# Development of a marine environmental classification optimised for demersal fish

NIWA Client Report: HAM2006-063
June 2006

NIWA Project: DOC06213

# Development of a marine environmental classification optimised for demersal fish

John Leathwick
Katie Dey
Kathryn Julian

*Prepared for*

Department of Conservation

NIWA Client Report: HAM2006-063
June 2006

NIWA Project: DOC06213

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:*

for T. Snelder

*Approved for release by:*

S. Elliott

*Formatting checked*

A. Bartley

# 1. Introduction

In contracts carried out over the last five years, NIWA staff have developed a classification of the marine environments around New Zealand, i.e., the Marine Environment Classification or MEC (Leathwick et al. 2004, Snelder et al. 2004), for use as a framework for biodiversity management. Parallel classifications are also available for terrestrial and freshwater environments.

The environmental classification of terrestrial environments (LENZ) was defined using environmental predictors chosen subjectively for their known relevance to the distributions of biota. However, during the development of marine and fresh-water classifications, we demonstrated that a tuning procedure built around the Mantel test (Mantel 1967) can improve classification strength by indicating optimal combinations of weighting and transformation of variables (Snelder et al. 2004, Snelder et al. 2005). In this context, classification strength refers to the ability of a classification to group together sites having similar biological composition.

In the existing marine classification, we applied this tuning approach using data from three biological groups (fish, chlorophyll-a, benthic invertebrates) to derive a product that would hopefully have broad relevance across a range of functional groups. In this project we explore the potential to develop a marine environments classification in which the weighting and transformation of variables is specifically tuned with respect to the distributions of demersal fish. In addition, we test the contribution of four additional variables that were not available when the existing marine classification was developed. Two of these describe the temperature and salinity at the sea floor, while the remaining two provide estimates of suspended particulate matter and coloured dissolved organic matter. Two other variables also developed for use in marine environmental classifications (sea-bed light-levels and calcium compensation depth) were not tested after initial assessment indicated that they had only negligible correlation with fish distributions.

# 2. Methods

## 2.1 Input variables

Input variables used in this analysis consisted of a subset of those previously used in the development of the Marine Environments Classification, along with four new variables (Table 1). Variables used previously in the MEC but omitted here consisted of those characterising conditions at the sea surface in a manner unlikely to be relevant directly to bottom-dwelling fish, i.e., winter and annual solar radiation, wintertime sea surface temperature, and the annual amplitude of sea surface temperature.

Table 1: Variables evaluated for inclusion in a fish-oriented environmental classification for New Zealand's Exclusive Economic Zone. New variables are indicated with an asterisk.

| Variable | Details |
|---|---|
| Depth | Average depth (m) |
| Slope | Sea-floor slope (°) |
| SST_grad | Sea surface temperature spatial gradient (°C km$^{-1}$) |
| SST_anom | Sea surface temperature anomaly (°C) |
| Tidalcurr | Depth averaged tidal current (m s$^{-1}$) |
| Orbvel | Wave-induced orbital velocity at the sea floor (m s$^{-1}$) |
| Chla_c2 | Chlorophyll-a concentration (ppm) |
| Temperature * | Estimated temperature at the sea floor (°C) |
| Salinity * | Estimated salinity at the sea floor (psu) |
| DisOrgMat * | Dissolved organic matter (dimensionless index) |
| SusParMat * | Suspended particulate matter (approximate to g m$^{-3}$) |

Two of the new variables describe the temperature and salinity at the sea-floor. Temperature estimates were the first to be derived, and this was done by combining quarter degree resolution data stored in the World Ocean Atlas 2001 with a 1 km resolution bathymetry layer to generate digital maps (Pinkerton et al. 2005). Salinity estimates were produced subsequently using the same procedure, prompted by the significant contribution of salinity in a separate analysis of demersal fish species richness (Leathwick et al in press).

The remaining two additional variables describe dissolved organic matter and suspended particulate matter. Both were derived from remote sensed data analysed using a case-2 algorithm designed for inshore waters and calibrated at three sites around the New Zealand coast (Pinkerton & Richardson 2005). Six years (1999-2004) of satellite imagery of the waters around New Zealand by the NASA SeaWiFS sensor were used in the production of these layers. Estimates of dissolved organic matter are expected to give a broad indication of terrestrial freshwater inputs, while suspended particulate matter will consist of both material washed in from terrestrial sources and sediments from the sea-floor re-suspended by wave action.

To preserve consistency with the remote sensed variables derived from the case-2 analysis of imagery, the original satellite-based description of chlorophyll-a (case-1) was replaced with a case-2 layer describing chlorophyll-a. This is expected to give better resolution of chlorophyll-a in inshore waters, and is very strongly correlated with the original chlorophyll-a estimates in offshore waters.

