Development of a demersal fish community classification for New Zealand's Exclusive Economic Zone

NIWA Client Report: HAM2006-062
June 2006

NIWA Project: DOC 06213

Development of a demersal fish community map for New Zealand's Exclusive Economic Zone

John Leathwick
Malcolm Francis
Kathryn Julian

Prepared for

Department of Conservation

NIWA Client Report: HAM2006-062
June 2006

NIWA Project: DOC 06213

National Institute of Water & Atmospheric Research Ltd
Gate 10, Silverdale Road, Hamilton
P O Box 11115, Hamilton, New Zealand
Phone +64-7-856 7026, Fax +64-7-856 0151
www.niwa.co.nz

Contents

Reviewed by:	*Approved for release by:*
P. Taylor	S. Elliott

Formatting checked

A. Bartley

1. Introduction

The intention of the New Zealand Government to provide comprehensive protection of the marine biodiversity values contained within its Exclusive Economic Zone (EEZ) is clearly signalled in the New Zealand Biodiversity Strategy (Anonymous 2000). This enunciates a broad vision for the management of New Zealand's biodiversity that includes the protection of a full range of marine habitats and ecosystems that are "representative of New Zealand's indigenous marine biodiversity". Priority actions identified to bring this vision to reality include the development and implementation of "a strategy for establishing a network of areas that protect marine biodiversity", with a specific target of protecting 10% of New Zealand's marine environments by 2010.

The development of such a network of protected areas is a complex task that requires consideration of a broad range of factors (see Margules and Pressey 2000). As Roberts et al. (2003) argue, information about biological distributions is fundamental to the process of reserve selection. That is not to say that other factors are unimportant, but rather that biological information should inform and underpin the subsequent decision-making that also takes into account some wider set of factors, including social and economic considerations.

Until recently, we have had only relatively limited access to spatially comprehensive information on the distributions of most of New Zealand's marine species. For some groups of species, particularly those in deeper waters, distributional data is sketchy or even largely lacking. One strategy to allow robust management of marine environments in the absence of such data has been the production of an environmental classification known as the Marine Environment Classification (Leathwick et al. 2004, Snelder et al. 2005). This classification is based on a set of eight environmental variables, selected for their functional relevance to biota. In addition, variables were weighted on the basis of their contribution to statistical models that used them as predictors in explaining the distributions of biological groups for which some distributional data was available. Multivariate classification tools were then used to combine these environmental variables to define a hierarchical classification that groups together sites having similar environmental character, on the premise that similar environments are likely to support similar biological assemblages.

For a few groups of species, distributional information of varying quality is available, and over recent years increasing effort has been made to integrate data stored in various locations and make it more readily available for general use (e.g., Ocean Biogeographic Information System - http://www.iobis.org/). For these species, recent advances in the statistical analysis of distributional data have also opened up new avenues for the integration of biological and environmental data to provide increased understanding of species distributions (e.g., Guisan and Zimmerman 2000). In this study, we demonstrate the use of such tools to develop distribution maps for a comprehensive set of demersal fish, the biological group for which we have probably the best quality distributional data. Using data collected as catch records from research trawls, we use statistical models to relate the distributions of 122 fish species to environment. We then use these individual models to make environment-based, spatially comprehensive predictions of individual species distributions across the entire EEZ, including sites from which trawl data are lacking. Finally, we use multivariate classification tools to define a fish community classification for the EEZ that defines the geographic distribution of particular communities, and describes both their composition, and the environmental conditions in which they occur.

We expect that both the individual species predictions, and the community classification will have a range of uses for both conservation and fisheries management. This includes their use with analytical tools specifically designed to provide robust, data-driven approaches to the identification of representative areas for protection. Exploration of the use of one of these with the predicted fish distribution data layers described here is described in a separate report (Leathwick et al. 2006a). A third report (Leathwick et al. 2006b) describes the production of an environmental classification optimised specifically to the distributions of demersal fish, using the same trawl data set as we use here.

2. Methods

The overall approach used here was to fit statistical models relating the distributions of 122 fish species (listed in Appendix 1) to a set of environmental variables, with the latter chosen for their functional relevance (see Leathwick et al in press). We then combined the statistical models with equivalent gridded environmental data at a resolution of 1 km and covering the entire EEZ to form predictions of individual fish distributions.

2.1 Input data

Fish distribution data were drawn from the *fish_comm* research bottom trawl database. This database is a groomed version of the Ministry of Fisheries trawl database of bottom trawl tows carried out by research vessels between 1979 and 2005. Grooming procedures placed special emphasis on the accuracy of species identification and the geographic coordinates of trawl tows. We enlarged the *fish_comm* database used previously by Francis et al. (2002) and Leathwick et al. (in press) by adding a further 4368 trawls that extended both the spatial and temporal extent of sampling. In particular, these extended the southern limit of the sampling to include the full extent of the EEZ, rather than stopping at 50°S as in Leathwick et al. (in press), and the date of the most recent sampling was extended from 1997 by including trawl survey records up to and including 2005. The total number of trawl records was 21,314.

Catch data describe the weight in kilograms by species of all species caught in 1% or more of trawls. A total of 29 vessels were used in collecting these samples, with the majority collected by the Kaharoa (3592) and Tangaroa (5714). Trawl parameters used in the analysis consisted of the trawl distance, trawl speed, and cod-end mesh-size. While it would have also been desirable to include factors such as the door spread and headline height, these parameters were not collected with sufficient reliability to allow their use.

Environmental predictors used in the analysis (shown in Table 1) were based on those used by Leathwick et al. (in press) with modifications as follows. Estimates of sea floor water temperature and salinity were derived from the World Oceans Atlas (Boyer et al. 2005) as described in Pinkerton et al. (2005). Estimates of suspended particulate matter and dissolved organic matter were derived from a case-2 analysis of satellite imagery (Pinkerton & Richardson 2005), and for consistency, estimates of chlorophyll-a concentration were also derived from a case-2 analysis.

Table 1: Environmental variables used in predicting the distributions of demersal fish species across New Zealand's Exclusive Economic Zone. New variables are indicated with an asterisk.

