Ministry for Primary Industries
Manatū Ahu Matua

Evaluation of tagging programme designs for SNA 1 and SNA 8

New Zealand Fisheries Assessment Report 2015/35

J.R. McKenzie
S.D. Hoyle
R. Bian
D. Parsons
A. Dunn
C. Williams

ISSN 1179-5352 (online)
ISBN 978-0-908334-71-1 (online)

June 2015

New Zealand Government

Growing and Protecting New Zealand

Requests for further copies should be directed to:

Publications Logistics Officer
Ministry for Primary Industries
PO Box 2526
WELLINGTON 6140

Email: brand@mpi.govt.nz
Telephone: 0800 00 83 33
Facsimile: 04-894 0300

This publication is also available on the Ministry for Primary Industries websites at:
http://www.mpi.govt.nz/news-resources/publications.aspx
http://fs.fish.govt.nz go to Document library/Research reports

Table of Contents

EXECUTIVE SUMMARY

McKenzie, J.R.; Hoyle, S.D.; Bian, R.; Parsons, D.; Dunn. A.; Williams, C. (2015). Evaluation of tagging programme designs for SNA 1 and SNA 8.

***New Zealand Fisheries Assessment Report 2015/35.* 80 p.**

Snapper (*Pagrus auratus*) is New Zealand's most valuable commercial coastal marine species and, by virtue of its high abundance around the populous regions of northern New Zealand, it is also the nation's most important recreational species.

A range of tagging technologies were evaluated for estimating snapper biomass, including internal transponder tags (PIT), coded wire tags (CWT), genetic tagging and photo identification. PIT and CWT tagging technologies had been used successfully in past SNA 1 and SNA 8 tagging programmes. PIT was found to be the better option because the technology was more amenable to bulk scanning. Genetic tagging has a strong advantage over PIT and CWT in that it is technically feasible to obtain tissue samples from snapper at-depth, thus eliminating barotrauma related tag losses (due to mortality associated with bringing the fish up from depth). However, a difficulty with genetic tagging is that the methodology has not been established for snapper and would require two to three years to fully develop. In addition, the costs of a genetic snapper tagging programme are likely to be much larger than for comparable PIT or CWT programmes. Photo-tagging would also require methodological development for use with snapper, and is likely to require a similar lead-in time to develop. The advice from experts in the field of photographic identification and image analysis was that photo imaging would still require bringing fish to the surface to ensure adequate identification. Hence, the method would not be likely to offer any advantage over PIT or CWT if developed.

Computer simulations were used to investigate possible trade-offs between different tag-release and recapture options. These suggested that a stratified random release design, with recaptures from the commercial fishery as a part of the normal operations, was likely to provide the best trade-off between cost and bias for both SNA 1 and SNA 8. A stratified release design, although not precluding the use of the commercial fishery to release tags in some areas, was likely to require the use of research or charter vessels. Simulations suggested that a tagging programme that relied solely on the commercial fishery to release and recover tags in proportion to the typical location of effort and catch would be likely to produce negatively biased biomass estimates of 25% or greater because the commercial fishery is likely to release and hence recover a disproportionately higher number of tags in high catch-rate areas. As it is unlikely to be practical to recover sufficient tags from SNA 1 or SNA 8 without scanning commercial catches (i.e., random recovery), to avoid bias, it will be necessary to release tags in the population relative to its spatial distribution (i.e., random release).

Computer simulations were also used to evaluate the statistical power offered by sequential tagging estimates of varying precision to track population biomass over time. A time series of tagging estimates of abundance were shown to have a high utility in estimating abundance change when change is consistently (predictably) increasing or decreasing. In such scenarios the time series can be used to predict the magnitude and direction of change. However, such time series of tagging estimates of abundance were shown to have little additional power over that from an individual estimate to predict the overall pattern of change, under circumstances where the population was fluctuating through time. Simulations therefore focused on the statistical power and cost of obtaining a "point-in-time" tagging estimate of abundance for time periods where the abundance of SNA 1 and 8 could be assumed not to have changed significantly — for example, intervals of two of so years.

Two options were considered for deriving "point-in-time" tag estimates of abundance of a given precision;

1. a single release event programme, or
2. a series of three annual lower precision release event programmes (labelled "triplet" in the text below).

The triplet release event programme would be conducted in three sequential years but when the results were combined together, would give rise to a single estimate of abundance, with the same overall precision as the single release programme.

There are some advantages to the triplet tagging approach. First, a better understanding of seasonal movement patterns would be obtained from repeated observations of fish movement, as tagged fish from the first and second release events would be observed over two or more annual cycles. Second, there would be greater opportunity to adaptively "correct" and adjust release and recovery designs for biases seen in the recoveries from the first and subsequent release events. Third, the costs associated with the programme could be spread out over a greater number of years. However, there are also some disadvantages, namely that there would be a delay in delivery of the combined "higher precision" result (about 4–5 years from the onset of the study compared with 2–3 for a single release programme), and there would be additional costs from running the programme over more than one year. A gross estimate of the additional cost was that a triplet study would be 10–30% more expensive than a single release study - although this is highly dependent on the actual level of savings that could be achieved in second and subsequent years from standardising and streamlining methods and processes.

In their cost-benefit evaluation of the numbers of tagged fish released and recapture tags required, McKenzie et al. (2011) showed that the likely highest practically achievable precision for a SNA 1 tag abundance estimate was CV = 0.1, equating to about 1500 tag recoveries. By comparison, the 1994 SNA 1 tagging programme achieved 530 tag recoveries and had an overall precision of CV = 0.2. It is possible to gain some appreciation on the relative value of two programmes by assessing their relative statistical power to resolve some of the key uncertainties in the 2013 SNA 1 assessment. Estimates of stock-status from the 2013 assessment had relatively high precision, with CV in the order of 0.1. If the 1994 tagging programme had achieved a higher precision than it did, for example a CV of 0.1, then it is reasonable to conclude that it was unlikely to have resulted in an appreciably improved precision of stock status in 2013 assessment. If this is the case, than there is no strong justification for choosing a higher precision design with the intention of increasing the precision on the overall assessment.

The main contributor to the uncertainty of the 2013 SNA 1 assessment was, however, not a lack of precision from estimates of abundance from the tagging study, but rather the model structural uncertainty. Specifically, the spatial extent of the three SNA 1 sub-stocks and the degree of mixing between them was likely to be the major uncertainty in the assessment outcome. This uncertainty is likely to lead to bias in the assessment model estimates of abundance. A tagging design with a CV of 0.1 is likely to yield three times the number of tag recoveries than the 1994 SNA 1 tagging programme and hence have much greater power to estimate the degree of spatial mixing.

1 INTRODUCTION

Mark-recapture (tag) methods can be used to estimate a range of important fish population parameters (Seber 1982; Quinn & Deriso 1999, Cooch & White 2013) including recruited stock numbers/biomass, area mixing rates (movement), gear selectivity, growth, and survivorship (total mortality *Z*).

Snapper are reasonably amenable to mark-recapture studies because: the species has a predominately shallow coastal distribution, meaning the majority of the population is accessible for tagging; a high proportion of fish survive the catch and release process; and a large commercial fishery makes it feasible to both tag large numbers over a spatially extended area, and subsequently scan large numbers of fish for released tags.

SNA 1 and SNA 8 are two of New Zealand's most valuable inshore finfish stocks. According to the Medium Term Research Plan for Inshore Finfish, the absolute biomass for SNA 1 and SNA 8 are estimated on a 10–15 year cycle using mark-recapture studies. Past SNA 1 and SNA 8 tagging programmes have been logistically challenging and relatively expensive to implement. As a consequence, there have been fewer tagging programmes for snapper than had been planned — the last SNA 1 tagging study occurred in the 1993–94 fishing year (McKenzie & Davies 1996) and the last SNA 8 programme took place in the 2001–02 fishing year (Davies et al. 2006).

Age structured population models have been used to assess the SNA 1 and SNA 8 snapper stocks since the mid-1990s. These models have become more complex over time, reflecting increased knowledge on the spatial complexity of northern snapper stocks and the need to account for the increasing amount of observational data (e.g., CPUE, age composition of the catch and tagging data). The spatially disaggregated age-structured model used in the 2013 SNA 1 assessment (Francis & McKenzie, in press) is the most complex snapper assessment model to date. A key point of difference between the Francis & McKenzie model and the earlier snapper assessment models is that the tag-release and tag-recapture observations were directly fitted in the model. In previous assessments, the tagging data were analysed outside the model with the resulting estimates of stock biomass and gear selectivity then provided to the model. Fitting the tagging data directly, as in the recent assessment, means that these data can be more appropriately weighted against other observational data and sub-stock movement can be estimated.

Despite the SNA 1 tagging data from the two previous tagging studies being 20 and 30 years old respectively, these data still strongly influenced the 2013 assessment results. If more recent tagging data had been available for the 2013 assessment, our understanding of current stock status is likely to have been more certain.

However, in a stock assessment, the use of tagging estimates of abundance require a number of assumptions (Seber 1982), and failure to meet or appropriately test for these can result in bias (McKenzie et al. 2011; Welsford & Ziegler 2012; Parker & Mormede 2012, Thorsteinsson 2002). Experience gained from past SNA 1 and SNA 8 tagging programmes should ensure that future programmes are better able to either mitigate bias in their designs or enable bias to be estimated and corrected for in the analysis.

Conventional tagging programmes for SNA 1 and SNA 8 are expensive. The purpose of this project was to investigate the relative costs of alternative tagging approaches, including mark-recapture methods and survey designs.

1.1 Project objectives and scope

This report presents results of an evaluation of design options for SNA 1 and SNA 8 tagging programmes intended to provide estimates of absolute biomass, movement, and selectivity for input to age structured stock assessment models. This report is the final deliverable on MPI project SNA201303 which had the following objectives:

Overall Research Objective:

1. To evaluate designs for tagging programmes for SNA 1 and SNA 8.

Specific Research Objectives:

1. To develop and evaluate alternative designs for tagging programmes for SNA 1.

2. To develop and evaluate alternative designs for tagging programmes for SNA 8.

The report draws heavily on work related to bias estimation and design optimisation for past SNA 1 and SNA 8 tagging programmes; and summarises many key results and conclusions from these investigations.

Because more historical tagging data exists for SNA 1, most of the design optimisation and cost benefit analysis in the report is specifically for SNA 1. However, many of the SNA 1 tag design recommendations also apply to SNA 8.

1.2 Lincoln-Petersen estimator specific assumptions

A key metric derived from tagging data is the ratio of marked to unmarked animals in the underlying population (see also Appendix 1).

$$\text{Population mark ratio} = \frac{Number\ of\ fish\ in\ population\ [N]}{total\ number\ of\ tagged\ fish\ [n_1]}$$

Assuming that tags are homogeneously mixed in the population, then samples of fish from the population (n_2) will contain an average of m_2 tagged fish such that:

$$\frac{N}{n_1} = \frac{n_2}{m_2} \qquad \textbf{Equation 1}$$

By rearranging the terms in Equation 1 it is possible to derive an estimate for the total population (N) (Equation 2), commonly called the Lincoln-Petersen estimate:

$$\widehat{N} = \frac{n_1 n_2}{m_2} \quad \textbf{Equation 2}$$

Some basic assumptions have to hold for Equation 2 to be an unbiased estimates of population size (N):

1. Population is closed (no immigration or emigration);
2. Fish do not lose their marks (no tag loss occurs);
3. All recaptured tagged fish are recognised (no failures in tag detection);
4. Tagged and untagged animals behave in the same way (i.e., a tagged animal behaves the same as an untagged animal, so that they are equally likely to be captured);
5. Tagged and untagged fish are either homogeneously mixed (Seber 1982) or randomly sampled so that the probability of recapture of a tagged animal remains the same no matter the location or timing of sampling.