## 2.2 Mantel-based testing of variables

Selection of variables, along with an assessment of their ideal degree of transformation and weighting, was carried out using a procedure based on the Mantel-test. This approach was originally developed in preparing the MEC, and was subsequently refined in the preparation of a fresh-water classification for New Zealand (Snelder et al. 2005). In brief, this procedure uses a set of test sites at which both biological and environmental data are available. Biological differences between these sites are described using a standard biological distance measure (Bray-Curtis - Legendre & Legendre 1998). Environmental distances between these test sites were then computed using various combinations of environmental variables, with the objective of finding the combination of environmental factors that maximises the correlation between the parallel sets of environmental and biological distances.

The biological dataset consisted of trawl catch data (weight in kg) for 127 fish species at 17101 sites, this data being drawn from the *FishComm* database, as described by Francis et al. (2002). All catch data were site standardized to allow for inconsistencies in catch size. Because of the memory requirements of the testing process, the trawl dataset was divided into 11 random subsets of 1500 sites for analysis. Biological distances were calculated using the Bray-Curtis distance measure, and any distances greater than 0.95 were re-estimated using a flexible shortest path adjustment method (details are provided in Snelder et al. 2005).

The environmental data consisted of 11 environmental variables at every site. Five transformations were considered for each variable: square, untransformed, square root, fourth root and log. At each step in the testing process, a set of environmental variables and transformations was used to calculate environmental distances between sites using the Gower metric (Gower 1971), a distance measure that standardizes the range of the input variables so that all have an equal contribution to the final distance estimates.

The testing process started by finding the single variable and transformation that gave the highest correlation between environmental and biological distances for the test sites. At each subsequent step, all remaining possible combinations of variable and transformation were tested to identify that which gave the greatest improvement in correlation between the environmental and biological distances. After every step, any variables with negative fitted coefficients were dropped, as variables cannot be included in the final classification with a negative contribution. The process continued until no more variables could be added without introducing negative coefficients. This process was repeated for all eleven subsets, and results were used to identify the most frequently occurring variables and transformations, of which the 15 most common are shown in Table 2.

A backwards-stepwise elimination process was then used to establish the relative importance of these 15 combinations of environmental variable and transformation (Table 2). This was achieved by fitting all fifteen combinations separately on each of the 11 data subsets, and progressively eliminating the variables that made the least contribution. We then eliminated from all models the variable that made the least contribution, averaged across all the subsets.

Table 2: Environmental variables and their transformations identified as important correlates of fish distribution from the Mantel-based testing procedure. Variables are ranked in order of decreasing importance.

| **Variable name** | **Transformation** |
|---|---|
| Avg depth | z |
| New temp | z.^2 |
| Avg depth | sqrt(z) |
| New salinity | z.^2 |
| Tidalcurr | z.^(1/4) |
| Sst anom | log10(z+1) |
| New temp | log10(z+1) |
| MEC sstgrad | z.^(1/4) |
| DisOrgMat | z.^2 |
| Tidalcurr | z ^2 |
| Tidalcurr | sqrt(z) |
| Sst anom | z |
| MEC sstgrad | z.^2 |
| Orbvel | z.^2 |
| Slope | z.^2 |

While the relative importance of variables identified by this process (Table 2) can be used as a guide to the order with which variables should be added to the final classification, we still needed to assess robustly both the optimal number of variables to use, and their weightings. This was achieved by fitting a progressively more complex model to each subset of data, starting with the most important variable in Table 2, and successively adding in variables and assessing the improvement in correlation between the biological and environmental distances. This procedure continued until no further variables could be added without the fitting of negative coefficients. Coefficients from these models were then averaged and used to calculate the weights to be applied in the final classification.

Results from this last stage of the testing indicated that there were marked improvements in correlation between environmental and biological distances with the addition of the first two variables (depth and temperature). However, subsequent gains in correlation were relatively small, and no more than six combinations of variable and transformation could be added without fitting negative coefficients. This suggested that we needed to test two classifications, with two and six variables respectively, and with weights defined from the coefficients fitted in the testing process.

To assess the adequacy of the testing process, we also defined a classification in which we included 10 variables used in separate analyses of the distributions of 122 fish species with respect to environment (Leathwick et al. 2006). Weightings of variables for this classification were based on their relative contributions to prediction outcomes for the fish species. A transformation was applied to only one variable, i.e., values for the variable describing orbital velocities were very strongly skewed, and this was reduced to a degree by applying a log10 transformation.

## 2.3 Defining the classification

All three classifications, i.e., using 2, 6, and 10 variables, were defined using PATN (Belbin 1995), multivariate classification software developed at CSIRO, Canberra. In all three cases, the module ALOC was used to define an initial non-hierarchical classification in which around 100 groups were defined using the Gower metric. A conventional agglomerative classification was then performed on the resulting group averages using the PATN module FUSE. Results were exported from PATN and imported into ArcView where they were combined with a text file indicating the . progressive fusing of groups. This allowed the display of the spatial distribution of groups at varying levels of classification detail.