Variable	Details
AvgDepth	Average depth (m)
Temperature*	Estimated temperature at the sea floor (°C)
Salinity*	Estimated salinity at the sea floor (psu)
ChlaCase2	Chlorophyll-a concentration - derived from case-2 algorithm (ppm)
SSTGrad	Sea surface temperature spatial gradient (°C km^{-1})
TidalCurr	Depth averaged tidal current (m s^{-1})
CodendSize	Mesh-size of the trawl cod-end (mm)
Distance	Trawl distance (nm)
Slope	Sea-floor slope (°)
Pentade*	Year of trawling grouped in five-year intervals
Speed	Trawl speed (knots)
SusPartMat*	Suspended particulate matter (approximate to g m^{-3})
OrbVel	Wave-induced orbital velocity at the sea floor (m s^{-1})
DisOrgMat*	Dissolved organic matter (dimensionless index)

2.2 Model fitting

Model fitting was made difficult by the highly skewed distributions of the fish catch data, these consisting for most species of a mix of many zeros and a very high variance/mean ratio in the positive occurrences (Venables & Dichmont 2004). To address this, we used a delta-lognormal approach in which two models were fitted for each species (Fletcher et al. 2005). The first model used all trawls with data reduced to presence/absence form, and the model was fitted using a binomial error term so that the fitted values predict the probability of capture. The second model used only those trawls in which a catch occurred for the species of interest, and the catch weight was log-transformed prior to fitting a model assuming normally-distributed errors - fitted values indicate the amount caught, conditional on a catch occurring.

All models were fitted using Boosted Regression Trees (BRT), sometimes referred to as stochastic gradient boosting (Friedman 2001). This is an ensemble method in which a large number of relatively simple regression trees are fitted in an adaptive manner, i.e., new trees focus on improving model fit for those observations poorly explained by those trees already fitted. Predictions from all the individual trees are combined at the end to form the fitted values for the model. The number of trees fitted is controlled by the 'learning rate', a parameter that controls the amount of weight applied to each of the tree terms. In this analysis, learning rates were adjusted species by species to give a minimum of 500 trees in the final model. The complexity of the individual trees can also be varied. Fitting trees that consist of only one decision rule results in the fitting of an additive model closely approximating that fitted by a conventional regression capable of fitting non-linear functions, e.g., a generalised additive model (GAM - Hastie & Tibshirani 1990). However, using trees with more than one decision rule allows the automatic fitting of interactions between predictor variables, with the potential complexity of interaction effects increasing as the complexity of the individual trees is increased. This has the potential of providing much greater explanatory power than is possible with a purely additive model. All models in this analysis were fitted with trees having a complexity of ten, i.e., each tree consisted of ten decision rules. A more extended description of the use of BRT in a marine context is provided in Leathwick et al. (in press).

Because of the large amount of data, we split the trawl samples in two prior to analysis, with one component ("training data" - 17 000 trawls) used for model fitting, and the balance ("evaluation data" - 4314 trawls) used for independent assessment of model performance. All models were fitted on the training dataset using a cross-validation procedure (see Hastie et al. 2001, Leathwick et al in press) that commenced with the identification of ten randomly selected, mutually exclusive subsets of trawls. Each of these subsets was dropped in turn, forming ten data subsets, each containing 90% of trawls, and models of increasing numbers of trees were fitted to each of these. With each progressive addition of trees, which were added in sets of 100, predictions were made from each model to its withheld data subset, and the average prediction error was calculated across the ten models. The prediction error typically decreases rapidly in the initial stages as the model complexity is increased, but gradually slows as more trees are added. At some stage a point is reached at which further addition of trees degrades rather than improves the predictive power of the model, and once this was reached, model fitting was stopped. The number of trees giving best predictive performance was then identified, and was used to specify a model fitted to the entire dataset.

Model performance was evaluated using both the cross validation statistics calculated during model fitting, and by making predictions of fish occurrence or catch to the independent evaluation dataset. As a consequence, all performance statistics indicate the performance of the model when predicting to new sites. This gives a much more realistic view of model robustness than statistics that compare the model fitted values with the training data used to generate the model. The latter tend to be overly optimistic because of the tendency of models to over-fit, i.e., they adapt to specific features of the training data that do not have generality in the wider sampling universe.

Both presence/absence and catchmodels were evaluated by calculating the deviance explained when predicting to either withheld subsets of the training data during cross-validation, or to the completely independent set of evaluation data. Because of the variation in the total deviance across species, these values are expressed as percentages of total deviance. Presence/absence models were evaluated using area under the receiver operating characteristic curve (AUC) statistic (Fielding & Bell 1997). This statistic indicates the ability of a model to discriminate between presences and absences. Values of 0.5 indicate that a model can do no better than chance at distinguishing presence and absences, while a value of 1 indicates perfect discrimination. Catch models were evaluated by calculating the correlation between raw and fitted values.

2.3 Predicting fish distributions

Predictions were formed by combining the statistical models for each species with 1 km resolution environmental data covering the entire EEZ down to the maximum trawl depth, i.e., 1950 m. Variables describing trawl characteristics were set at fixed values for all grid cells as follows: distance - 2.3 nautical miles, speed - 3.2 knots, codend mesh size - 75 mm, and year - 2000. This gave separate predictions, standardized for variation in trawl conditions, of the probability of capture for that species, and the amount of fish caught when a non-zero catch occurred. The latter were back-transformed from the log-scale in which the model was fitted, onto the original measurement scale of kilograms, using a "smearing" correction factor (Duan 1983) to account for the bias introduced by back-transformation. Multiplying the predictions of fish catch by the predicted probabilities of capture gave the final predictions of fish catch. These were then imported into ArcView for checking against the raw distribution data - example maps are shown in Figs 1 and 2.