These assumptions would need to be addressed in a tag release and recapture design, as violations would lead to bias in resulting abundance estimates.

In the inverted form the population mark ratio (Equation 1) is more often referred to as the population "mark rate" (Equation 3).

$$\frac{n_1}{N} = \frac{m_2}{n_2} \quad \textbf{Equation 3}$$

For previous SNA 1 and SNA 8 tagging programmes, catch scanning for tags has taken place over the entire stock area, typically for between 12 to 18 months after the initial tag release, and it has been assumed that the tags were homogeneously mixed.

If the homogeneity assumptions are violated, then biases in abundance estimates may be introduced. There are two options for obtaining an unbiased estimate of the overall population mark-rate — either by ensuring that the tags are released into the population randomly in proportion to the local abundance, or if the tag recovery sample is a random sample from the population. Note that the abundance estimates will be unbiased if either of these conditions are met, and it is not necessary to meet both.

Scanning a significant proportion of the commercial and or recreational annual catch is likely to be the best practical option for recovering tags in SNA 1 and SNA 8 tagging programmes. However, using the commercial and recreational fisheries to recover tags is very likely to introduce bias because these sectors tend to direct fishing effort mainly to high catch-rate areas. Although there may be more random sampling options for recovering tags from SNA 1 and SNA 8 (for example, use of baited underwater scanners), these methods have not be used in large-scale tagging studies and would be likely to require considerable development before they could be used. Hence, given that almost all tag recoveries will come from commercial and recreational fisheries (i.e. tag recovery is likely to be biased), it becomes imperative that the tags are randomly distributed in the population during the release phase.

It is possible to correct for some spatial differences in the tag-release rates by post-stratifying the data, for example by dividing the area up into regions of comparable mark rates. Evidence of heterogeneity has been seen in all previous SNA 1 and SNA 8 tagging programmes (see Gilbert & McKenzie 1999, and Davies et al. 2006), thus all programmes have required post stratification to either correct for the differences in mark-rate or to determine the range of uncertainty that resulted from this effect (McKenzie & Davies 1996). However, post stratification as a means to deal with mark-rate heterogeneity can be problematic due to introduced bias from post-stratification as well as the bias in the Lincoln-Petersen estimator when the number of tag recoveries is small (typically less than five — see Chapman 1951). This is due to the fact that tag observations are integers — at low frequencies each individual tag scales (pursuant to Equation 2) to represent a large proportion of the population so the effect of seeing or not seeing a given tag results in large variability on the final estimate. McKenzie et al. (2011) simulated the effect of not accounting for spatial heterogeneity in tagging data, and concluded that less bias occurred when spatial heterogeneity was ignored, unless there were a relatively large number of tag recoveries available to adequately account for it.

A key design decision in a SNA 1 or SNA 8 tagging programme is how much bias (heterogeneity) to allow during the tag release phase. A strategy of random release in proportion to underlying abundance would require the use of dedicated research or charter vessels to release tags over the spatial extent of the population, and would be less prone to bias as a result of violating the homogenous mixing assumption. Alternatively, "opportunistic" tagging from, for example, the commercial fishery as a part of its usual fishing operations is likely to result in a non-random (heterogeneous) distribution of tags; the risk being that the number of tag recoveries would be insufficient to account for the bias through post stratification. Further, if the fishery did not release tags in all locations where the population was present, then post stratification would not be feasible. In summary, the "opportunistic" approach would be likely to be less expensive than the stratified random release strategy, but at the risk of introducing bias that would be difficult to measure and mitigate.

2 TAGGING TECHNOLOGY EVALUATIONS

Methods of marking or identifying fish can be divided into two categories:

1. Biological and chemical: marking patterns, morphometric traits, genetics, parasite signatures, chemical marking of hard structures (e.g., spines, otoliths), and mutilation (e.g., branding, fin clipping).
2. Physical tags (Figure 1): These can be classified into: external (e.g., dart tags, disc tags subcutaneous implant tags, satellite tags); or internal (e.g., PIT tags, coded wire tags, radio and acoustic tags)

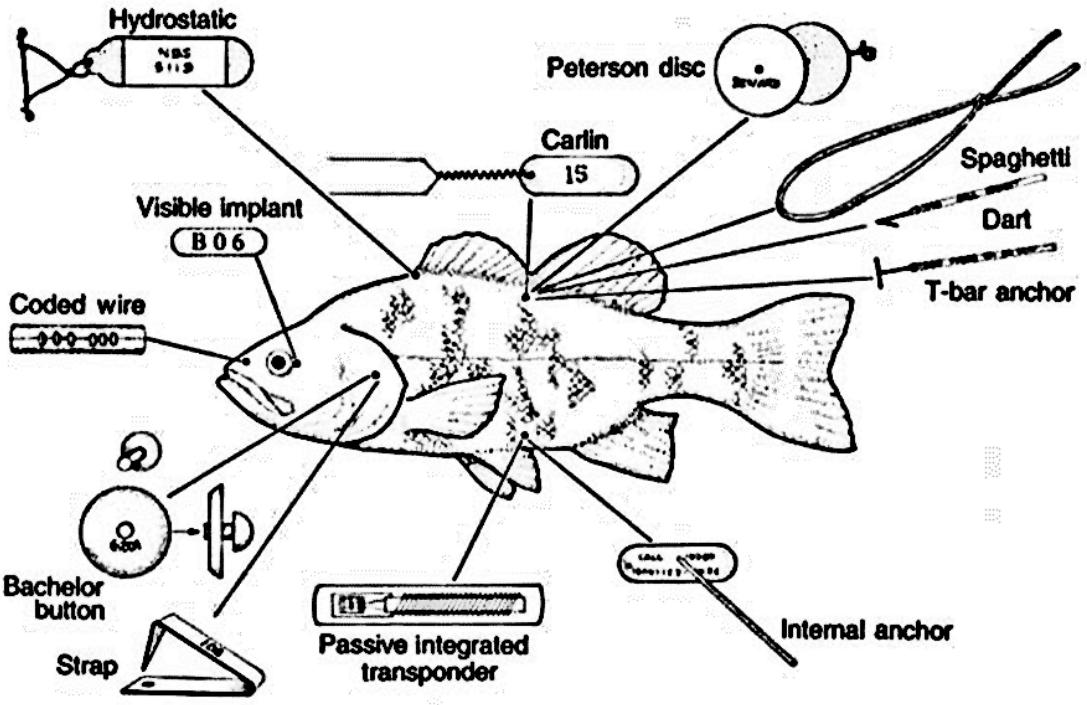


Figure 1: Some examples of common physical tagging technologies (Reproduced from Wydoski and Emery, 1983).

Future SNA 1 and SNA 8 tagging programmes are likely to require tagging technology that is relatively low cost, easy to apply, and that will uniquely identify each tagged fish. This is so that both estimates of movement as well as population abundance can be made from the resulting observations. The requirement for unique identification rules out most of the biological and chemical marking options for snapper, except for genetics and body pigmentation typing (i.e., photo identification).

Active physical tags such as hydro-acoustic, radio, and satellite tags are relatively expensive types of tags, and not suited to large scale tagging studies These technologies are designed primarily to assess local scale abundance or to investigate movement patterns, and are therefore neither an appropriate nor cost effective method for mass tagging snapper for the purpose of abundance estimation.

This reduces the list of potential tagging methods for SNA 1 and SNA 8 as either low cost "passive" external tags, internal coded wire tags (CWT), internal Passive Integrated Transponder (PIT) tags, genetics, or photo pigmentation identification methods.

A significant drawback of physical tagging (for example, the external, internal CWT, or internal PIT tags) is that all practical applications require first capturing an individual fish and then physically bringing it to the surface for tagging. The trauma from such activity may kill the fish or significantly alter its future behaviour. On the positive side, external, CWT, and PIT tags have all been used in previous SNA 1 and SNA 8 tagging programmes, and many of the mortality effects of these tags are well understood for snapper, and can therefore be accounted for (McKenzie & Davies 1996; Gilbert & McKenzie 1999).

Genetic and pigmentation methods have the potential advantage over physical tags in that it may be possible to collect tissue samples or photo-tag snapper in-situ without the need to bring fish to the surface (see Sections 2.3 and 2.4). However neither method has been applied to New Zealand snapper, and it is likely that a relatively long and expensive developmental phase would be required if these methods were to be applied.

2.1 External tags

The main advantage of external over internal tags is they are relatively easy to apply and most require no specialist equipment to identify the presence of a tag.

The two main disadvantages of external tags are that they are more prone to dropping out or causing ongoing health issues in the fish than internal tags, and also require fishers to both recognise and return tags. The high reliance on fishers to return tags, and the associated difficulties in quantifying differences in reporting rates is a key reason external tags have not been used in SNA 1 and SNA 8 tagging programmes since 1990. While there has been some development of methods to evaluate possible differences in reporting (see Mormede & Dunn, 2013), this approach is only able to distinguish relative differences in detection rates between vessels, and requires considerable overlap of vessel activity over space and time.

2.2 Internal tags

Both CWT and PIT tagging technologies are cryptic, meaning that it is generally not possible to determine if a fish is carrying a tag without "scanning" it using specialist electronic equipment. The recovery of tagged fish is typically by way of a dedicated catch scanning programme. This type of scanning has the advantage that both the number and size composition of fish passing through scanners can be controlled, and rates can be optimised. The cryptic nature of these types of tags means it is much more likely that differences in tag detection rates can be restricted to differences in equipment and the part of the catch being scanned, rather than to human factors which are much more difficult to quantify. Although the tag scanning process is typically not 100% effective at detecting tags (McKenzie & Davies 1996; Davies et al. 2006), tag seeding experiments allow for the estimation and hence correction of scanner detection rates (McKenzie & Davies 1996).

2.2.1 PIT tags

PIT tags are programmed in the factory with one of about 34×10^9 unique code combinations. An electronic scanning device polls the tag, and then detects its resulting radio frequency broadcasts and decodes its number (Figure 2). The tag gets the energy for RF broadcast from an electrical field generated by the scanning device. In the absence of an electrical field the PIT tag does not broadcast any signal as it has no power source of its own. The maximum distance a tag can be detected from is dependent upon the power of the detection device and the size of the tag. PITs are produced in a variety of shapes and sizes to accommodate the wide range of applications in which the technology is used (e.g., factory component tracking, security and enforcement, livestock identification, and animal tracking). Scanning devices range from small handheld units, which can only read a tag over a distance of a few

centimeters, up to large stationary systems that can read a tag in a cattle beast as it passes through a race. PITs are an electronics growth area and there are many manufacturers and many models and products available; the technology is constantly evolving and improving.

Many of the technical difficulties associated with the use of PIT technologies were resolved prior to the 2001 SNA 8 tagging programme, when the technology was first applied to snapper. Some key developments were:

- Use of "food-safe" plastic PIT encapsulation materials (prior to 2001 glass was primarily used);
- Tag trials to determine tag retention success and level of biological rejection in snapper;
- Reusable single and multi-shot tag applicators;
- Development of bulk scanners suitable for use in commercial fish factories.

Details of the development of PIT plastic encapsulation and subsequent tag testing on snapper can be found in McKenzie et al. (2006). The general conclusions from the McKenzie et al. (2006) investigations were that 12 mm plastic PIT tags (Figure 2) injected inter-peritoneal had a largely benign effect on snapper health, and that once in place annual tag retention rates are likely to be better than 99%. These findings were consistent with the use of PIT tags in other fish species.

Figure 2: A 12 mm PIT tag from the 2001 SNA 8 tagging programme being scanned.

In order to detect a 12 mm PIT in a 35 kg bin of snapper, two technical challenges needed to be overcome:

1. Reading a weak RF transponder signal at 50 cm distance from the aerial (signal strength issue);
2. the RF blind spot that occurs when the transponder is orientated at 90° to the aerial (directionality issue).

