maury et al. 2007a; Maury et al. 2007b), the majority of trophic models in marine ecosystems categorise the system using species and functional groups, with the underlying assumption that individuals within these groups are similar (Pauly et al. 2000; Polovina 1984). By contrast, a few models have considered animal size (e.g. biomass), with the assumption that size is the most important factor driving the strength of interactions in the ecological system (Brown and Gillooly 2003; Fulton et al. 2011; Shin and Cury 2001; West and Brown 2005). These models reasonably assume that most predators are larger than their prey, determining that the predator–prey relations are mostly based on the size of the predator (Jennings et al. 2001; Jennings et al. 2002; Scharf et al. 2000). Furthermore, Jennings et al. (2001) showed 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 in the North Sea. The *Consume* model indicates that body size, rather than taxonomy, is the most important predictor of which prey types are being consumed by shallow water marine fishes. This predictive diet model should therefore enable improved assessments of the trophic compartments occupied by individuals of a species through its lifespan.

Although taxonomic identity (at the species, genus or family level) was not as important as size when predicting the type of food consumed by fish, it was critical for predicting the size of prey consumed. This was identified in the original Western Port study, and was here confirmed at a continental scale. Previous studies have reached a similar conclusion: that taxonomic identity was related to prey size and that, within species, prey size was related with the size of the predatory fish (Andersen and Beyer 2006; Juanes 1994; Law et al. 2012; Mittelbach and Persson 1998; Scharf et al. 2000; Soler et al. 2016). In the current study, the size of the consumer fish (i.e. wet weight) was also a good predictor for prey size. The correlation between the predicted mean prey size (with no taxonomic information) and wet weight of the consumer fish was higher than the correlation between the observed

mean prey size and wet weight (Fig. 5). A possible explanation why the observed size correlation is lower is that the empirical field data are affected by a range of other factors not considered in the models, which add noise to the relationship.

The standard deviation of sizes of items in stomach contents was better predicted by taxonomic information than body size. This indicates that certain species are feeding from a larger prey size array than members of other species of similar size; species with relatively small mouths, such as mullet (Mugilidae), were found to consume a smaller range of prey sizes than those with larger mouths. In a study conducted in the northeast US where the diet of 18 fish species was considered, the maximum and minimum prey sizes, as well as the breadth of prey size, were also found to be principally related to the taxonomic identity of the predatory fish (Scharf et al. 2000).

Another important result from this study was that the loss in accuracy of the predictions was small when using the family of the consumer fish as the taxonomic identity, instead of the species. This has important implications for generality of the results, suggesting that prey type and size can be accurately predicted for other previously unassessed members of the families included to train the model. Any available information from published studies on diets of other species in the family would add further to the accuracy of predicted diets for unstudied species. Clearly, diets of species are constrained by morphological rigidity at the family level – most families are visibly recognisable because of similar sets of particular morphological characteristics. The results suggest there is an equivalent degree of rigidity in diet amongst species. Exceptions clearly exist, however, particularly amongst very large families such as wrasses and gobies, where great morphological diversity occurs. A study of New Caledonian marine fishes reported significant intra-family variation in diets (in Lethrinidae), as well as prey consumption similarities within other families (Kulbicki et al. 2005).

Geographical location was found to contribute only marginally to diet predictions for both prey type and prey size – consumer fishes of a particular family and body size had similar diets regardless of location. Considering the patchiness in the availability of each particular food type across the marine domain, a number of possible mechanisms could contribute to this result. It is possible that local representatives of each family associate with particular sets of conditions or microhabitats that contain similar mixes of potential prey types, regardless of location. Diet type categories used in this study were relatively broad, and are more likely to occur across multiple regions than had prey type been more finely partitioned, but they will still occur in varying proportions at different locations. It may also be that fishes selectively remove particular prey items, regardless of density in the seascape, and such selectivity has an element of consistency within families. While further work is needed to determine this, the results of this study most likely reflect a combination of these and other mechanisms.

In conclusion, the *Consume* model generated accurate predictions of prey type and prey size for consumer fishes at a continental scale. The model was sufficiently robust to predict across a geographical domain

using training data from a single location, although accuracy substantially improved, particularly for prey size, when broader-scale training data were utilised. The most essential predictors for the diet estimates were the size of the consumer fish followed by family identity, albeit these elements contributed differently when predicting prey type versus prey size. Community-level estimates using this dietary model, where diet predictions based on body size and family identity are aggregated for all individuals at a location, should provide a novel perspective to food web studies, and can capitalise on more readily available community-structure data.

### Acknowledgements

This study was supported by the Australian Research Council (LP100200122), a Tasmania Graduate Research Scholarship (to G.A.S.), and the Marine Biodiversity Hub, a collaborative partnership funded under the Australian Government's National Environmental Science Programme. We would like to thank Justin Hulls for designing the map and Dr Kilian Stehfest for supporting the development of the model. We would also like to thank Dr Simon Wotherspoon and Dr Andre Punt for their input and comments in developing the predictive diet model. The authors have no conflict of interest to declare. [SW]

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.jembe.2016.03.017>.