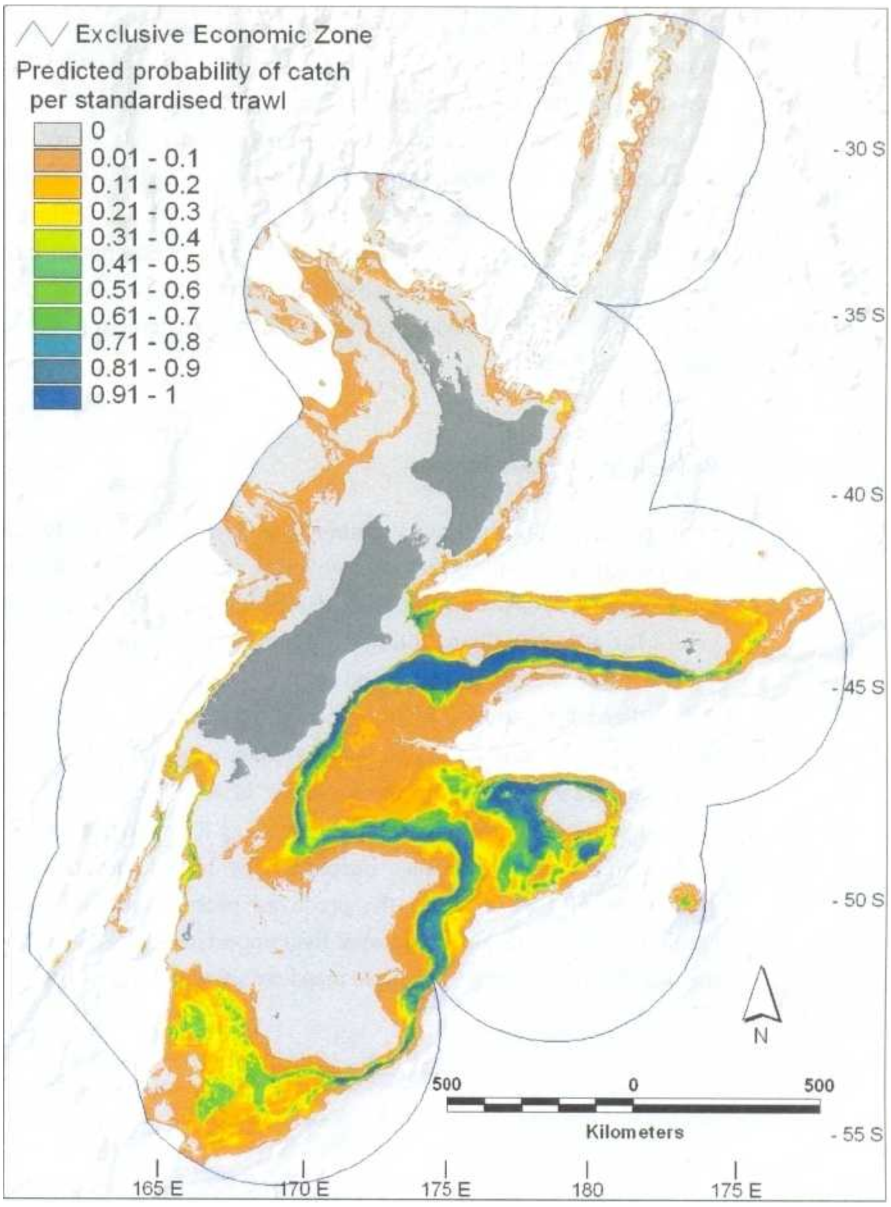


Figure 1: Geographic variation in the probability of capture from a standardised trawl for black oreo dory. Predictions are only made for grid squares whose depth is less than the maximum trawl depth recorded in the *fish_comm* database (1950m).