The bin scanner developed by Sanford NZ Ltd. for the 2001 SNA 8 tagging programme incorporated three high gain aerials arranged in different planes to counter the tag orientation and signal strength issues (Figure 3). This scanner proved to be 85% effective in recovering tags. However, it is highly likely that advances in PIT scanning technology since 2001 mean that scanning success will be significantly improved. Development of an improved bulk scanning process for use on commercial vessels or processing plants will be a requirement if PIT tags are used again in future SNA 1 and SNA 8 tagging programmes.

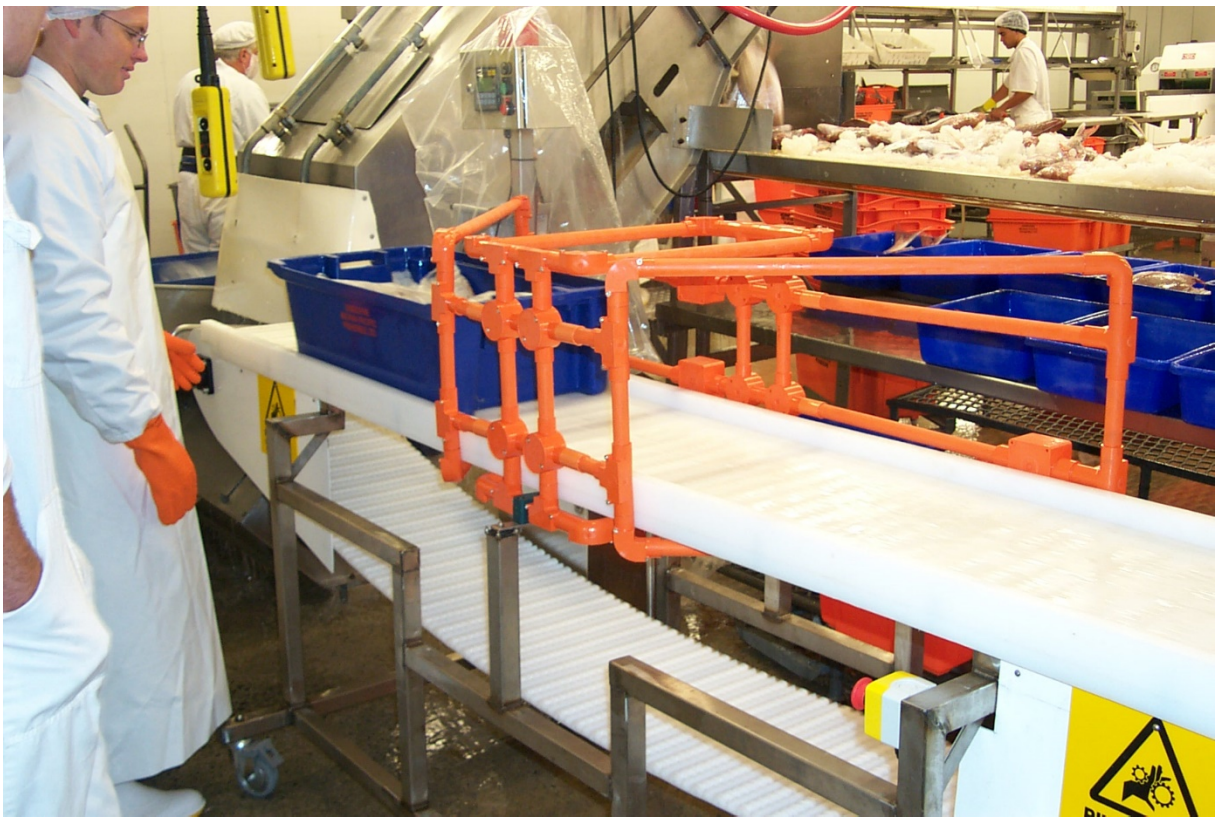
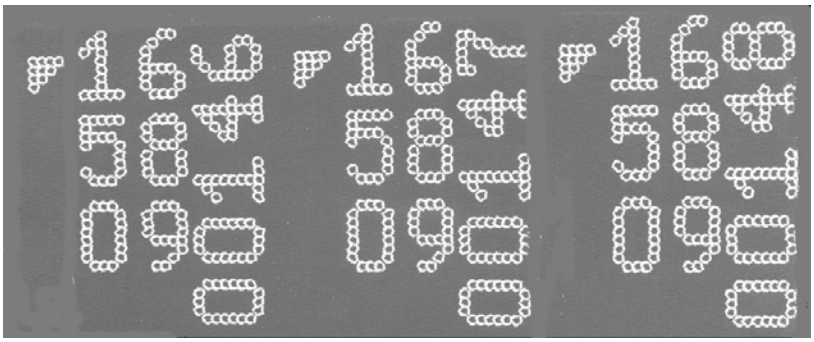


Figure 3: PIT tag fish-bin scanner developed by Sanford NZ Ltd for the 2001 SNA 8 tagging programme. The orange frames enclose three RF antenna arranged in different planes.

2.2.2 Coded wire tags

The CWT is constructed of 0.2 mm magnetised stainless steel wire and is typically 1 or 2 mm in length (Figure 4a). Each 1 mm length of wire (regardless of where it is taken from the spool of tag wire) will contain one unique sequential code. The technology is protected by worldwide patent and is exclusive to one North American company (Northwest Marine Technologies (NMT) Ltd., Washington State). Number sequencing on CWTs were previously by way of grey-scale binary codes but are now laser etched decimal codes (Figure 4b). CWT codes are unknown at the time of application; however, as all codes are sequential, the collection of a reference tag between each application provides for the unique identification of subsequently recovered fish. The tag is inserted into the body of the fish, with the aid of an applicator device, either intra-muscularly or just beneath the skin.



a. CWT shown to scale b. Form of laser etched decimal code (sequential)

Figure 4: Coded wire tag showing form of decimal etched code.

Each CWT is magnetised as it is inserted into the animal. NMT manufacture a range of detection devices, all of which work by registering the small magnetic field associated with the tag. Handheld scanners were used to examine catches in the 1994 East coast snapper tagging programme (McKenzie & Davies 1996). Although this programme successfully examined 1000 t of snapper over a one-year period, the associated labour costs were high, due to the requirement to individual wand-scan fish to recover tags, collect reference tags for each released fish, manually read every reference and recovered tag under a microscope, and subsequently manually enter the data. CWT technologies are likely only to be a viable alternative to PIT if the above limitations could be overcome through technological improvement.

As an alternative to the manual wand scanner, NMT market a high through-put tunnel scanner; the detection success of which, the company claims, is close to 100% (R9500 tunnel scanner, Figure 5). The R9500 tunnel scanner is likely to be impractical for use on commercial snapper catches as it would probably require unpacking fish to pass them through the scanner and then repacking them. Although it would be technically feasible to build a larger version of the R9500 scanner that would accept whole fish bins, this technology does not currently exist. Developing this scanner would require convincing NMT of its commercial viability.

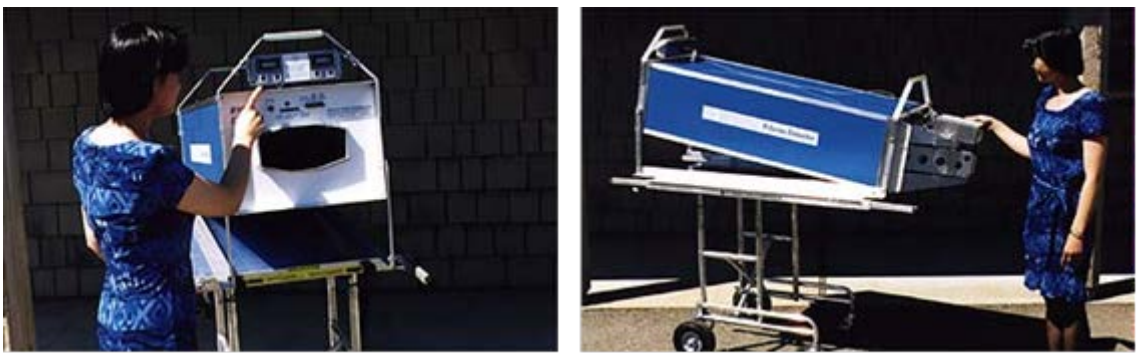


Figure 5: Northwest Marine Technologies R9500 Coded Wire Tag Tunnel detector.

2.3 Genetic tags

2.3.1 Overview

Conventional mark-recapture for snapper using internal tags is a proven approach, but some associated problems are difficult to resolve. In particular, there is mortality associated with the tagging process, which adds uncertainty to resulting estimates. It is difficult to precisely estimate rates of release mortality, since retaining tagged fish in holding tanks permits only limited sample sizes. Many confounding covariates that may be related to survival increase the difficulty. These include fish size, capture method, capture depth, and fish condition on release. Furthermore, poor recapture rates of snapper that were tagged below 80 m depth in past tagging programmes suggest that fish tagged from these depths rarely survive, so there are potentially significant parts of the stock for which conventional tagging is not possible (Section 4.2). Genetic mark-recapture, or gene-tagging, provides a way to avoid problems with release mortality by using in-situ tagging (Ovenden et al. 2013), thus opening up the possibility of resolving some of these issues.

Genetic mark-recapture is now standard practice in wildlife studies (Luikart et al. 2010). It is used when animals are captured directly, and also with remotely harvested material such as hair or faeces (Banks et al. 2003, Bellemain et al. 2005). Genetic mark-recapture methods can mimic the same approach as used for physical tags, except that is identifies the individual based on its unique DNA fingerprint. By avoiding inserting a physical tag into the animal, the possibility of tag loss is eliminated.

DNA fingerprints are derived from variable genetic markers, such as microsatellites or single nucleotide polymorphisms (SNPs) (Ovenden et al. 2013). The DNA fingerprints are chosen so that they uniquely identify individuals. The fingerprints must have a very low error rate with regards to identifying, or the population size will be overestimated (Mills et al. 2000).

Within New Zealand, genetic mark-recapture has been used in marine environments for Maui's dolphins (Baker et al. 2013) and southern right whales (Carroll et al. 2011).

Genetic mark-recapture for fish has been successfully applied to several species, based on identifying parents and offspring. With salmon, offspring are 'tagged' as fry by genotyping spawners in hatcheries (e.g., Steele et al. 2013). Recaptures of the offspring are made by genotyping samples from the catch. Genetic methods initially used microsatellites, but more recently have moved to using SNPs.

Genetic mark-recapture with parent-offspring pairs has recently been developed for southern bluefin tuna (Bravington et al. 2014), and in this case the analysis method is more complex because both parents and offspring are randomly sampled from the wild population. When the genotypes of the parents are unknown it is more difficult to identify parent-offspring pairs, and so higher resolution in the genetic data (lower probability of identity) is required than when parent genotypes are known.

Genetic mark-recapture for *P. auratus* is being used in a project currently under way at the University of Auckland (see http://tinyurl.com/mmdtykf). This project involves estimating the contribution of reproduction within a marine reserve to recruitment in surrounding areas, and the gene-tagging is therefore targeted at identifying parents and offspring. Tagging of potential spawners within the reserve occurs by catching individuals with barbless hooks, and taking a fin clip. Genetic identification is by a panel of microsatellites.

Remote hooking will be a key technology in gene-tagging snapper, since it permits tagging fish without bringing them to the surface, eliminating barotrauma and potentially reducing tagging mortality and behavioural effects to negligible levels. Remote hooking technology for gene-tagging fish was initially developed for troll-caught Spanish mackerel (*Scomberomorus commerson*, Buckworth et al. 2012). Technology for remote hooking using baited hooks for benthic species has been developed for rockfish in the US (Tomich et al. 2010).