One difficult issue that remains unresolved is that the trawl data only sample waters with depths down to 1950 m, whereas parts of the New Zealand's exclusive economic zone (EEZ) exceed 10,000 m in depth. As a consequence, the weightings and transformations identified from the Mantel based testing can only be applied robustly across approximately 47% of the total area of the EEZ. That is they have no relevance to deeper waters, and applying weightings and transformations defined from an environmental subset is likely to give potentially misleading results, particularly given the five-fold extrapolation of depth required for deeper parts of the EEZ. At this stage, we therefore restricted our classifications to waters with depths less than or equal to 1950 m.

## 2.4 Evaluating classification strength

As the primary objective of this analysis was to define a classification giving maximum discrimination of variation in fish composition, each of the three classifications was assessed using a statistic (c) (Van Sickle 1997), computed as:

$$c = \frac{b - w}{b}$$

where b is the mean biological distance between sites in different classification groups, and w is the mean biological distance between sites in the same classification groups. This provides a measure of the degree to which biological samples (trawls) falling within the same environmental group are more similar to each other than to biological samples located within other environmental groups. This test can be applied at varying levels of classification detail, with results plotted as a function of the number of classification groups.

# 3. Results

Three classifications were tested using two, six, and ten variables as described above. As a comparison, we also assessed the classification strength of the original, general-purpose MEC, and a multivariate classification of the predicted distributions of 122 fish species, i.e., that groups together sites having similar predicted fish composition (Leathwick, Julian & Francis 2006). The predicted layers used as an input to this classification were produced from boosted regression tree models relating fish presence/absence and catch to environment, species by species.

Test results from ANOSIM for all four classifications indicate that there is a progressive increase in classification strength as the number of classification groups increases. This is because reduction in the within-group variation as the number of classification groups is increased is matched by a parallel decrease in the variation in biological composition of the trawl samples occurring in each of the groups.

However, despite this similarity there is also a clear difference in the performance of the different classifications at different levels of classification detail. Generally, the use of a small number of variables gives superior classification strength at low numbers of groups, but the classifications defined using more variables outperformed the two-variable classification at higher classification levels. This is evident in Fig. 1 where the six-variable classification provides superior performance to its two-variable counter-part when compared at the 40-group level, but both of these classifications are outperformed by the 10-variable classification above a 60-group level of classification.

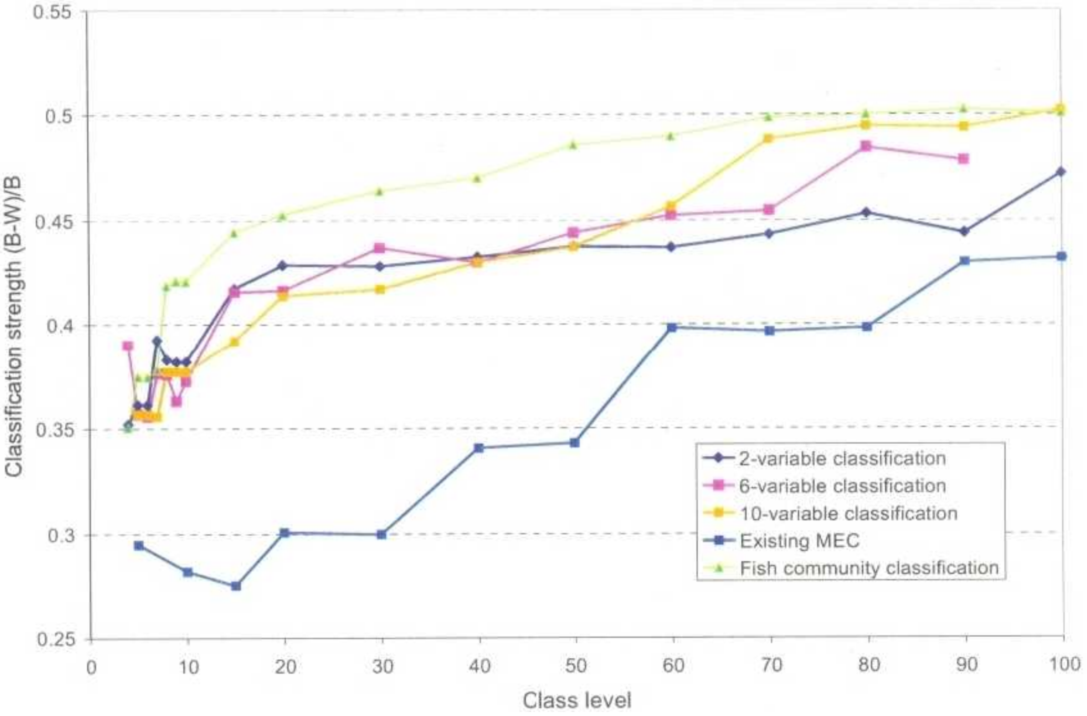


Figure 1: Classification strength of different classifications as a function of the number of classification groups. Increasing values along the vertical axis indicate increasing discrimination between biological samples in different classification groups. The environmental classifications are defined using 2, 6 or 10 variables, while the biological classification is based on predicted fish distributions for 122 species. The existing marine environment classification (MEC) is also shown for comparison.