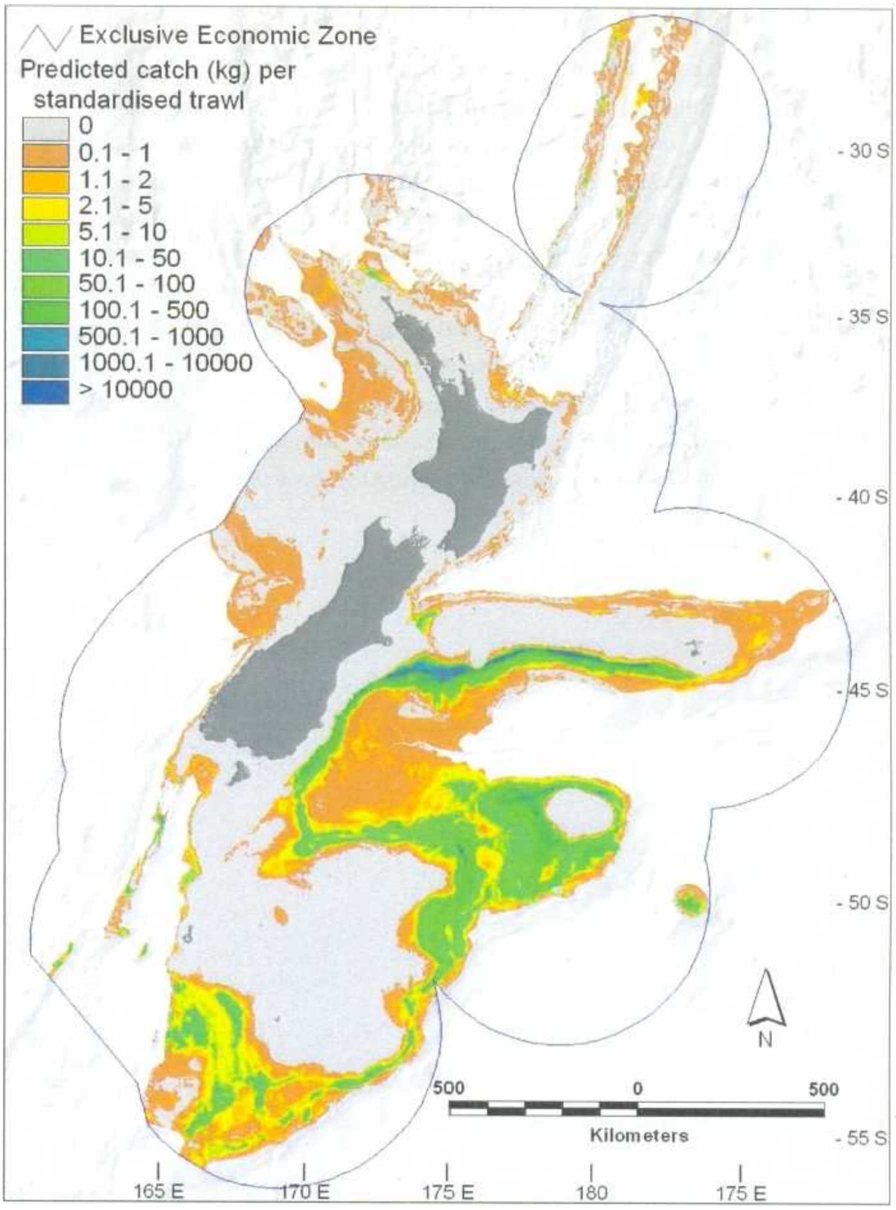


Figure 2: Geographic variation in the predicted catch (kg) from a standardised trawl for black oreo dory, derived from separate predictions of probability of catch, and catch weight conditional on a catch occurring. Predictions are only made for grid squares whose depth is less than the maximum trawl depth recorded in the fish comm database (1950m).

2.4 Defining a fish classification

Once predictions had been made for all species, they were exported and clustered using the multivariate statistical software PATN (Belbin 1995). This was performed in two stages. First, we performed a non-hierarchical classification in which grid cells were aggregated into 202 groups based on similarities in species composition as measured by the Bray-Curtis distance measure. A reduced data matrix describing the average composition for each of these groups was then analysed using conventional hierarchical classification. Results from the second phase of the analysis describe relationships between the groups from the initial clustering, and allow the overall classification results to be viewed at any level of classification detail from 2 to 202 groups. Results were imported back into ArcView and combined with a table summarising group membership, allowing rapid inspection of the geographic distributions of groups at varying levels of classification detail.

3. Results

3.1 Model performance

Performance statistics, shown in Table 2, indicate that the occurrence of most species can be predicted with a high degree of reliability. On average, the presence/absence regression models explained a little over 50% of the total deviance, and AUC statistics for these models averaged 0.95, indicating excellent discrimination of presences from absences. There was a weak relationship between model performance and species prevalence, i.e., while models for all species occurring in 10% or more of trawls tended to perform close to the average, there was more variation in performance in models for species with lower levels of occurrence. In fact, species with the lowest performing models (humpback rattail, viper fish, plunkets shark, longnose deep-sea skate, and Ray's bream) all had levels of occurrence of 5% or less. However, this trend was not consistent across all low prevalence species, and models for many of these performed close to or even above average.

Table 2: Performance of regression models relating fish presence/absence and catch to environment. Table entries indicate the mean performance statistic, averaged across all species, its standard error (in brackets - CV estimates only), and the range. Performance statistics consist of the percentage of deviance explained (both models), the area under the receiver operating characteristic curve (AUC) (presence/absence models), and the correlation between raw and fitted values (catch models). Values are shown both as calculated during the cross-validation procedure used to set model complexity (cv), and when predicting to 4314 independent trawls (independent).

	Presence/Absence	Catch
% deviance (cv)	51.9 (1.7), 16.0-85.6	30.3 (4.2), 0.0 to 67.4
deviance (independent)	52.3, 17.7-85.5	29.0, 0.0 to 67.1
AUC (cv)	0.95 (0.003), 0.84-0.99	na
AUC (independent)	0.95, 0.87-0.99	na
Correlation (cv)	na	0.529 (0.028), 0.05 to 0.82
Correlation (independent)	na	0.522, 0.05 to 0.82

Although similar patterns are apparent in the predictive models of fish catch, the average predictive success was lower for the catch models than for the presence-absence models. Regression models on average explained 29% of the total deviance, and cross-validated correlations between the raw and fitted values averaged a little over 0.5. The model for one species (longnose deep-sea skate) had virtually no explanatory power, explaining only 0.3% of the total deviance, and effectively predicting the mean catch for all sites. In addition the relationship between predictive performance and mean catch was stronger for the catch models than for the presence-absence models, with both the percent deviance explained and correlations between raw and fitted values increasing steadily with the mean catch of the fitted values. Models for species with catches averaging around 1 kg typically explained 20% of the deviance and had cross-validated correlations between the raw and fitted values of around 0.4, but this steadily increased with average catch weight so that models for species with an average catch of 100 kg explained 40-50% of the deviance and had correlations of 0.6 to 0.7.

When predictions were made for the completely independent set of 4314 trawls, results provide strong support for the robustness of the cross-validation procedure used in fitting the regression models. For example, the average percentage deviance explained when predicting to the independent trawls fell well within the range of variation around the average cross-validated mean deviance from the model fitting for both presence/absence and catch models. Similar close concordance was also apparent for the presence-absence model AUC scores and the catch model correlations between raw and fitted values.

3.2 Contributions of predictor variables

Average depth made the largest overall contribution to model outcomes (Table 3), but its proportional contribution for the presence-absence models (32.8%) was nearly twice that in the catch models (17.9%). Temperature and salinity were the next most important variables in both sets of models, followed by chlorophyll-a. Predictors describing variation in trawl methods made larger contributions to the catch models than to the presence-absence models.

Table 3: Percentage contributions of predictor variables to presence/absence (P/A) and catch models, averaged across 122 demersal fish species. Predictors are ranked in decreasing order with respect to their maximum contribution across the two sets of models.

Predictor	P/A	Catch	Average
AvgDepth	32.8	17.9	25.4
Temperature	14.0	12.7	13.4
Salinity	8.3	10.6	9.4
ChlaCase2	6.8	7.6	7.2
CodendSize	5.0	7.3	6.2
TidalCurr	4.8	6.8	5.8
SSTGrad	4.8	6.3	5.5
Pentade	4.1	5.8	4.9
SusPartMatter	4.0	5.4	4.7
Distance	3.7	4.6	4.2
Slope	3.5	4.6	4.1
OrbVel	3.2	4.2	3.7
Speed	3.1	3.4	3.2
DisOrgMatter	2.0	2.8	2.4

3.3 A demersal fish community classification

After exploration of several different levels of classification, we chose the 16-group level as giving a suitable balance between detail and good overall description of broad-scale patterns at the EEZ scale. At the broadest scale (Figs 3, 4, Table 4), there is a strong separation between groups 1-5 that occur in coastal and shelf environments, groups 6-13, that occur further offshore, predominantly in upper[1] and mid slope environments, and groups 14-16, that occur in the deepest waters accessible to trawling. In shallow waters there is further strong separation between the first two groups, which occur in northern and southern inshore waters respectively, and groups 3-5, which show similar latitudinal separation in deeper waters of the continental shelf. Similarly, groups 6-13 can be subdivided into those occurring predominantly on shallow parts of the upper slope (6-10) and those occurring predominantly on mid-slopes (11-13), i.e., at depths of 500 metres or more.