2.3.2 Genetic mark-recapture methods

In this section we describe the methods for implementing genetic mark-recapture and compare them with the PIT tag approach. The four stages are tagging fish in the water, screening the catch, using genetic methods to determine the identities of tagged and screened fish, and statistical analysis of the data.

Tagging

Remote sampling hooks provide a method of obtaining a genetic sample of fish without the need to catch or bring the fish to the surface. These take a small flesh sample from an individual fish. The hooks are baited and deployed on a longline in a similar way to standard hooks. A DNA sampling hook has been developed by the US National Marine Fisheries Service (Tomich et al. 2010), and may be suitable for use with snapper (Figure 6). When a fish bites the hook and puts pressure on the barb, the hook is triggered and the barb retracts inside a sleeve that encloses the shaft. After the longline has been recovered, barbs are removed from triggered hooks and placed in a sample storage medium.

Figure 6: DNA sampling hook, with the sampling barb deployed (bottom left) and with the barb retracted inside the enclosure (top right).

The longline would be set up with both conventional hooks and DNA sampling hooks, with ratio to be refined, but probably about 1:1 initially. The catch on the conventional hooks can be used to identify the species composition and size composition of the population being fished, while the DNA sampling hooks provide the individual genetic tagging information. There could also be potential to tag and release the conventionally hooked fish with both genetic and PIT tags, which would allow estimation of both conventional hooking mortality and the relative selectivity of conventional and DNA sampling hooks.

This approach is similar to PIT tag mark-recapture in that it involves fishing with longlines. It differs in that fish are not brought to the surface, and hence can avoid many of the associated trauma and mortality effects.

Screening

Like PIT-tag mark-recapture, the recapture method for genetic tagging involves screening part of the catch for tagged fish. This approach has the advantage that the number of fish examined is known, and there are fewer issues with estimating rates of tag detection. In experiments that rely on recaptures being identified and reported by eye, estimating the reporting rate is problematic (owing to both non-detection and non-reporting) and is a source of considerable uncertainty.

The fish to be screened should be a random sample from all the catch in a fishery, since the harvest rate estimate is associated with a fishery in the stock assessment. During screening a flesh sample is taken from each fish and placed in a sample storage medium. The flesh sample should be large enough to eliminate risk of contamination via slime from other fish, such as a large fin clip.

Genetic assay methods

Genetic assay technologies are rapidly advancing, and there are a number of possible ways to analyse the DNA in order to identify individuals. The three main options are 1) microsatellites, 2) single nucleotide polymorphisms (SNPs), and 3) next generation sequencing. All approaches examine segments of the nuclear genome, but there are advantages and disadvantages for each one.

The aim of the genetic analysis is to uniquely identify the individual, and careful planning is needed to ensure that the genetic analysis has sufficient power to do this. Each individual snapper has unique nuclear DNA, but the genetic analysis only looks at a very small part of the nuclear genome – i.e., a discrete number of loci (the genetic marker). If too few loci are examined, several individuals may have the same alleles at those loci, and be wrongly thought to have the same identity –- resulting in a false recapture. The risk of two randomly chosen individuals having the same DNA signature at a set of loci is known as the 'probability of identity' for that set (Waits et al. 2001).

2.3.3 Issues to resolve

Field development

There are a number of issues associated with the field component of the approach. Some of these will require dedicated field experiments.

Estimate success rates

Mark-recapture experiments require that the recapture rate is high enough to derive precise estimates of population parameters. There is therefore a need to release enough tags into the population to ensure a high recapture rate. Planning requires an estimate of the number of tags released per unit of effort. Pilot studies would be required to estimate this rate. Factors affecting the tagging rate include 1) the proportion of DNA sampling hooks that are triggered, 2) the proportion of triggered hooks that retain a flesh sample, 3) the proportion of flesh samples that are snapper, and 4) the proportion of snapper flesh samples that can be successfully genotyped.

Hook development

The US National Marine Fisheries Service (NMFS) DNA sampling hooks are designed for rockfish, and may require adaptation to work successfully for snapper. The hooks will need to be tested to see if they retain snapper flesh samples, and if the design can be improved. Tank testing would allow observation of snapper interactions with the hooks. Input from industry will also be useful to improve hook design and usability.

Hook manufacturing

The DNA sampling hooks are not manufactured commercially, and would need to be manufactured specifically for this project. NMFS has agreed to provide plans for hook design. Manufacturers will need to be identified.

Hook selectivity

Knowing the size selectivity of the hooks is important for modelling the size distribution of tagged fish. Field and tank trials will be required in order to estimate the relative selectivities of DNA sampling hooks and conventional hooks.

Cross-contamination

When screening the catch for recaptures, it will be important to avoid cross-contamination, where genetic material from several fish is mixed together. When this happens, a sample may contain more than one individual, or a single individual may be DNA fingerprinted more than once. Either situation may bias mark-recapture estimates, or reduce sample sizes and increase costs. Cross-contamination can be avoided by appropriately designing the sample collection procedure and the sample type. Experiments may need to test for cross-contamination with different sample types.

Sample management

Genetic mark-recapture experiments involve examining very large numbers of samples for a small number of recaptures. The probability of each screened fish being a tagged fish is very low, and the rate of false positives must be kept much lower. Sample management is therefore critically important. Sample numbers would be very high, which makes it almost inevitable that there will be some errors in labelling, sorting, transcribing, or other elements of sample management. Protocols would need to be developed that minimise the probability of error.

2.3.4 Choice of genetic fingerprinting method

Choice of technique

Selecting the genetic fingerprinting method (or genetic marker) will be a key step. The two main approaches use microsatellites and single nucleotide polymorphisms (SNPs). A further approach is next generation sequencing.

Microsatellite arrays for snapper are currently available (Shane Lavery, University of Auckland, pers. comm.) and have been used to identify snapper. The arrays are currently being used for identifying parent-offspring pairs. These require very high resolution and are likely to have sufficient resolution to be used for mark-recapture with a large population. There are, however, some issues working with microsatellites. The analysis and interpretation has not been fully automated and requires subjective human input. This may cause problems for a large project that requires multiple operators and may also make it difficult to transfer genetic analyses between laboratories.

Analysis of SNPs is a more recent development. SNP libraries have not yet been developed for snapper. However, marker scoring can be automated, which makes the process portable between laboratories and suitable for scaling up with few compatibility problems. The error rate is likely to be lower than microsatellites, and the costs may be similar. The required DNA quality is also similar.

Next generation sequencing involves automatic sequencing of a section of the genome. Marker scoring is automated, so the approach is portable between laboratories. Error rate is likely to be similar to SNPs, but cost may currently be in the order of double the costs for SNPs and microsatellites. However, costs are highly variable between laboratories and depending on the quantity of samples being analysed. Next-generation sequencing is a newer technique than the others, and costs are likely to decline more rapidly.

2.3.5 Design panels for probability of identity

The probability of identity (pID) of a set of markers is the probability that two individuals randomly selected from the population have the same DNA fingerprint at those markers — in other words, that they appear to be identical. pID does not depend on the population size, but larger populations require lower pID, because there are more individuals to potentially share the same markers.

Ensuring sufficiently low pID is an important part of the genetic marker development process. It will require pilot studies of the allele frequency in the target population of the microsatellites, or the variability of the SNPs. It also requires estimates of the relatedness in the target population.

The design process will also require investigation of error rates (such as allelic dropout), and whether sample quality (e.g., the time on the longline, and the water temperature) affects error rates and the probability of successful fingerprinting.

The design process differs depending on whether the analysis uses microsatellites, SNPs, or next-generation sequencing.

2.3.6 Identify efficient low-cost operators

Fingerprinting large numbers of fish can be very expensive, and genotyping will be a large component of overall costs (all scanned and all recaptured fish have to be fingerprinted). It will therefore be important to identify ways to reduce genotyping costs. There are companies in South Korea and southern China that offer low cost analyses of high volume genetic data and may provide substantial savings. However, it may be more difficult to deal with problems that occur when working with companies at a distance. For example, DNA from remote sampling hooks are likely to be of inconsistent quality, and this can cause difficulties for some analysis systems. Identifying suitable operators will require trials and specialist expertise and would require on-site management, at least in the initial stages.

2.3.7 Comparison of gene-tagging with PIT tagging

Genetic mark-recapture has a number of potential advantages over PIT tagging, but there are also some significant disadvantages.

Advantages

Minimal release mortality
Release mortality introduces both bias and uncertainty to mark-recapture experiments (Arnason & Mills 1987). Mortality rates are difficult to estimate precisely, and can vary with covariates such as depth, water temperature, and treatment during the tagging process (Bartholomew & Bohnsack 2005; Hoyle et al. 2015). It is a significant problem with mark-recapture experiments that use PIT tagging. Mortality due to gene-tagging is likely to be considerably lower than PIT tag mortality, and may even be

negligible. First, barotrauma causes substantial mortality for PIT tagged fish, whereas remotely gene-tagged fish will not experience any depth displacement due to the tagging process. Secondly, the gene-tagging hook folds up after taking a sample and the barb retracts beneath a cover, leaving a simple tube that cannot gut hook or otherwise damage the fish (see Figure 6). Thirdly, gene-tagged fish are released in situ rather than at the surface, and so are not subject to increased risk of predation mortality which can be associated with displacement, exhaustion, and separation from the school (Raby et al. 2013).

Can tag anywhere
Mortality due to barotrauma makes it much less feasible to PIT tag at depths greater than about 75 m, which leaves significant parts of the fish habitat with no released fish. Gene-tagging has no such depth restrictions, so tagging can be applied in all areas in the same way.

Minimal behaviour change
Tagging models assume that tagged fish are representative of all fish, so if tagging changes individual behaviour it can introduce substantial bias to tagging estimates. Behavioural response to the tagging process can involve changed catchability, if the experience teaches the fish to avoid certain types of fishing gear, often termed trap shyness. This may have occurred in snapper in previous tagging studies (Gilbert et al. 2001). Tagging can also affect movement patterns, with the fish avoiding the location where it was tagged. Gene-tagging is less likely than PIT tagging to affect the fish's behaviour, since the trauma will be considerably lower. The fish would experience some immediate effects from the tagging hook, but this would be relatively superficial and brief compared to being caught by a longline hook. The gene-tagged fish would also not be physically moved, removed from the water, handled, injected with a tag, or dropped back into the water and required to swim back down.

Minimal effects on growth
Tagging affects the growth of many species, regardless of tag type (Murray & Fuller 2000). These effects may occur due to stress and damage associated with the tagging process, or due to the effects of the tag itself. PIT tagging does not avoid initial effects associated with tagging, but the use of internal PIT tags may avoid ongoing effects as they are too small to significantly affect the fish. As described above, gene-tagging involves little trauma, and the tag is not a physical object, so effects on growth are likely to be minimal.

Known reporting rate
The reporting rate can be difficult to estimate, and is similar to release mortality in that it directly scales fishing mortality rate estimates. Mark recapture with PIT tags is a relatively effective way to deal with the problem of estimating reporting rate, but genetic mark-recapture may be slightly better. PIT tag detectors may not detect all the tags in a fish bin, and detection rate may depend on the orientation of the fish in the bin. There may also be uncertainty about the numbers of fish passing through the scanner. Gene-tagging in contrast provides precision about the number of fish screened, effectively eliminating this source of uncertainty and bias.

Tags don't affect food industry
PIT tags have been approved for use in food products, but the need to have them approved has been a significant cost for previous snapper tagging projects. Genetic tagging has no such concerns since no artificial tag is left in the animal. However, the screening process may require development to address any food quality concerns.