Perhaps not surprisingly, the fish community classification was clearly superior to the environmental classifications at all levels of detail above about six groups, except at a 100 group level, where the 10-variable environmental classification had similar classification strength. All the new classifications out-performed the existing Marine Environment Classification, particularly at lower levels of classification detail.

### 3.1 The ten-variable classification

Given the results described above, and the superior strength of the fish-community classification in particular, we present only a broad-scale description of classification results for the best-performing environmental classification, i.e., that defined using ten environment variables weighted according to their contribution to the fish distribution models. We expect that in most circumstances, the fish community classification would be the preferred option where a spatial classification is required for managing issues related to demersal fish.

The classification we describe here was summarised at a ten-group level of detail for all of the EEZ down to the maximum depth recorded in the research trawl database (1950 m). To provide coverage in deeper waters, we also classified one km grid data across sites with depths greater than 1950 m using the same variables and weightings as were applied in the trawlable waters, and merged the two classifications together. This has an artificial disjunction at 1950 m - in waters down to this depth, variable selection is fully supported by species distribution data held in the Research Trawl database. In deeper waters, for which we lack trawl distribution data we can only assume that variables have similar importance as they do in shallower waters.

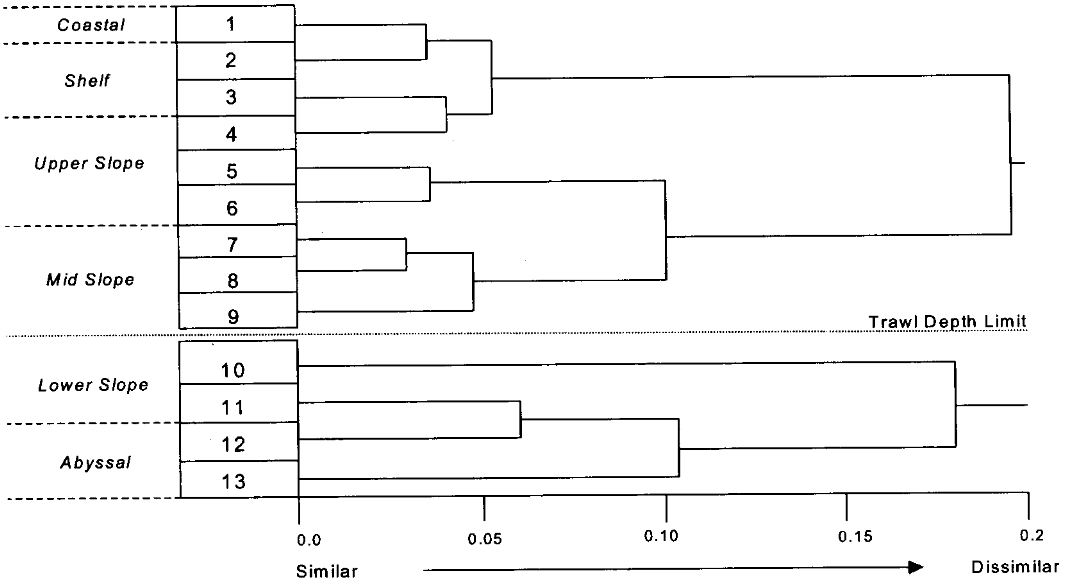


Figure 2: Dendrogram showing environmental relationships between thirteen environmental classification groups. Groups 1-9 within trawl depths limits (<1950 m) and groups 10-13 below trawl depth limits (>1950 m).