---

[1] We use the following definitions for these terms: upper slope -150-1000 m mid slope - 1000-2000 m; lower slope - 2000-4500 m; abyssal -4500 m.

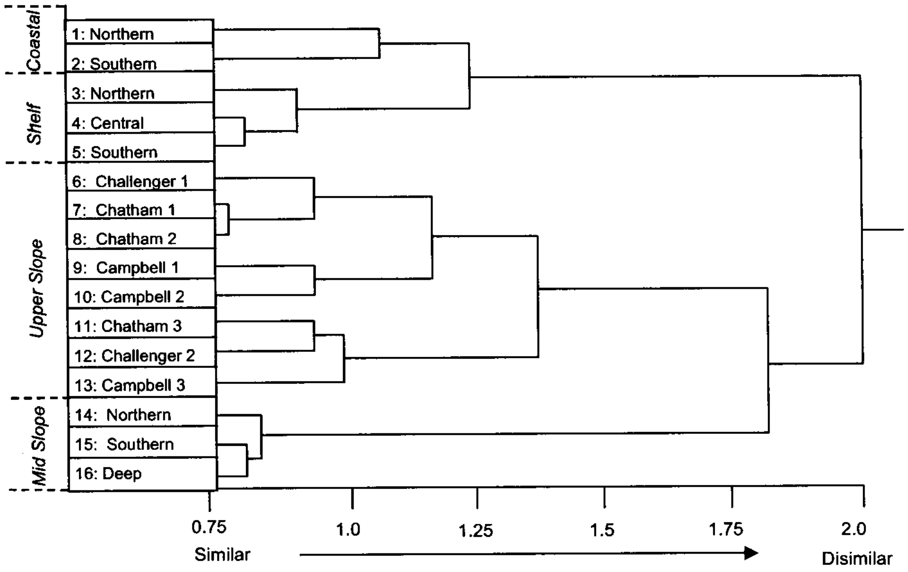


Figure 3: Dendrogram showing relationships between fish assemblages at a 16-group level.

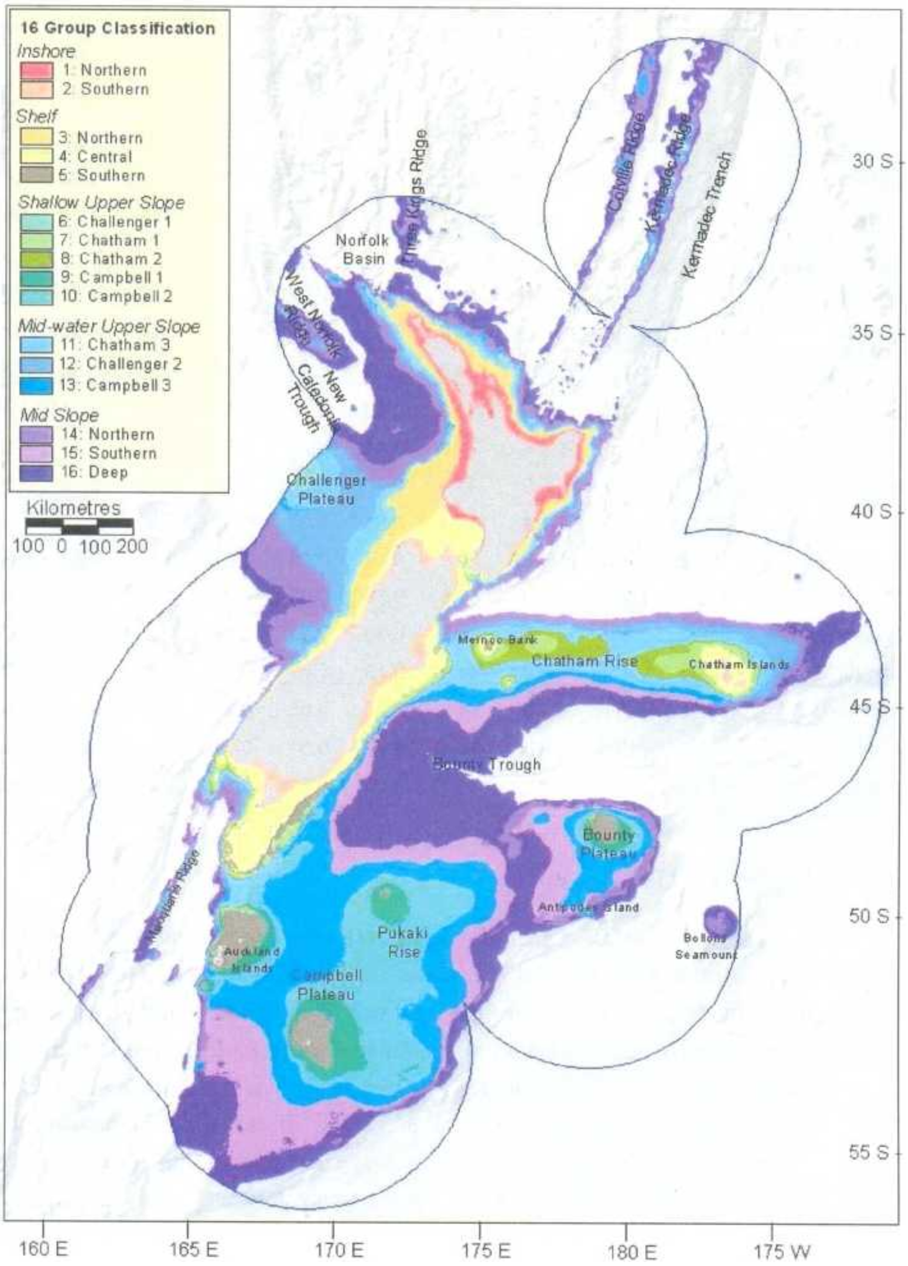
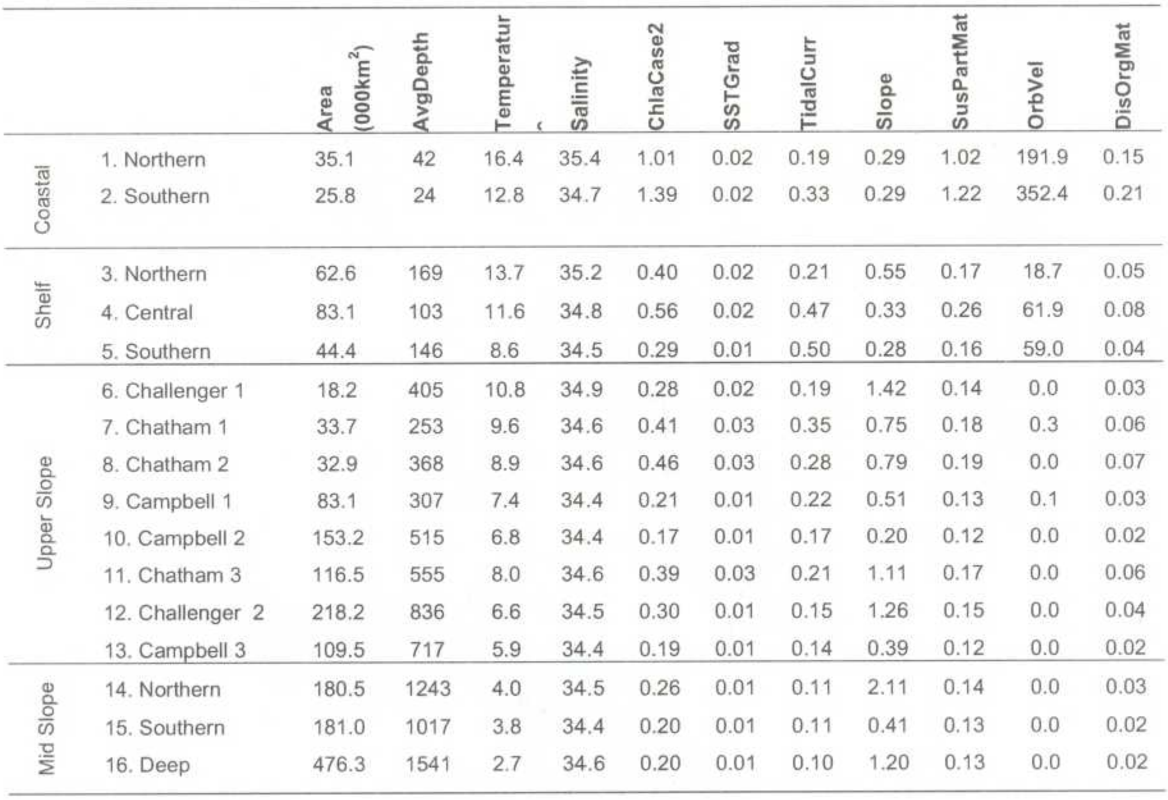