Screening can be done anywhere
While hand-held scanners are available, PIT tag screening has in the past been restricted to fish processing plants, where high quality scanners can be installed and large numbers of snapper processed efficiently. As well as the higher unit cost of screening lower sample numbers, handheld scanners may have lower detection rates that may bias the tag reporting rate. In contrast, sampling the catch for genetic tags can be undertaken with the same high detection probability in any location including boat ramps and smaller fish factories. However, as for PIT tags the labour costs of genetic screening will be higher at boat ramps and smaller factories, since fewer samples will be available.

No tag loss
Immediate tag loss has the same effect on tag abundance estimates as release mortality. And, as well as immediate tag loss there may also be an ongoing tag loss from tag failure or other tag loss processes. This introduces a slightly different type of bias into tagging models. Tag loss with PIT tags is generally thought to be low, but is non-zero. Tag loss with gene-tagging is zero, since the tag cannot be lost.

Population genetic information
Gene-tagging large numbers of snapper would generate an unprecedented quantity of population genetic information about the SNA 1 stock. Such data can be highly informative about the population, and would be available for later application. For example, estimates of the genetic effective population size at varying temporal and spatial scales may be informative about genetic drift, population connectivity, and the size of the spawning population (Hare et al. 2011). The relatedness within the population and spatial aggregation of related individuals may be informative about rates of population mixing. This may help to identify areas that provide more spawners than others. Information about differentiation within the population between locations, seasons, and age classes may permit improved management and higher yields, given better understanding of population processes.

Can store samples and data for later use
Genetic techniques and the statistical techniques used in analysis of genetic data are continually improving, and genetic analysis costs are continuing to decline. Methods available in future may permit much more information to be extracted from existing samples, potentially at significantly lower cost. Genetic samples are small and can be stored for significant periods relatively cheaply, and well-designed databases are similarly long-lived.

Disadvantages

The number tagged is unknown at time of tagging
Previous snapper mark-recapture experiments have tried to tag the same proportion of the population in each area/stratum, based on prior estimates of population density from longline CPUE. Tagging in an area ends when enough individuals have been tagged. However, the number of gene-tagged fish in an area depends not only on the number of hooks triggered, but also on the proportion of triggered hooks that retain usable flesh samples, and the proportion of samples that are snapper. These proportions can only be estimated in the laboratory, since usable samples would often be too small to be visible. Understanding of likely proportions would develop with experience, and in future educated guesses based on local conditions may be adequate.

Lengths of individuals at release are unknown
Tagging models use the length of fish at time of release to estimate its probability of recapture, given the size selectivity of different fishing gears. PIT tagged fish are brought to the surface and measured, so their sizes at release are known quite accurately. However with gene-tagging there is no information on the size of tagged fish. One approach to resolving this is to intersperse gene-tag hooks with conventional hooks, and use the fish from the conventional hooks to estimate the size distribution of the gene-tagged fish. This assumes that the size distributions from the two types of hook are the same, but this does raise further issues. First, the selectivity of gene-tag and the conventional hooks are likely to be different, since the physical hook would have different properties. The size of the difference would need to be estimated. Second, this approach provides a much less precise estimate of the size distribution of the released fish than can be obtained from direct measurement.

Various one-off development issues – costly
Gene-tagging involves techniques that are new to New Zealand fisheries, some of which would require significant development work. This development and the associated costs are outlined below in Section 2.3.8, stage 1.

Timing
Genetic tagging and screening will be a slower operation than PIT tag tagging and screening.

a. Bulk scanning is slower than with PIT tags, and potentially disruptive
At the sampling location, screening PIT tags can be relatively fast and efficient, because fish are screened electronically as they pass along a conveyer belt. Genetic tagging in contrast requires a flesh sample to be taken from every fish. This could take some time to process each fish, as a suitable clean sample must first be taken, placed in a storage medium, and then the associated paperwork filled out. The process of sampling may also be an unacceptable interruption to normal commercial processing operations.

b. Information feedback is considerably delayed
Analysis of both the tagging and the screening samples would take significant amounts of time, given the need to collect sufficient samples, ship them to the genetics laboratory, process the samples, determine the identity, and check the results. This complex logistical operation is likely to take considerable time, particularly in the early stages when protocols are being developed.

c. Risk of time blowout
Gene-tagging involves techniques that are relatively new to New Zealand fisheries, and there are likely to be unforeseen problems. Some of these problems may be difficult to resolve. There is therefore more potential for significant delay at the start of a gene-tagging project than a PIT tag mark recapture project.

d. Delayed start to tagging
Experiments to develop techniques required for gene-tagging would involve significant amounts of time, and are likely to delay the start of a project by several years.

Risk of failure
The two previous gene-tagging experiments in fisheries that used remote tagging hooks have succeeded in developing many aspects of the approach, but neither experiment successfully estimated fishing mortality, due to unforeseen logistical issues that arose during the studies. Given the untried nature of gene-tagging for snapper with remote sampling, there is a high risk of unforeseen factors effecting the success of the study.

Costs may be two times higher than PIT tagging
Initial indications are that gene-tagging may be more expensive than PIT tagging although this may be offset by the higher precision of estimates from gene-tagging permitting lower sampling intensity. Current estimates put the cost of a gene-tagging at approximately twice that of a comparable PIT tagging programme. However cost estimates for gene-tagging are currently very imprecise, since they depend on unknown factors such as the proportions of deployed hooks that obtain snapper samples, and the costs of genotyping large sample numbers.

2.3.8 Proposed approach

If genetic tagging were to be used to estimate biomass of New Zealand snapper stocks the following staged approach would be required.

Stage 1: Method development

Stage 1 would be aimed at resolving the important unknowns associated with genetic mark-recapture for snapper. Key activities are development of hooks and genetic markers. Based on discussions with geneticists, and taking into account additional experiments to test and develop hooks, it is possible that the method development would take 12–18 months and cost in the order of $500 000–$1 000 000.

Development of Genetic Markers
For initial work the best approach may be to use the panels of microsatellites developed by the University of Auckland. Research providers could seek to collaborate with the project currently under way in the Leigh Marine Reserve. This would reduce costs through the use of existing information, including genotyping error rates, estimates of relatedness, and probability of identity. In parallel, it would be useful to investigate the use of SNPs, which is likely to be more efficient in the long term. As a part of the sampling program, it may be prudent to store enough of each flesh sample to allow future reanalysis using SNPs.

Tank trials
Initial work would require at least 20 DNA sampling hooks, which may be obtained from the US National Marine Fisheries Service, Northwest Fisheries Science Centre. Tank trials could be conducted to resolve questions around methodology and suitability.

In particular the following is recommended:
- Observe a number of interactions between the fish and the hook, in order to understand the tagging process
- Trial alternative baiting methods to optimize rates of bait contamination
- Estimate rates of sample collection per triggered hook
- Estimate levels of damage to fish that interact with the hook
- Consider potential design improvements
- Estimate DNA usability as a function of soak time and water temperature

Hook manufacture
We recommend exploring options for manufacturing DNA sampling hooks both in moderate (5 000–10 000) and large (50 000–150 000) quantities. Design drawings for the current hooks are available.

Field trials
A series of field experiments is necessary to develop and test approaches for gene-tagging fish. Given the existing genetic mark-recapture experiments by the University of Auckland, using boat-based tagging, and initial field trials may be carried out adjacent to the Leigh marine reserve.

Trials should include the following:
- Trial the tagging process, and develop and trial protocols for sampling and data recording
- Estimate the optimal number of hooks per set, and the proportion of conventional hooks
- Estimate the proportions of hooks triggered per set
- Retain sampling heads from all triggered hooks in order to estimate the sampling rate per triggered hook, and the rate of snapper per sample.
- Estimate the proportion of snapper samples that produce a viable DNA fingerprint
- If feasible, estimate the relative selectivity of the DNA sampling hooks and conventional hooks

Screening
The collection of samples from the catch by sampling at fish processing plants and at recreational boat ramps would need to be tested. These screen trials should include:

- Trial the screening process, and develop and trial protocols for sampling and data recording
- Liaise with industry to identify potential food safety issues, and any costs that may be associated with visible effects of sampling on fish
- Test alternative sample types (e.g., fin clips or whole fins) to identify cross-contamination of DNA between individuals

Stage 2: Small-scale trial

Methods development should be followed by a small-scale trial. One approach could be to undertake a mark-recapture experiment in a limited area for both tagging and screening, but with higher spatial densities of both tagging and screening than in a full experiment. It may be preferable to link the trial with the existing study at the Leigh Marine Reserve, as working in a system that is better understood reduces the chances of unexpected events leading to failure. Recapture of about 50 snapper would allow some understanding of likely effects. Such an experiment would be likely to take about 18–24 months to complete with a cost of between $1.0–1.5 million.

Stage 3: Full SNA 1 or SNA 8 programme implementation

The full-scale SNA 1 or SNA 8 genetic tagging programme could be undertaken once developmental stages 1 and 2 were successfully completed. The overall design of the experiment would be similar to the approach used in the conventional tagging experiments, but with the addition of tagging in areas deeper than 80 m.

The relative effort allocated to tagging and screening will depend on the relative costs of genotyping and vessel time. When genotyping costs are lower, a higher proportion of the overall costs come from tagging, so that for the same number of recoveries costs can be reduced by increasing screening and reducing tagging.

2.4 Photo tags

The use of spot patterns to uniquely identify snapper (Figure 7) was recently suggested by David Ashton, Nicholas Tuckey, and Alistair Jerrett from Plant & Food Research Ltd.

Figure 7: *Pagrus auratus* spot patterns are likely to be unique for each fish and these could be used to uniquely identify individual animals.

In common use in cetacean research, photo id marking is a relatively untried method in marine fish and has no precedent for New Zealand snapper. And, unlike genetic tagging where there is a significant body of published research available to at least establish "proof of concept", additional time and expenditure would be required to "proof of concept" photo tagging as method for snapper. Some of the component steps to establish "proof of concept" and to develop the technique to meet the requirements of SNA 1 and SNA 8 tagging programmes are:

1. Establish that spot patterns can be used to identify individual snapper, especially from very large samples sizes i.e. like human finger prints.
2. Establish that individual snapper spot patterns remain largely unchanged through time as the fish ages and grows.
3. Establish that snapper spot patterns are still discernible in dead and chilled trawl caught fish.
4. Determine the feasibility and practicalities involved with photo imaging live snapper, either at depth or on board research vessels at the surface.
5. Determine the feasibility and practicalities of photo-scanning dead and chilled snapper in commercial fish sheds or on commercial fishing vessels.
6. Develop portable fish photo imaging hardware and image processing software for photographing snapper from research and charter vessels.
7. Develop high throughput photo scanning hardware and image recognition software suitable for use in commercial fish processing plants.

Dr Paul Pang, Associate Professor from the department of computing at Unitec Institute of Technology and who specialises in decentralised machine learning intelligence, was asked to comment on the feasibility of a mark reading scanner for snapper. Although technically feasible, in Dr Pang's opinion scanning technology suitable for deployment in fish processing plants would take about 2–3 years

research and development time and be likely to cost in the region of $1.0–2.0 million to develop. Dr Pang was sceptical that it would be technically feasible to photo image live snapper at-depth in such a manner as to obtain an image that could be consistently and reliably identified. Current imaging techniques require a well-lit subject in a consistent orientation and camera viewing angle for optimum performance. This would suggest that fish are likely to need to be brought to the surface to image, thus negating the main potential benefit the technology has over physical tagging. Fish would also need to be individually scanned, instead of bulk scanning in bins as is possible with PIT tags, which would be likely to cause unacceptable disruption to processing shed operations

3 OPTIMUM SNA 1 TAGGING DESIGNS

In this section we look at various tagging design options for achieving given levels of estimate precision. Specifically investigated are:

1. The levels of bias associated with random verses "opportunistic" or non-random release designs;
2. The number of tag recoveries required to achieve the best "practicable" precision on the population estimate;
3. A cost benefit evaluation of frequent verses infrequent tagging events for deriving point-in-time population estimates.