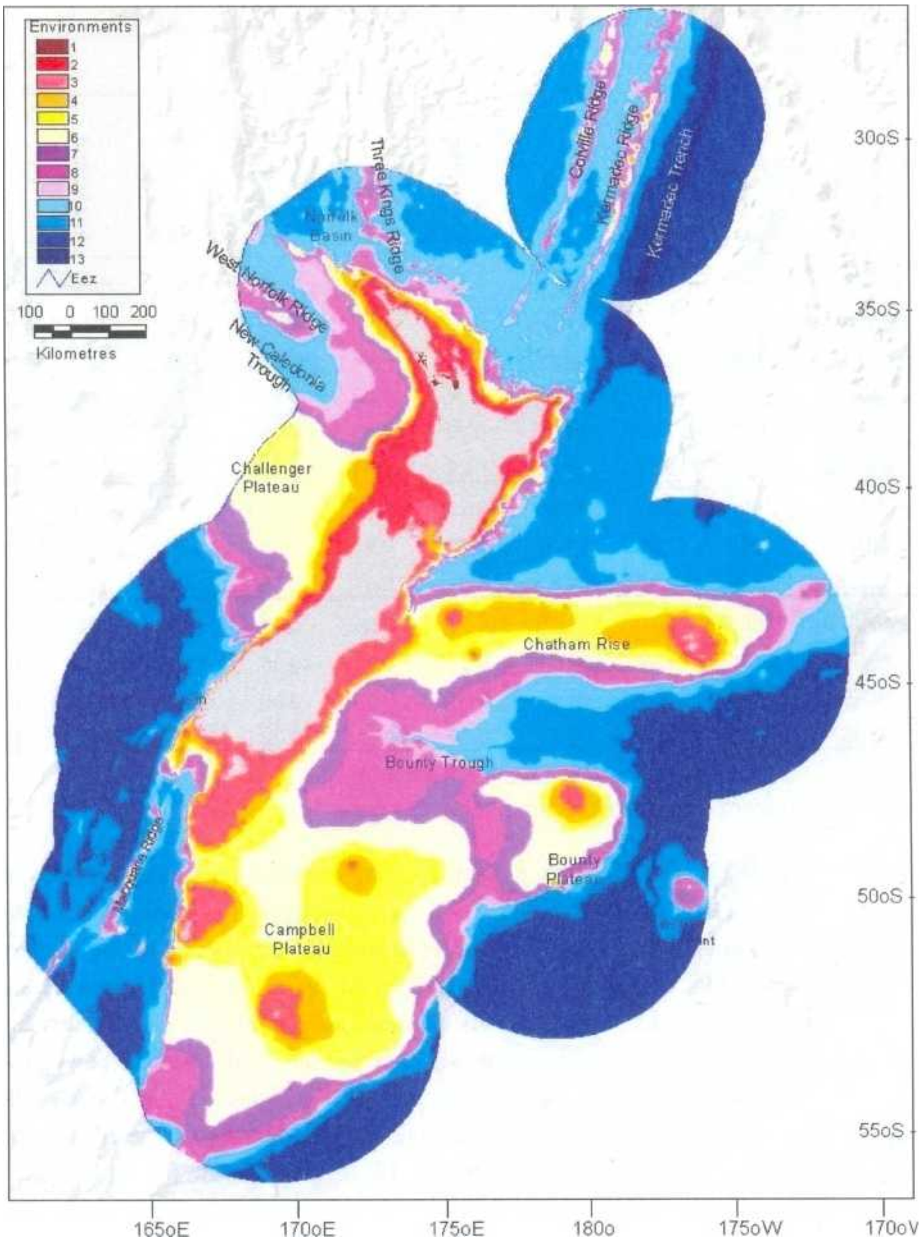
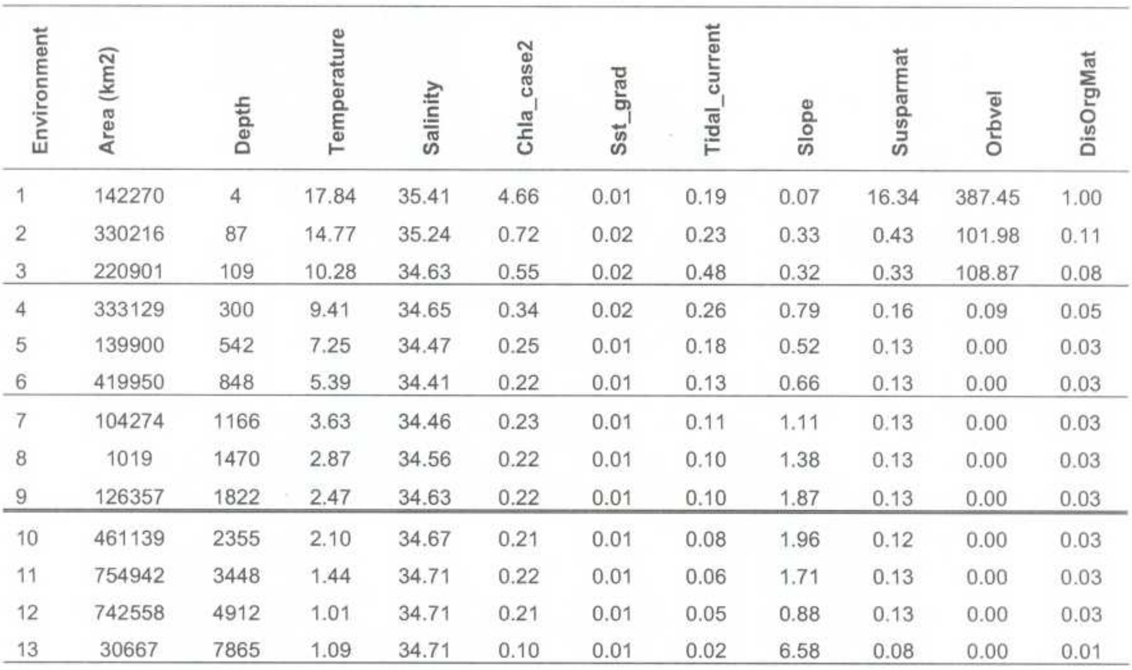