Figure 4: Geographic distribution of 16 fish assemblages, defined on the basis of similarity in fish composition for sites with an average depth less than the maximum depth recorded in the *fish_comm* research trawl database (1950 m).

Table 4: Geographic extent and average environmental conditions for the 16 fish assemblages



		Area (000km^2)	AvgDepth	Temperatur	Salinity	ChlaCase2	SSTGrad	TidalCurr	Slope	SusPartMat	OrbVel	DisOrgMat
Coastal	1. Northern	35.1	42	16.4	35.4	1.01	0.02	0.19	0.29	1.02	191.9	0.15
	2. Southern	25.8	24	12.8	34.7	1.39	0.02	0.33	0.29	1.22	352.4	0.21
Shelf	3. Northern	62.6	169	13.7	35.2	0.40	0.02	0.21	0.55	0.17	18.7	0.05
	4. Central	83.1	103	11.6	34.8	0.56	0.02	0.47	0.33	0.26	61.9	0.08
	5. Southern	44.4	146	8.6	34.5	0.29	0.01	0.50	0.28	0.16	59.0	0.04
Upper Slope	6. Challenger 1	18.2	405	10.8	34.9	0.28	0.02	0.19	1.42	0.14	0.0	0.03
	7. Chatham 1	33.7	253	9.6	34.6	0.41	0.03	0.35	0.75	0.18	0.3	0.06
	8. Chatham 2	32.9	368	8.9	34.6	0.46	0.03	0.28	0.79	0.19	0.0	0.07
	9. Campbell 1	83.1	307	7.4	34.4	0.21	0.01	0.22	0.51	0.13	0.1	0.03
	10. Campbell 2	153.2	515	6.8	34.4	0.17	0.01	0.17	0.20	0.12	0.0	0.02
	11. Chatham 3	116.5	555	8.0	34.6	0.39	0.03	0.21	1.11	0.17	0.0	0.06
	12. Challenger 2	218.2	836	6.6	34.5	0.30	0.01	0.15	1.26	0.15	0.0	0.04
	13. Campbell 3	109.5	717	5.9	34.4	0.19	0.01	0.14	0.39	0.12	0.0	0.02
Mid Slope	14. Northern	180.5	1243	4.0	34.5	0.26	0.01	0.11	2.11	0.14	0.0	0.03
	15. Southern	181.0	1017	3.8	34.4	0.20	0.01	0.11	0.41	0.13	0.0	0.02
	16. Deep	476.3	1541	2.7	34.6	0.20	0.01	0.10	1.20	0.13	0.0	0.02

3.4 Coastal Groups

1. Northern Coastal: occurs in shallow northern waters (mean depth = 42 m) south to about latitude 43°S. While it is extensive in the north, occurring to depths of 100 m or more, in the south it becomes progressively confined to shallow coastal waters. The most southerly occurrences are scattered around the northern and western coast of the South Island. Water temperatures are moderately warm and the salinity is very high. Orbital velocities are also high, as are levels of suspended particulate matter and chlorophyll-a. This group is relatively species poor- red gurnard are the most widespread species, followed by snapper, john dory and barracouta (Appendix 11). Catch rates are highest for snapper and barracouta. Spiny dogfish are occasionally caught in large amounts.