It was not possible to provide absolute estimates of cost within the scope of this report, as key component factors such as charter costs and scanning costs were not known. We provided instead a cost benefit analysis tool (see Section 3.4) that enables reliable estimation of relative costs of the various alternatives, and will allow stakeholders and managers to determine the real overall cost of various tagging designs once the main unit cost components have been determined.

3.1 Opportunistic versus stratified (proportional) release designs

An "opportunistic" tagging approach would see tags being released from commercial fishing vessels during the course of normal fishing events. Such a strategy is likely to result in a higher proportion of tags being released in high catch-rate (i.e., high CPUE) areas and therefore lead to a heterogeneous distribution of tags in the population. Commercial CPUE is usually used to design and evaluate retrospectively the spatial and temporal distribution of tagged fish.

A bootstrap simulation approach was used to investigate the likely degree of bias introduced by disproportionate tagging in high catch rate areas.

3.1.1 Methods

The bootstrap population model was stratified into spatial areas or strata of differing size, with each stratum assigned the same number of fish. The density of snapper within each stratum therefore varied according to stratum size. Bootstrap runs were undertaken in which each stratum was randomly assigned a catch-rate as a random normal deviate in log-space, with the mean a function of the density. The level of catch-rate variation between strata was altered by adjusting the coefficient of variation (CV) on the random draws. The fact that the stratum catch-rates varied whereas the stratum population sizes were constant meant that stratum-area was inversely proportional to catch-rate in the simulations.

Three simulations were undertaken:

1. Stratum allocation of tags in proportion to population number, i.e., stratum mark-rate range = 1:1 (Stratified random allocation)
2. Stratum allocation of tags in proportion to catch-rate, i.e., stratum mark-rate (Equation 3) range ≈ catch-rate range (opportunistic tagging)
3. Stratum allocation of tags in proportion to area, spatial allocation tagging, i.e., stratum mark-rate range ≈ stratum-area range.

For all simulations it was assumed that tag recoveries would come from the commercial fishery, therefore the allocation of the scanned catch across strata was in proportion to the stratum catch-rate. The expected number of tag recoveries in each stratum was derived by multiplying the stratum mark-rate by the stratum catch. The tagging population estimate was derived from the sum of all tag releases, tag- recoveries and catches ignoring strata. The proportional bias for each bootstrap run was simply the tag estimate divided by the true population number minus 1.

3.1.2 Results

The median level of bias for all random allocation simulations (simulation 1) was 0.0 (unbiased) despite the catch recovery allocation being biased toward high catch-rate strata (i.e., non-random) (see also Section 1.2 above).

Opportunistic tagging, where the number of tagged fish in each stratum was directly proportional to catch rate (simulation 2), produced negatively biased population estimates across all scenarios (Table 1). The level of bias was strongly determined by the CPUE range across strata and the number of strata (Table 1). Bias was low (less than 6%) when the range in stratum catch-rates was less than 1:2; the bias reducing as the number of strata increased (Table 1).

Both the 1985 and 1994 SNA 1 tagging programmes used stratified random tag-release approaches. Analyses in Gilbert & McKenzie (1999) suggest that the range in the mean tags-per-tonne ratio (i.e., mark-rate) across all five 1985 SNA 1 tagging programme areas was 1:4, declining to 1:2 in the eight area 1994 programme. Given our improved understanding of the relative spatial distribution of the SNA 1 fishery, through fine scale catch and effort reporting (see Section 4.4), the likely level of spatial heterogeneity in future SNA 1 tagging programmes using stratified random release designs should be lower than 1:2.

It is unknown what range in mark-rate would be achieved through opportunistic tagging by the commercial fishery across SNA 1. The expectation is that the mark-rate range would be greater than achieved in 1985 using a stratified tag release design (i.e. four times). Based on the level of spatial variation seen in commercial longline CPUE across SNA 1 (see Section 4.3); we believe mark-rate heterogeneity in excess of 10 fold is likely under a purely opportunistic tag release strategy (this equating to negative bias in the order of 25–40%; Table 1)

Table 1: Median percentage bias associated with opportunistic release where mark-rate is proportional to catch-rate for 2, 10 and 50 fold range in median catch-rate across strata.

Stratum	Number of strata			
mark-rate range	5	10	20	40
1:2	-6%	-4%	-3%	-2%
1:10	-39%	-30%	-27%	-25%
1:50	-56%	-56%	-53%	-48%

Opportunistic tagging based on stratum size/area (simulation 3) was positively biased in all scenarios (approximately equivalent to Table 1, but in a positive direction). The level of bias was strongly determined by the range in area across strata, and the number of strata.

3.2 Optimum number of SNA 1 tag recoveries from a single tagging event

Mark-recapture data can provide an estimate of absolute population size. The utility of a single estimate to inform management is largely determined by its precision — highly uncertain tagging estimates have lower utility than more certain ones.

McKenzie et al. (2011) used a three-area SNA 1 tagging programme simulation model to identify designs capable of producing biomass estimates at a range of precision levels. The McKenzie et al. simulator was divided into three spatial areas representing the three SNA 1 stock areas (East Northland [ENLD]; Hauraki Gulf [HAGU]; Bay of Plenty [BPLE]). The “true” recruited population size in each simulated area was set at the upper 95% percentile of the estimated 1999–2000 fishing year spawning stock biomass from the 2000 SNA 1 assessment model (Gilbert et al. 2000). The simulator incorporated inter-area tag movement in accordance with a home fidelity movement dynamic (McKenzie et al. 2011).

Estimator or tag design performance was evaluated in reference to the level of improvement in the coefficient of variation (CV) expressed as Root Mean Square Error (Appendix 2).

McKenzie et al. (2011) showed that the relationship between the number of tag recoveries and CV tended to an asymptote (flatten) at a CV of 0.1, this being the likely practicable upper limit on tagging programme estimate precision. The number of simulated tag recoveries required to achieve CVs in the range 0.1 to 0.3 are given in Table 2.

Table 2: Tag recovery targets (number of tags) necessary to achieve a given CV by SNA 1 stock area as derived from McKenzie et al. (2011) simulations.

Stock area	CV			
	0.1	0.15	0.2	0.3
ENLD	400	200	150	75
HAGU	700	300	200	100
BPLE	400	200	150	75
SNA1	1500	700	500	250

3.3 Utility of multiple tagging events when the net rate of population change is log-linear

The additional value of two or more tagging surveys is in their power to give information on the relative change in stock size through time. There are two useful metrics that can be used to evaluate tagging designs:

1. The precision on the individual tagging estimate;
2. The precision on the estimate of relative change in stock size, as derived from two or more tagging estimates.

Log-linear interpolation through multiple population estimates over time can lead to more precise point-in-time population estimates and greater power to detect relative change in population size between the start and end of the series. For log-linear interpolation to be a valid approach the population over the analytical period has to be either: static; consistently (monotonically) increasing or consistently decreasing.

We used a simulation approach to investigate the power of a log-linear interpolation approach to:

a. Maximise precision on the point-in-time population estimates;
b. Estimate the magnitude of relative change in population size between the start and end of the series.

3.3.1 Methods

Maximise precision on the point-in-time population estimates

For the purpose of the simulation we made the assumption that the lowest tagging programme frequency is 1 survey every 10 years. We looked at the effect of varying the number of surveys in the ten year period up to the maximum number of annual surveys (n=11) and varying the precision on the individual survey estimates.

The "true" biomass in the starting year was arbitrarily set and the biomass in each of the 10 subsequent survey years derived through logistic regression assuming a constant rate of change, i.e., the "true" pattern of population change in the simulations was log-linear.

For each survey year in the simulation, 1000 random survey estimates were generated assuming log-normal error about the "true" biomass pursuant to a given CV. For each individual bootstrap, the expected biomass in all 11 possible survey years could be derived by fitting a logistic curve through the bootstrap survey estimates. The performance criterion used was the improvement in RMSE (expressed as a CV; Appendix 2) on the survey year biomass estimates derived using log-linear interpolation.

Power to detect and estimate relative biomass change

The premise for the simulations was that two programmes spaced 10 years apart should be able to detect a significant change in abundance with greater than 95% confidence under the assumption of log-linear change. It follows that introducing more surveys in the 10 year period should increase the power to detect the change.

Projections from the 2013 SNA 1 assessment (Francis & McKenzie, in press) for the Bay of Plenty provided guidance as to the likely range in stock change to investigate in the simulations. Projecting forward 10 years under the assumption of mean recruitment the Bay of Plenty stock is likely to reduce by 50%, whereas the projection assuming recent recruitment had the stock increasing by 1.5 times; therefore the range of proportional stock-size change investigated was 0.5–1.5. The 10-year stock size ratio changes investigated in the simulations were: 0.5, 0.75, 0.875, 1.125, 1.25, 1.5.

It was assumed that the lowest (best) practicable and feasible tagging survey CV was 0.1, and that the highest acceptable survey CV was 0.30. These CVs defined the range of survey design options (Table 2). We investigated two options for introducing new years in the series: equidistant spacing (Table 3a); and alternate (end-to-end) spacing (Table 3b). The number of surveys investigated in the simulations were: 2; 3; 6; and 11 (Table 3).

Performance criteria used in the simulations were:

a. The proportion of bootstraps that predicted that a change in biomass had occurred in the "correct" direction;
b. The RMSE (expressed as a CV; Appendix 2) on the estimate of relative change between the first and last survey year.

Table 3: Two options for spacing surveys over a ten year period

a. Equidistant spacing

Number of surveys	Survey year 0	1	2	3	4	5	6	7	8	9	10
2	✔	×	×	×	×	×	×	×	×	×	✔
3	✔	×	×	×	×	✔	×	×	×	×	✔
6	✔	×	✔	×	✔	×	✔	×	✔	×	✔
11	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔

b. Alternate spacing

Number of surveys	Survey year 0	1	2	3	4	5	6	7	8	9	10
2	✔	×	×	×	×	×	×	×	×	×	✔
3	✔	✔	×	×	×	×	×	×	×	×	✔
6	✔	✔	✔	×	×	×	×	×	✔	✔	✔
11	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔

3.3.2 Results

1. Precision on the estimate of relative stock biomass change over 10 years

The simulation for two high precision (CV 0.10) tagging programmes was 100% successful at detecting a ±50% change in stock-size and approximately 80% successful at detecting ±12.5% stock-size changes (Table 4; Figure 8).

The simulation for two 0.30 precision surveys was 95% successful at detecting a ±50% change in stock-size but only about 60% successful at detecting ±12.5% stock-size changes (Table 4;Figure 8).

The RMSE CV on the ratio estimate from the two surveys was 0.15 for the high precision (0.1) survey and 0.5 for the low precision (0.3) survey (Table 4; Figure 8).