Figure 3: Distribution of thirteen environmental groups defined for New Zealand's EEZ using ten environmental variables, weighted according to their contribution to regression models relating the occurrence of 122 fish species to environment.

Table 3: Environmental attributes of the thirteen environments defined by multivariate classification.



| Environment | Area (km2) | Depth | Temperature | Salinity | Chla_case2 | Sst_grad | Tidal_current | Slope | Susparmat | Orbvel | DisOrgMat |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 142270 | 4 | 17.84 | 35.41 | 4.66 | 0.01 | 0.19 | 0.07 | 16.34 | 387.45 | 1.00 |
| 2 | 330216 | 87 | 14.77 | 35.24 | 0.72 | 0.02 | 0.23 | 0.33 | 0.43 | 101.98 | 0.11 |
| 3 | 220901 | 109 | 10.28 | 34.63 | 0.55 | 0.02 | 0.48 | 0.32 | 0.33 | 108.87 | 0.08 |
| 4 | 333129 | 300 | 9.41 | 34.65 | 0.34 | 0.02 | 0.26 | 0.79 | 0.16 | 0.09 | 0.05 |
| 5 | 139900 | 542 | 7.25 | 34.47 | 0.25 | 0.01 | 0.18 | 0.52 | 0.13 | 0.00 | 0.03 |
| 6 | 419950 | 848 | 5.39 | 34.41 | 0.22 | 0.01 | 0.13 | 0.66 | 0.13 | 0.00 | 0.03 |
| 7 | 104274 | 1166 | 3.63 | 34.46 | 0.23 | 0.01 | 0.11 | 1.11 | 0.13 | 0.00 | 0.03 |
| 8 | 1019 | 1470 | 2.87 | 34.56 | 0.22 | 0.01 | 0.10 | 1.38 | 0.13 | 0.00 | 0.03 |
| 9 | 126357 | 1822 | 2.47 | 34.63 | 0.22 | 0.01 | 0.10 | 1.87 | 0.13 | 0.00 | 0.03 |
| 10 | 461139 | 2355 | 2.10 | 34.67 | 0.21 | 0.01 | 0.08 | 1.96 | 0.12 | 0.00 | 0.03 |
| 11 | 754942 | 3448 | 1.44 | 34.71 | 0.22 | 0.01 | 0.06 | 1.71 | 0.13 | 0.00 | 0.03 |
| 12 | 742558 | 4912 | 1.01 | 34.71 | 0.21 | 0.01 | 0.05 | 0.88 | 0.13 | 0.00 | 0.03 |
| 13 | 30667 | 7865 | 1.09 | 34.71 | 0.10 | 0.01 | 0.02 | 6.58 | 0.08 | 0.00 | 0.01 |

### 3.2 Environment descriptions

At a broad scale (Figs. 2 & 3, Table 3), there is a strong separation between the coastal and shelf environments (groups 1-4), upper[1] and mid slope environments (groups 5-9) that occur further offshore, and environments that occur in the deepest waters inaccessible to trawling (groups 10-13). At more local scales there is also strong separation in the shallow waters between environments 1 and 2, which occur in northern estuaries and sheltered waters and shelf waters respectively, and environments 3 and 4, which occur in deeper continental shelf waters. Similarly, environments 5-9 can be subdivided into groups occurring on the upper slope (5 & 6) and mid slopes (7-9). Environments 10-13 occur below the trawl depth limit (at depths below 1950 m) and are subdivided predominantly by depth, with environments 10 -11 and 12-13 occurring in lower slope and abyssal waters respectively.

#### 3.2.1 Coastal and shelf environments

**Environment 1:** occurs inside the harbours of the central and upper North Island in very high salinity warm waters approximately 5 m in depth, north of about latitude

---

[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.

38°S. Both chlorophyll-a and suspended particulate matter occur in very high concentrations, dissolved organic matter levels are high, indicating significant inputs of terrestrial freshwater, and orbital velocities are high.