### References

- Andersen, K.H., Beyer, J., 2006. Asymptotic size determines species abundance in the marine size spectrum. *Am. Nat.* 168, 54–61.
- Barber, D., 2011. *Bayesian Reasoning and Machine Learning*. Cambridge University Press.
- Brown, J.H., Gillooly, J.F., 2003. Ecological food webs: high-quality data facilitate theoretical unification. *Proc. Natl. Acad. Sci.* 100, 1467–1468.
- Conway, D., White, J., 2012. *Machine Learning for Hackers*. O'Reilly Media, Inc.
- Cushing, D.H., 1975. *Marine Ecology and Fisheries*. Cambridge University Press.
- Dickie, L.M., Kerr, S.R., Boudreau, P.R., 1987. Size-dependent processes underlying regularities in ecosystem structure. *Ecol. Monogr.* 57, 233–250.
- Edgar, G.J., Shaw, C., 1995a. 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, G.J., Shaw, C., 1995b. 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.
- Fulton, E.A., Link, J.S., Kaplan, I.C., Savina-Rolland, M., Johnson, P., Ainsworth, C., Horne, P., Gorton, R., Gamble, R.J., Smith, A.D.M., Smith, D.C., 2011. Lessons in modelling and management of marine ecosystems: the Atlantis experience. *Fish Fish.* 12, 171–188.
- R-Core-Team, 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Jennings, S., Warr, K.J., 2003. Smaller predator–prey body size ratios in longer food chains. *Proceedings of the Royal Society of London. Ser. B Biol. Sci.* 270, 1413–1417.
- Jennings, S., Pinnegar, J.K., Polunin, N.V., Boon, T.W., 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., Pinnegar, J.K., Polunin, N.V., Warr, K.J., 2002. Linking size-based and trophic analyses of benthic community structure. *Mar. Ecol. Prog. Ser.* 226, 77–85.
- Juanes, F., 1994. What Determines Prey Size Selectivity in Piscivorous Fishes? In: Stouder, D.J., Fresh, K.L., Feller, R.J. (Eds.), *Theory and application in fish feeding ecology*. University of South Carolina Press, Columbia, pp. 79–100.
- Kulbicki, M., Bozec, Y.-M., Labrosse, P., Letourneur, Y., Mou-Tham, G., Wantiez, L., 2005. Diet composition of carnivorous fishes from coral reef lagoons of New Caledonia. *Aquat. Living Resour.* 18, 231–250.
- Law, R., Plank, M.J., Kolding, J., 2012. On balanced exploitation of marine ecosystems: results from dynamic size spectra. *ICES J. Mar. Sci.* 69, 602–614.
- Maurly, O., Faugetas, B., Shin, Y.-J., Poggiale, J.-C., Ari, T.B., Marsac, F., 2007a. Modeling environmental effects on the size-structured energy flow through marine ecosystems. Part 1: the model. *Prog. Oceanogr.* 74, 479–499.
- Maurly, O., Shin, Y.-J., Faugetas, B., Ari, T.B., Marsac, F., 2007b. Modeling environmental effects on the size-structured energy flow through marine ecosystems. Part 2: simulations. *Prog. Oceanogr.* 74, 500–514.
- Mittelbach, G.G., Persson, L., 1998. The ontogeny of piscivory and its ecological consequences. *Can. J. Fish. Aquat. Sci.* 55, 1454–1465.
- 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.
- Peters, R.H., 1991. *A Critique for Ecology*. Cambridge University Press, Cambridge, U.K.
- Polovina, J.J., 1984. Model of a coral reef ecosystem. *Coral Reefs* 3, 1–11.
- Rastetter, E.B., Aber, J.D., Peters, D.P., Ojima, D.S., Burke, I.C., 2003. Using mechanistic models to scale ecological processes across space and time. *Bioscience* 53, 68–76.
- Rudolf, V.H., Rasmussen, N.L., 2013. Ontogenetic functional diversity: size structure of a keystone predator drives functioning of a complex ecosystem. *Ecology* 94, 1046–1056.
- Russ, G.R., 1980. Effects of predation by fishes, competition, and structural complexity of the substratum on the establishment of a marine epifaunal community. *J. Exp. Mar. Biol. Ecol.* 42, 55–69.
- Scharf, F.S., Juanes, F., Rountree, R.A., 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, Y.-J., Cury, P., 2001. Exploring fish community dynamics through size-dependent trophic interactions using a spatialized individual-based model. *Aquat. Living Resour.* 14, 65–80.
- Soler, G.A., Thomson, R.J., Stuart-Smith, R.D., Smith, A.D.M., Edgar, G.J., 2016. The contributions of body size, habitat and taxonomy to predictions of temperate Australian fish diets. *Mar. Ecol. Prog. Ser.* 545, 239–249.
- Spalding, M.D., Fox, H.E., Allen, G.R., Davidson, N., Ferdaña, Z.A., Finlayson, M., Halpern, B.S., Jorge, M.A., Lombana, A., Lourie, S.A., Martin, K.D., McManus, K.D., Molnar, J., Recchia, C.A., Robertson, J., 2007. Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *Bioscience* 57, 573–583.
- Werner, E.E., Gilliam, J.F., 1984. The ontogenetic niche and species interactions in size-structured populations. *Annu. Rev. Ecol. Syst.* 15, 393–425.
- West, G.B., Brown, J.H., 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.