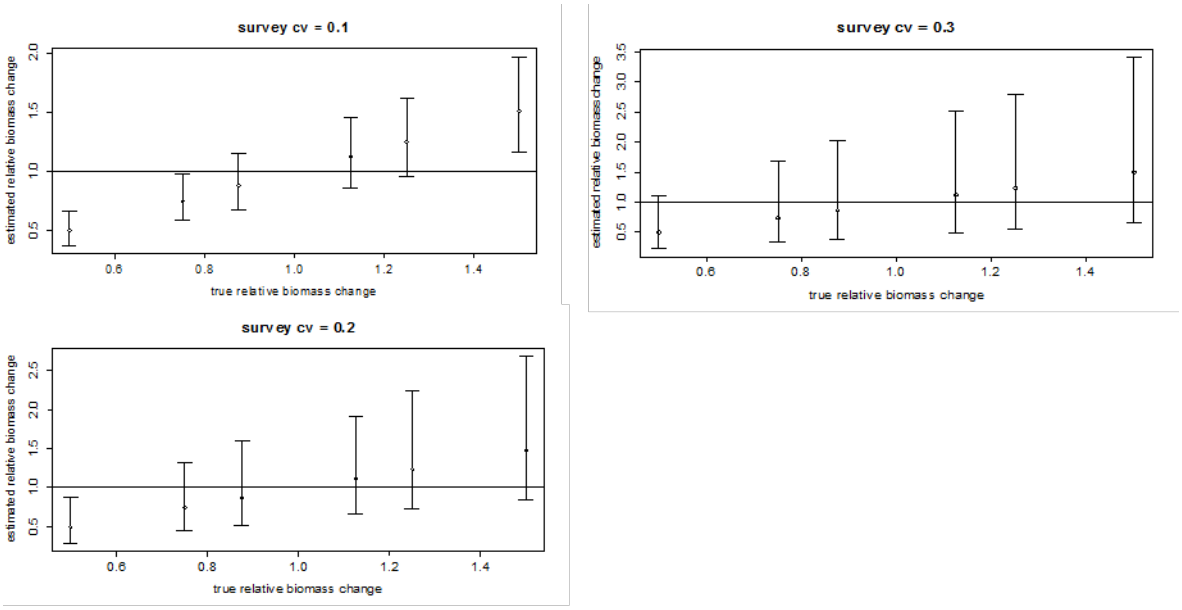


Figure 8: Estimated bootstrap median 10-year biomass change relative to "true" from two 0.10, 0.20 and 0.3 CV surveys showing 95% bootstrap confidence intervals.

Table 4: Probability of detecting a stock size change of given magnitude from two 0.1, 0.2 and 0.3 CV tagging surveys. RMSE CVs on the ratio estimates are in brackets.

True biomass change	Survey precision (CV)		
	0.1	0.2	0.3
0.5	1 (0.15)	1 (0.3)	0.95 (0.5)
0.75	0.98 (0.15)	0.85 (0.3)	0.75 (0.5)
0.875	0.83 (0.15)	0.7 (0.3)	0.63 (0.5)
1.125	0.79 (0.15)	0.65 (0.3)	0.61 (0.5)
1.25	0.95 (0.15)	0.8 (0.3)	0.7 (0.5)
1.5	1 (0.15)	0.93 (0.3)	0.83 (0.5)

Increasing the number of surveys over the 10 year period increased the RMSE precision on the ratio estimate (Table 5). Alternate end spacing is slightly more effective than equal spacing for estimating the relative change ratio, with minimal improvement in precision between 6 and 11 tagging programme scenarios (Table 5).

Table 5: Improvement in RMSE CV on the predicted 10-year biomass ratio by increasing the number of surveys over the 10 year period (individual survey CV 0.10).

Number of surveys	Equidistant spacing	Alternate spacing
2	0.15	0.15
3	0.14	0.13
6	0.12	0.10
11	0.10	0.10

2. Precision on the annual biomass estimates as derived from log-linear interpolation

Using log-linear interpolation precision on the first and last year survey estimates is also significantly improved by adding more surveys, the biggest precision gain being achieved by adding a third survey (Table 6).

Table 6: RMSE CV on the initial survey year (year 0) biomass estimates derived using log-linear interpolation relative to the total number of alternately spaced surveys undertaken and the precision on the individual survey estimates (CV).

Number	Survey CV			
of surveys*	0.1	0.15	0.2	0.3
2	0.10	0.15	0.20	0.30
3	0.07	0.11	0.15	0.23
4	0.07	0.11	0.15	0.23
5	0.07	0.11	0.15	0.23
6	0.07	0.10	0.13	0.19
7	0.06	0.09	0.13	0.19
8	0.06	0.09	0.12	0.17
9	0.06	0.09	0.12	0.17
10	0.06	0.08	0.11	0.16
11	0.06	0.08	0.11	0.16

* Alternate spacing

Log-linear interpolation provides biomass estimates for all years over the tagging interval, and not just the survey years. The bootstrap simulations resulted in higher RMSE CVs on the intermediate interpolated (non-survey) years than on the start and end (true-survey) years (Table 7). For two surveys of 0.10 precision spaced 10 years apart the RMSE CV on the interpolated biomass was highest for the mid-point year (0.07 year 5; Table 7). This relationship was found to be independent of the time period between the two tagging estimates, e.g., the RMSE CV on the predicted mid-point biomass estimate was always 0.07 for surveys of precision 0.10.

Table 7: RMSE CV on all interval year predicted biomass derived from log-linear interpolation through start and end year survey estimates of 0.10 (CV) precision (shaded).

Survey year	CV (RMSE) on ratio estimate
0	0.10
1	0.09
2	0.08
3	0.08
4	0.07
5	0.07
6	0.07
7	0.08
8	0.08
9	0.09
10	0.10

Increasing the number of surveys over the 10 year period increased the precision on the interpolated survey estimates, including the individual survey year estimates, when recalculated as interpolated values (Figure 9). Alternately spaced surveys generally resulted in slightly higher precision in the start and end period, than equidistant surveys, when the number of surveys was even (compare alternate and equidistant RMSE plots for six surveys in Figure 9) — the exception being three surveys, where equidistant spacing appears preferable to alternate (compare alternate and equidistant RMSE plots for three surveys in Figure 9).

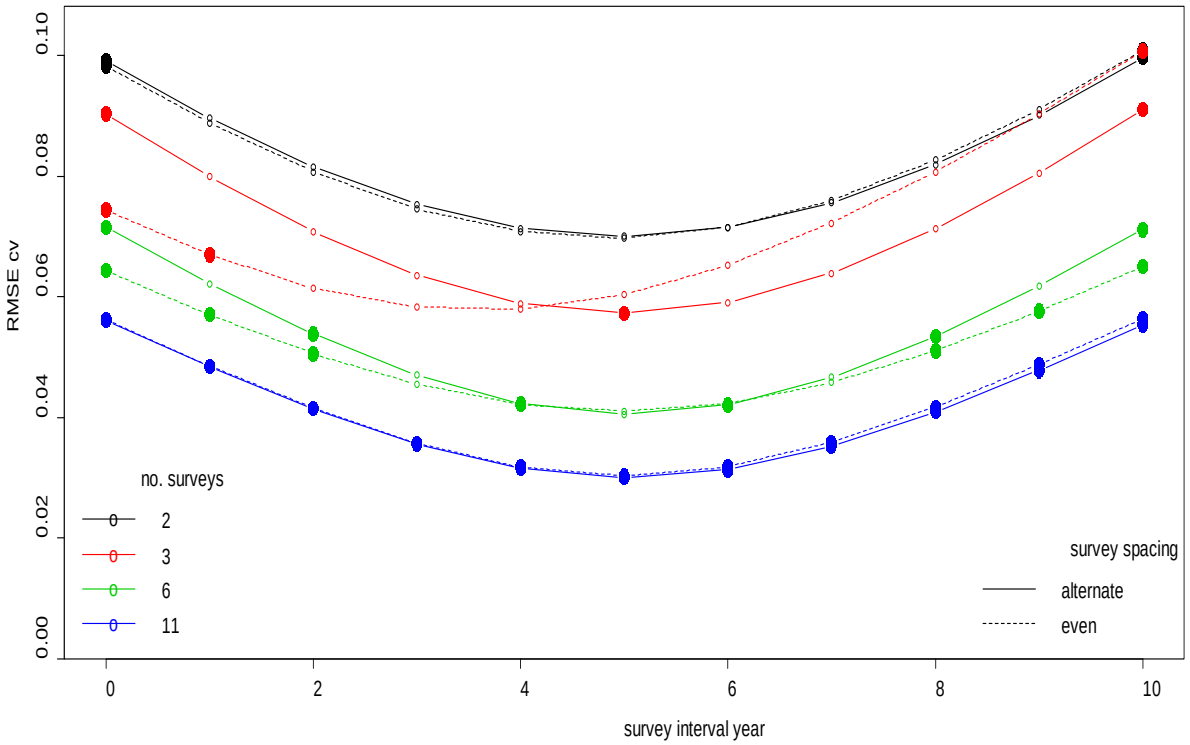


Figure 9: RMSE CV on interpolated biomass estimates through 0.10 precision survey estimates (solid points denote actual survey years); solid lines alternate and dashed lines equidistant (even) survey spacing.

3.3.3 Validity of the log-linear population change assumption for SNA 1 and SNA 8

All the above interpolated precision conclusions are predicated on the assumption that the rate-of-change over the intervening period between surveys is approximately log-linear. As the biomass trajectories from the recent SNA 1 assessment (Francis & McKenzie in press) are largely inconsistent with a 10-year log-linear rate of change hypothesis, we conclude that it is inadvisable to adopt a 10-year tagging strategy for SNA 1 (and 8) based on linear-change optimisations. The question then becomes: what would be an alternative interval over which a log-linear analytical approach would be valid? A useful way of addressing this question is to firstly determine the minimum time period over which a multiple tagging approach could be "usefully" applied in SNA 1 and SNA 8, and then to determine if the log-linear rate-of-change hypothesis is "reasonable" over this interval. Results given in Table 6 suggest that the minimum number of surveys needed to increase the precision on the biomass estimate through log-linear interpolation is three. Three annual tag release events could be conducted over the space of two years in SNA 1or SNA 8. The rate of change seen in the Francis & McKenzie SNA 1 biomass trajectories at the scale of two-years is not inconsistent with a log-linear hypothesis, and hence a triplet tagging strategy would be valid for obtaining a point-in-time biomass estimate in SNA 1. Given current knowledge, it is difficult to provide an objective basis to test the log-linearity assumption for SNA 1 at the scale of two years, and alternative approaches may lead to different conclusions.

Pursuant to a requirement of obtaining a point-in-time SNA 1 population estimate of precision 0.1, there are two options for achieving this:

1. One survey with precision CV = 0.10;
2. Three consecutive surveys conducted over two years of sufficient precision to achieve RMSE on the middle survey close to CV = 0.10 as derived through log-linear interpolation.

The simulation results indicate that three 0.2 CV surveys are likely to provide an estimate on the middle survey year of close to 0.1 through log-linear interpolation (RMSE CV 0.11; Table 8).

Table 8: RMSE CV for biomass estimates relative to three equidistant spaced time periods as derived by log-linear interpolation through 2 and 3 precision 0.20 (CV) survey estimates.

Number of surveys	Survey year 0	1	2
2	0.20	0.14	0.20
3	0.18	0.11	0.18

3.4 A cost-benefit analysis of SNA 1 tag designs

3.4.1 SNA 1 tagging programme design cost/benefit analysis tool

McKenzie et al. 2012 showed that there was a direct inverse relationship between the number of tag releases required to achieve a given number of tag recoveries, and the number of fish examined (or scanned) for tags. Thus there are potentially a broad range of programme designs capable of meeting tag recovery targets, but there is likely to be only one optimum design for each target in terms of cost. However, the optimum cost design depends on the relative cost of releasing verses recovering a tag, and cannot therefore be determined unless these specific cost trade-offs are known.

As part of this project we developed a cost-benefit analysis tool that would enable managers and stakeholders to come up with the optimum programme design for achieving a SNA 1 population estimate for a given target precision and to provide an overall cost for the programme.

The tool performs a cost benefit optimisation; maximising the number of tag recoveries and minimising the cost trade-off between the number of tags released and the number of fish scanned for tags. The input table used to describe the relationship between releases, recapture and scanning numbers (Appendix 3) comes from McKenzie et al. (2011). The tool requires the desired number of tag recoveries from each of the three SNA 1 sub-area stocks (East Northland, Hauraki Gulf, and Bay of Plenty) as input, as well as the choice between multiple and single release event tagging approaches. The multiple tag optimisation routine takes into account that scanning after the first release year is likely to recover tags from earlier release events. The optimisation process also factors in the loss of tags from the population due to a user specified total mortality (i.e., fishing and natural). In the example runs below the loss of tagged fish from the population from total mortality one year after release was assumed to be 30%.