**Environment 2:** occupies coastal and shelf waters to depths of 90 m or more between latitudes 34°S and 45°S, mostly surrounding the North Island but also along the west coast of the South Island. This environment is typified by moderately warm water temperatures and high salinity-orbital velocities are often high, chlorophyll-a concentrations are moderate. Tidal currents and sea surface temperature gradients are both moderate.

**Environment 3:** occupies southern continental shelf waters between latitudes 40°S and 54°S, occurring extensively around the South Island (except along the West Coast), the southwest of the North Island, and at shallow depths on the Chatham Rise and Bounty and Campbell Plateaux. The average depth is a little over 100 m, and waters are generally of moderate temperature and low salinity. Tidal currents are very strong, orbital velocities are high, and sea surface temperature gradients are moderate.

## 3.3 Upper slope environments

**Environment 4:** occurs between latitudes 33°S and 54°S, both as a narrow, off-shore band surrounding the North and South Islands, and as more extensive patches on the Challenger and Bounty Plateaux and the Chatham Rise. Depths average 300 m, water temperatures are moderately cold, and salinity is moderate. Both sea surface temperature gradients and tidal currents are moderate.

**Environment 5:** occurs over a similar latitudinal and geographic range to the previous environment, but in deeper waters (mean = 542 m). It forms a narrow strip on the upper slope around much of New Zealand's continental shelf, and also on the Chatham Rise, Campbell Plateau and shallow parts of Challenger Plateau. It is typified by moderately cold water temperatures, low salinity and weak tidal currents.

**Environment 6:** is the most extensive of the three upper-shelf environments, and occurs over the widest latitudinal range (26–54°S). It is widespread at intermediate depths (850 m) along the Chatham Rise, and on the Challenger, Campbell and Bounty Plateaux. It also occurs as a narrow strip off the continental shelf around the North and South Islands, and along the Colville, Kermadec, Three Kings and West Norfolk Ridges in the north, and along the Macquarie Ridge in the south. This environment has generally, cold low salinity waters, with low tidal currents, and low concentrations of chlorophyll-a.

### 3.3.1 Mid-slope environments

**Environment 7:** occurs on the mid slopes at an average depth of 1166 m over a wide geographic range (latitudes 26–55°S), including around the margins of the continental shelf that surrounds the North and South Islands, along the flanks of the Chatham Rise, and around the margins of the Bounty and Campbell Plateaux in the south and Challenger Plateau in the west. More scattered occurrences are found on the uplifted ridges to the north of the North Island, surrounding the Bollons seamount, and along the Macquarie Ridge. Conditions at the sea-floor are cold, with low salinity, and moderately gentle slopes.

**Environment 8:** occupies very cold, low salinity waters with an average depth of 1470 m. Although it occurs across a similar geographic range to the previous environment, it is more geographically extensive, occupying large areas in the gently sloping basins at the heads of the Bounty and New Caledonia troughs. It also occurs as a narrow band around the continental plate and plateau boundaries around the entire New Zealand land-mass, and in scattered locations along the Colville, Kermadec, Three Kings and West Norfolk Ridges.

**Environment 9:** this is the deepest of the mid-slope environments, occurring in very cold, low salinity waters at an average depth of around 1820 m. It forms locally extensive areas in the head of the New Caledonia Trough (north east of the North Island), off the eastern end of the Chatham Rise, and in the Bounty Trough in the southeast. It also occurs as a narrow fringe around the continental shelf and plateau boundaries. Average slopes reach their highest values in this environment.

## 3.4 Environments below the trawlable limit

**Environment 10:** is most extensive to the north of the North Island, including in the New Caledonia Trough and Norfolk Basin, and between the Kermadec and Colville Ridges. Elsewhere it occurs mostly as a narrow band around the New Zealand landmass, although with more extensive areas around the Chatham Rise and Bounty Trough. It occurs at depths of around 2300 m in waters having very cold temperatures and moderate salinity. Average slopes are moderately steep compared to other groups.

**Environment 11:** occurs at average depths of nearly 3500 m, in waters that are very cold and of moderate salinity-slopes are moderately steep. It is most extensive in the far north (west of the Colville Ridge), between the east coast of the North Island and the Chatham Rise, in the Bounty Trough, along both sides of the Macquarie Ridge, and southwest of the Challenger Plateau.

**Environment 12:** occurs in very cold abyssal waters with an average depth of nearly 5000 m. It is extensive to the east of the Kermadec Trench, to the southeast of the Challenger and Bounty Plateaux, and on the Tasman Plain off the southwest coast of the South Island.

**Environment 13:** is the most geographically restricted of these environments, occurring in the deepest waters (average depth = 7865 m) found in New Zealand's Exclusive Economic Zone, i.e., the Kermadec Trench. Very cold temperatures, moderate salinity and very steep slopes typify this environment.