2. Southern Coastal: occurs in coastal waters from latitudes 41°S to 47°S, in depths of around 25 m, and is most extensive around the South Island coast, but with smaller areas around the southern tip of the North Island, and around the Chatham Islands. The environment is typified by intermediate temperatures and moderate salinity, with high concentrations of chlorophyll-a and suspended particulate matter, strong tidal currents and high orbital velocities. Spiny dogfish, barracouta, red cod and elephant fish occur most frequently and are caught in high volumes. Smaller quantities of red gurnard, school shark and common warehou are also caught.

3.4.1 Shelf Groups

3. Northern Shelf: occupies the inner continental shelf around the North Island (latitudes 33–44°S) at average depths of 170 m and in waters having moderately warm temperatures and high salinity. Orbital velocities, sea surface temperature gradients and tidal currents are all moderate. This group is relatively species poor-tarakihi and barracouta are the most widespread species, followed by carpet shark, school shark, spiny dogfish and frostfish. Catch rates are highest for spiny dogfish, barracouta and tarakihi. Silver dory, sea perch, horse mackerel, scaly gurnard and cucumber fish are also caught frequently but in small quantities.

4. Central shelf: occupies the inner continental shelf waters around central New Zealand (latitudes 39–49°S), and is widespread around the southwest of the North Island and the entire South Island - small patches occur in shallow waters on the Memoo Bank and around the Chatham Islands. It occurs at an average depth of around 100 m, where the environment is typified by intermediate temperatures, moderate salinity, strong to very strong tidal currents, and moderate orbital velocities and sea surface temperature gradients. Spiny dogfish and barracouta are the most widespread species, followed by tarakihi, school shark and carpet shark. Scaly gurnard and hapuku are caught frequently but only in small amounts.

5. Southern Shelf: occurs at depths of around 150 m in southern continental shelf waters between latitudes 42°S and 54°S, and is widespread as a narrow band along the south-eastern continental shelf of the South Island, on the Campbell Rise, Pukaki Rise and Bounty Plateau, and around the Auckland Islands. The environment is typified by moderately cold temperatures, low salinity, strong to very strong tidal currents, and moderate orbital velocities. Spiny dogfish and barracouta occur most frequently and have the highest catch rates followed by redbait, carpet shark and hapuku. Smaller volumes of silver warehou, silverside, silver dory and witch are also commonly caught.

3.4.2 Shallow upper slope groups

6. Challenger Plateau 1: occurs at depths of around 400 m along the upper slope mostly in the west and north (latitude 33–43°S). While it is most extensive on the

Challenger Plateau and off the western South Island coast, it also occurs as a narrow strip around the upper North Island and in scattered locations along the Kermadec Ridge. Environmental conditions are typified by intermediate temperatures and moderate salinity and sea surface temperature gradients. This group is relatively species poor - javelin fish, sea perch, ling, capro dory and hoki are the most widespread species with ling and hoki having the highest catches. Silver warehou and spiny dogfish are occasionally caught in large quantities.

7. **Chatham Rise 1:** occurs at depths of around 250 m, mostly on the Chatham Rise, but with more restricted occurrences along the east and south of the South Island, particularly in the northern end of the Solander Trough, and with more scattered occurrences south to about latitude 51°S. These waters have moderately cold temperatures and low salinity; strong tidal currents are common and sea surface temperature gradients are high. Spiny dogfish and silver warehou occur most frequently and are caught in high volumes. Red cod, barracouta, and silver dory are less frequently caught. Hoki are occasionally caught in large amounts.

8. **Chatham Rise 2:** occurs mostly adjacent to previous group at depths averaging 370 m across the Chatham Rise, and in a narrow band along the shelf edge off the eastern and southern South Island. The environment is typified by moderately cold temperatures, low salinity, moderate tidal currents and high sea surface temperature gradients. This group is relatively species rich-hoki, spiny dogfish, javelin fish and ling are the most widespread species followed by sea perch, lookdown dory, silverside and white warehou. Catch rates are highest for hoki and spiny dogfish.

9. **Campbell Plateau 1:** occurs at average depths of around 310 m from latitudes 47-54 °S, and is most extensive on the Campbell Rise, Bounty Plateau, Pukaki Rise, and around the Auckland Islands and the Antipodes. These waters have moderately cold temperatures, low salinity and moderate tidal currents. Silverside, spiny dogfish, southern blue whiting, ling and javelin fish are the most widespread species. Catch rates are highest for spiny dogfish, ling, southern blue whiting and hoki. Smaller catches of red cod, white warehou, witch and silver dory are also taken.

10. **Campbell Plateau 2:** occurs at average depths of 515 m in the southeast (latitudes 47–54°S); it is most extensive on the Campbell Plateau, around the Auckland Islands, and on the Bounty Plateau. The environment is typified by moderately cold temperatures and low salinity. Javelin fish, ling, hoki, silverside, hake and southern blue whiting are the most frequently caught species, while catch rates are highest for ling and southern blue whiting. Spiny dogfish, banded rattail, pale toadfish and longnose spookfish also occur frequently but with smaller catch rates.

3.4.3 Mid-water upper slope groups

11. Chatham Rise 3: occurs at depths of around 550 m over a wide latitudinal range (33–51°S). It forms a narrow band around much of the North and South Islands, but is most extensive on the Chatham Rise with a smaller area on the Challenger Plateau. It is more restricted in the south, occurring mostly around the Auckland Islands. The environment is typified by moderately cold temperatures, low salinity, high sea surface temperature gradients, and moderate tidal currents. Javelin fish, hoki and ling are the most widespread species followed by lookdown dory, sea perch and pale ghost shark. Ribaldo, hake, lucifer dogfish and capro dory also occur frequently but are caught in smaller amounts. Catch rates are highest for hoki and ling, with shovelnose dogfish and spiny dogfish occasionally caught in large amounts.