## 4. Discussion and conclusions

One of the most important features of this analysis is its reinforcement of recent lessons learned in the definition of an environmental classification for freshwater environments, and that is the difficulty in defining classifications that will work well across a range of classification detail (Snelder et al in review). In effect, a classification based on a few environmental variables that have a dominant effect at broad spatial scales will perform best at relatively low levels of classification detail, where it will effectively partition the broad-scale environmental features driving variation in biological composition. By contrast, where it is desirable for a classification to be used at a high level of classification detail, this will be best achieved using a classification based on a mix of variables. This should include both those that are important globally, as well as variables that perhaps show strong variation only locally, but which define distinctive environments that become important at finer levels of subdivision.

In this respect, it appears that our Mantel-based test procedure is working best at defining classifications for broad scale application using just a few classification groups. If a classification is required to work at higher levels of classification detail, then other procedures may be more effective guides to variable selection and weighting. Here, the use of the relative importance of variables in regression models relating fish abundance to environment proved to have the highest classification strength from a 60-group level on.

The other important feature to emerge from this analysis is the markedly superior performance of a community classification of predicted fish distributions. This raises important questions about the justification for continued use of an environmental classification for management of conservation issues related to demersal fish, even when this is tuned to this biological group. The rationale for use of environment-based products is generally built around the inadequacy of biological data, but given the richness of trawl survey data for New Zealand's EEZ, our results suggest that the more robust approach would be to use a biological classification, rather than using an environmental surrogate. However, this approach is clearly not going to work well in deeper environments from which biological data are largely lacking. In these waters other factors such as the high solubility of calcium and low availability of oxygen are also likely to become important influence on ecosystem character, and would require consideration.

## 5. Acknowledgements

This work was initiated as a result of discussions between Clinton Duffy (Department of Conservation) and Ton Snelder (NIWA), both of who have supported and encouraged its subsequent development. Matt Pinkerton provided advice on the interpretation of case-2 products describing suspended sediments and dissolved organic matter, Mark Hadfield interpolated the temperature and salinity layers, and Ude Shankar assisted with the re-projection of geographic data layers into a common format.

## 6. References

Belbin, L. (1995). PATN analysis package. Division of Sustainable Ecosystems, CSIRO, Canberra. 237 p.

Clarke, K.R.; Warwick, R.M. (2001). Change in Marine Communities: An Approach to Statistical Analysis and Interpretation. 2. PRIMER-E, Plymouth, UK. 172 p.

Francis, M.P.; Hurst, R.J.; McArdle, B.H.; Bagley, N.W.; Anderson, O.F (2002). New Zealand demersal fish assemblages. *Environmental Biology of Fishes 65*: 215-234.

Gower, J.C. (1971). A general coefficient of similarity and some of its properties. *Biometrics 27*: 857-871.

Leathwick, J.R.; Elith, J.; Francis, M.P.; Hastie, T.; Taylor, P (in press). Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series.*

Leathwick, J.; Julian, K.; Francis, M. (2006). Development of a demersal fish community map of New Zealand's Exclusive Economic Zone. *NIWA Client Report: HAM2006-062*. 21p.

Leathwick, J.; Image, K.; Snelder, T.; Weatherhead, M.; Wild, M. (2004). Definition and test of Marine Environment Classifications of New Zealand's Exclusive Economic Zone and the Hauraki Gulf. NIWA *Client Report CHC2004-085*. 64 pp.

Legendre, P. & Legendre, L. (1998). Numerical ecology. Elsevier, Amsterdam, The Netherlands. 853 pp.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research 27*: 209-220.

Pinkerton, M.H.; Richardson, K.R. (2005). Case 2 Climatology of New Zealand: Final report. *NIWA Client Report WLG2005-49*. 16 pp.

Pinkerton, M.; Hadfield, M.; Hunter, K.; Macaskill, B. (2005). Three additional layers for Marine Environment Classification: (1) Sea bed temperature; (2) Calcium compensation depth; (3) Sea-bed light levels. *NIWA Client Report WLG2005-50*. 40 pp.

Snelder, T.; Dey, K.; Leathwick, J.R. (2005). Definition of the multivariate environmental river classification - "Freshwater Environments of New Zealand". *NIWA Client Report CHC2005-49.* 86 pp.

Snelder, T.; Dey, K.; Leathwick, J.R. (in review). Sensitivity of multivariate environmental classification performance to choice of input variables.

Snelder, T.; Leathwick, J.; Image, K.; Weatherhead, M.; Wild, M. (2004). The New Zealand Marine Environment Classification. *NIWA Client Report CHC2004-071.* 86 p.

Van Sickle, J. (1997). Using mean similarity dendrograms to evaluate classifications. *Journal of Agricultural, Biological and Environmental Statistics 2:* 370-388.