12. Challenger Plateau 2: occupies deep upper slope waters (836 m) between latitudes 33°S to 43°S. It occurs around the entire North Island and as a thin strip along the western coast of the South Island, and in offshore waters is extensive on the Challenger Plateau and on the northern and eastern Chatham Rise, with scattered occurrences along the Kermadec and West Norfolk Ridges. Moderately cold temperatures, low salinity and gently sloping bathymetry typify the environment. Ribaldo, hoki, javelin fish, spineback and shovelnose dogfish occur most frequently. Orange roughy are occasionally caught in high volumes. Smaller quantities of pale ghost shark, serrulate rattail, spiky oreo and black oreo are also caught.

13. Campbell Plateau 3: occurs in depths of around 720 m on the upper slope in the south to about latitude 54°S. It is extensive throughout the Campbell and Bounty Plateaux, and also occurs in a narrow band to the east and south east of the South Island and along the southern Chatham Rise. Scattered occurrences are also found on the Kermadec and Colville Ridges, around the upper North Island and on the Challenger Plateau. This group is relatively species poor-javelin fish, hoki and pale ghost shark are the most widespread species followed by banded rattail, ribaldo, ling and spineback. Catch rates are highest for javelin fish and hoki. Black oreo and small-scaled brown slickheads, are occasionally caught in large quantities.

3.4.4 Mid slope groups

14. Northern Mid Slope: occupies the mid slope occurring in depths around 1250 m around the North Island and the western and southern South Island, also occurring extensively on the Challenger Plateau in the northwest, and along the eastern and northern Chatham Rise. Small areas also occur around the Bounty Plateau and Bollons seamount in the southeast, and along the western edge of the Campbell Plateau and Macquarie Ridge in the south-west. This group is relatively species poor-basketwork eels are the most widespread species, followed by Johnson's cod, orange roughy, violet cod, longnose deep-sea skate and serrulate rattail. Catch rates are highest for orange roughy but smooth oreo are occasionally caught in large amounts.

15. Southern Mid Slope: occurs at depths of around 1020 m to the east and south of the South Island, extending south to latitude 55°S. It occurs in a band along the southern Chatham Rise, off the east coast of the South Island and in extensive areas around the margins of the Campbell and Bounty Plateaux. Basketwork eels are the most widespread species followed by kaiyomaru rattails, smooth oreo, baxters dogfish, small-scaled brown slickheads and violet cod. Catch rates are highest for smooth oreo. Ridge scaled rattails, Johnson's cod, and four-rayed rattails occur frequently but in small amounts. Black oreo and orange roughy are occasionally caught in large amounts.

16. Deep Mid Slope: this is by far the most extensive group, and occurs in the deepest fishable waters (average = 1540 m) where conditions are typified by very cold temperatures and low salinity. It occurs throughout the New Zealand region, but with more extensive areas around the upper North Island and to the east and south of the South Island. This group supports a relatively depauperate fish fauna-violet cod and basketwork eel are the most widespread species, followed by small-scaled brown slickheads, ridge scaled rattails, Johnson's cod and big-scaled brown slickheads. Smooth oreo are occasionally caught in large quantities, while lighthouse fish, spineback and Baxters dogfish are caught frequently in small volumes.

4. Discussion

Results from this analysis provide a wealth of information about the fish assemblages occurring in New Zealand's Exclusive Economic Zone, including descriptions of their species composition and geographic distribution, and the environmental conditions under which they occur. Sitting behind this is a further collection of spatial data layers describing the distributions of 122 individual fish species in terms of both their probability of catch and their expected catch in kilograms per standardized trawl. Associated model results describe the preferences of these species with respect to a functionally-relevant set of environmental predictors, as well as describing variation in their catchability under varying trawling conditions and over time.

While some of these results could have been produced using the research trawl records alone (as in Francis et al. 2002), results from this study demonstrate that the production of spatially comprehensive predictions, including for sites not directly sampled by trawling, is very feasible given recent advances in statistical and environmental modelling. In addition, the approach we used here has the advantage of building up a description of community patterns based on analysis of the distributions of individual fish species. This gives greater insight into the ecology of the individual species than a community-level analysis, particularly for those species that show unusual or idiosyncratic responses to particular environmental conditions. Such insights would have been largely obscured had we simply classified the research trawl records into groups of similar composition and then interpolated these results across the trawlable parts of the EEZ.

Given the reliance on statistical modelling in this study, we paid particular attention to the robust assessment of predictive error. Rather than assessing model performance using statistics that compare model predictions against the data used to define the model, we used both cross-validation and an independent set of samples for all assessments. This avoids the optimism inherent in performance statistics when they are computed on the training data (e.g., Leathwick et al in press). The close concordance we demonstrate in this study between the cross-validation performance estimates and those calculated using a large set of independent data provides a satisfying demonstration of the robustness of cross-validation as a tool both for setting model complexity and for assessing performance.

In many respects, the work described here has only explored a little of the potential of this type of study - our project was run under relatively tight financial and time constraints, and with strongly focussed objectives. Our overall conclusion from this work is that information-based approaches such as we used here have considerable potential to contribute to the robust management of New Zealand's marine biodiversity. Our analysis of the relative performance of fish-tuned environmental classifications versus the fish community classification described here is that the latter provides the most robust available classification for general conservation management of demersal fish at EEZ scales (Leathwick et al. 2006a). In addition, our subsequent work in which we analysed the individual fish distribution layers with reserve selection software, indicates a considerable potential to develop a robust, informed approach to the identification of a representative network of marine protected areas (MPAs) for New Zealand's Exclusive Economic Zone (Leathwick et al. 2006b).

5. Acknowledgements

Neil Bagley facilitated access to the fish trawl data, the collection of which was funded by the Ministry of Fisheries. Grooming of the *fish_comm* data was carried out under Foundation for Research Science and Technology project CO1830. Jane Elith from the University of Melbourne provided substantial input into the statistical routines used to fit the boosted regression tree models, the implementation of the split sample approach, and the subsequent evaluation of model performance. Paul Taylor provided comment and advice on the analysis of fish distributions throughout the project.

Continue to next file: MCU2a