The user is required to provide fixed and variable tag and scanning related costs (see Appendix 4 for an example of the cost specification input table).

3.4.2 Some important cost considerations governing the use of the tool

As a note of caution; the main purpose of cost-benefit simulation tool is to provide relative costs pursuant to a set of fairly broad design criteria; it should not be used to predict precise costs for use in detailed financial planning. However, provided with a set of reasonable cost parameters the simulator is likely to provide cost estimates within ±$125 000–$250 000 of the "true" cost. Although not precise, these estimates allow comparison of the relative trade-offs of different approaches and methods under different assumptions for various tagging and recapture scenarios.

The simulator works by optimising the required number of tags released against the scanning targets for a given target number of recoveries. Thus there are only two unit costs: the mean cost of tagging a fish; and the mean cost of scanning a fish. This simplifies the costs down to a few variables, but note that

there are likely to be many component costs going into the derivation of the mean per-tag release and scan costs. Some of these will relate to the physical cost of tagging and scanning, while others will relate to the need to test for mitigating bias (see Section 4), e.g., need for good spatial coverage during the release and scanning phases.

3.4.3 Optimum designs relative to genetic and PIT tagging costs

We were unable to provide absolute or precise estimates of the total costs for the genetic and PIT tagging programme designs, as many of the component costs (e.g., vessel charter costs) were unknown. Release costs from the 1994 SNA 1 tagging programme and PIT tag and scanning costs from the 2001 SNA 8 tagging programme put the cost ratio between releasing an individual fish and scanning an individual fish at 1.00:0.02. Our gene-tagging investigations (Section 2.3) suggest that the release and recovery cost differential is in the order of 1.00:0.55. Applying these ratios in a design optimisation produced different release and recovery strategies for PIT and genetic tagging programmes; genetic tagging would require marking in the order of five times the number of fish required by PIT tagging and conversely sampling (scanning) five times fewer fish (Table 9).

Table 9: PIT and genetic 500 tag recovery tagging programme design optimisation as determined by the respective tag:scan cost ratios.

Tag type	No. tags released	No. fish scanned (% TACC)	No. tags recovered	Cost ratio tag:scan
PIT	33 500	1 700 000 (32%)	500	1:0.02
Genetic	174 000	330 000 (6%)	500	1:0.55

3.4.4 Relative cost of 1500 recovery and 500 recovery PIT tag programmes

The optimal designs for achieving 500 (precision 0.2) and 1500 (precision 0.10) SNA 1 tag recoveries for a PIT tag:scan cost ratio (1:0.02) are given in Table 10. The optimisations showed that tripling the number of target recoveries only increased release and recovery targets by a factor of 1.73 (Table 10). This result illustrates the relationship between tagging programme design cost and the number of tags recovered is not 1 to 1. As a generality, tagging programme costs increase as a square-root of the increase in the predicted number of recovered tags, e.g., $\sqrt{3}$ or approximately 1.73 (see Appendix 5).

Table 10: Predicted release and scanning targets required to achieve a given number of SNA 1 PIT tag recoveries. Costs (000's) are derived relative to unit cost of tagging one fish being 1.

	No. tags recovered	No. tags released	No. fish scanned (% TACC)	Scan cost	Tag cost	Total cost
	500	33 500	1 700 000 (32%)	33	33	66
	1 500	58 000	3 000 000 (54%)	58	58	116
Ratio	1:3	1:1.73	1:1.73	1:1.76	1:1.76	1:1.76

3.4.5 Relative cost comparisons between single 1500 and triplet 500 PIT tagging strategies

In Section 3.3.2 it was shown that three (triplet) consecutive CV = 0.2 precision surveys (500 recoveries) would provide a point-in-time tagging estimate of equivalent precision to one CV = 0.1 precision survey (1500 recovery), assuming that the rate of population change over the two years that the three surveys were conducted is approximately log-linear.

From the equations given in Appendix 5, the cost of conducting three 500 tag recovery programmes is approximately 1.73 times greater than the cost of a single 1500 tag recovery programme. However this simple calculation does not take into account cost savings due to the scanning effort for the later surveys also contributing to tag recoveries from the earlier release events. There are also likely to be savings in the second and third survey years due to the reduced set-up costs and efficiencies of running the operation over longer time period.

To gain an appreciation of the level of savings that might be expected in running three sequential tagging programmes we ran the following cost optimisations:

Optimisation 1: annual proportional tag loss at start of the second and third scanning year after tagging = 0.3;
Optimisation 2: tag loss as above; 25% reduction in tagging and scanning costs after year 1.

Consistent with Appendix 5 predictions, three independent 500 tag PIT programmes cost 71% more than a single 1500 tag programme (Table 11). In contrast, a triplet approach that achieved savings in the number of fish needed to be scanned in the second and third years (assuming that annual tag loss is 0.3) cost 33% more (Table 11). Achieving a 25% reduction in the tagging and scanning costs after year 1 brought the comparative cost of triplet strategy down further to within 10% of the cost of a single larger programme (Table 11).

Table 11: Cost of various 500 tag single and triplet PIT tagging programme strategies expressed as a % of a single 1500 tag recovery programme. Costs (000's) are derived relative to unit cost of tagging one fish being 1.0 in the first survey year.

Strategy	rectag	reltag	Nscan	Scan cost	Tag cost	Total cost	% cost of single 1500
1 × single 1500	1 500	58 000	2 915 200	58	58	116	100%
3 × single 500	1 500	100 700	5 050 000	99	99	198	171%
Triplet 500 tag loss (opt. 1)	1 500	100 700	2 700 000	55	99	154	133%
Triplet 500 tag loss + 25% reduced costs after year 1 (opt. 2)	1 500	100 700	2 700 000	43	83	126	109%

For reasons given in Appendix 6, the minimum cost tagging programme design is always one in which the variable scanning and tagging component costs are equal. This means that the comparison results in Table 11 will be the same for any unit tag:scan cost ratio, i.e., the scanning cost savings expected under a triplet design will be the same for all tag types and scanning technologies.

3.5 Conclusions and recommendations

Under an opportunistic tagging approach, low catch-rate areas and areas where the commercial fishery have been excluded (e.g., inner Hauraki Gulf) are likely to receive very few or even no tags. We believe that the potential level of negative bias (under estimation) associated with an opportunistic tag release strategy in SNA 1 and SNA 8 could be in excess of 25%. There is also no way of determining the magnitude of the bias. For this reason we do not recommend opportunistic tagging as a release strategy for SNA 1 and SNA 8.

Instead we recommend that spatial allocation of tags is carried out using a stratified design based on an *a priori* understanding of spatial abundance, and that the number of tags released in each stratum be monitored and controlled. This approach does not necessarily preclude opportunistic tagging by the commercial and recreational sectors that is carried out in conjunction with dedicated tagging (e.g., research vessel or science charter) in areas that would not be adequately covered by opportunistic tagging.

If the pattern of change in the fishery is log-linear (i.e., stable or monotonically increasing/decreasing) then the simulation results provide an indication of the number of surveys needed to detect population change. However, as a generality, the more unpredictable the pattern of change in the stock the greater the number of surveys needed to describe it. Therefore the optimum frequency of tagging events cannot be determined without an understanding of the degree of variability in population change the surveys are intended to monitor. Although a series of tagging estimates would provide an insight into the pattern of relative abundance change in SNA 1 and SNA 8, there are significantly less expensive options than tagging for monitoring relative stock abundance, e.g., trawl surveys or CPUE. We recommend that a key consideration for evaluating SNA 1 and SNA 8 tagging designs should be the precision of the point-in-time absolute population estimate, and the precision on the estimate of relative stock change between two independent point-in-time absolute tagging estimates.

The simulation results from McKenzie et al. (2011) indicate that the highest practicable estimate of precision achievable through tagging in SNA 1 is a CV = 0.10, which equates to about 1500 tag recoveries. The 1994 SNA 1 tagging programme achieved about 530 recoveries. The McKenzie et al. simulation results suggest that a comparable programme to the 1994 SNA 1 programme would produce an estimate of precision CV = 0.20.

Obviously a CV = 0.1 is preferable to CV = 0.20 if precision were the only consideration, but does the improvement justify a factor of 1.73 increase in cost? Similarly, 1500 tag recoveries will provide greater insight on bias and movement, but would the information warrant the additional cost? It is possible to gain some appreciation of the relative value of the higher precision result by looking at its power to address some of the fundamental uncertainties from the 2013 SNA 1 assessment.

Despite the biomass signal in 1994 tagging data being over twenty years old; by anchoring the model biomass trajectory in 1994, these data strongly influenced the 2013 SNA 1 stock assessment estimate of current stock status. The level of precision on model estimates of stock status from the 2013 assessment was relatively high (in the order of CV = 0.1 based on the published confidence intervals; Francis & McKenzie in press). It is likely that there would have been little to be gained, in terms of the SNA 1 2013 assessment's overall precision, had the 1994 tagging programme been precision CV = 0.1.

However, the main uncertainty of 2013 SNA 1 assessment was not a lack of precision, but rather the structural uncertainty in the model — specifically the spatial extent of the three SNA 1 sub-stocks and the degree of mixing between them (Francis & McKenzie in press; MPI 2013). The information available to the 2013 SNA 1 assessment model to estimate sub-stock movement rates was the 1985 and 1994 tagging data, and the level of interchange between the Hauraki Gulf and Bay of Plenty was only described by the 1994 tagging programme data.

The 2013 assessment model prediction that the Bay of Plenty sub-stock in 2013 was below the hard-limit of 10% virgin biomass was strongly determined by only 40 tag observations from 1994. On the basis of structural uncertainty in the 2013 SNA 1 model, the Northern Inshore Fisheries Assessment Working Group (and the Plenary meeting) recommended combining the Bay of Plenty and Hauraki Gulf model yield estimates. A tagging programme design with CV = 0.1 would be likely to yield three times the number of tag recoveries than the 1994 SNA 1 tagging programme, and consequently have three times the "power" to estimate movement and describe stock spatial structure.

As discussed in Section 3.1, failure to distribute tags in the population in proportion to spatial abundance can result in biased population estimates. Despite the use of stratified release designs, evidence of heterogeneous or unequal mixing of tags was present in all previous SNA 1 tagging programmes (Gilbert & McKenzie 1999; Section 3.1.2), suggesting that biomass estimates from past tagging programmes are likely to be negatively biased. A common inherent limitation in past SNA 1 tagging programmes has been that there was an insufficient number of tag recoveries to detect, estimate, and correct for spatial bias. Another significant area of uncertainty in the 2013 SNA 1 assessment was the degree to which weighting (importance) placed on the tagging data allowed for possible negative biases. The need to test-for and estimate bias in future tag recovery data *a posteriori* is another strong justification for choosing a higher precision tagging design than was achieved in the 1994 SNA 1 programme.

Under the triplet tagging design, tags from the first release event are observed over three consecutive annual cycles, and those from the second event observed over two annual cycles. There are a number of advantages in the triplet tagging approach for SNA 1 and SNA 8, however, the triplet approach assumes that the rate of biomass change over the two year period the three surveys are undertaken is either negligible and/or approximately log-linear (monotonic). Given the stock dynamics for snapper in SNA 1 and SNA 8, this assumption would appear reasonable.

A triplet tagging programme has an advantage over a single release event high precision programme in that recoveries during the earlier years can be used to modify the release design of the later release events, potentially allowing for management and reduction of spatial allocation bias. Triplet event tagging is likely to provide better insight into seasonal movement dynamics than a single release limited to one seasonal cycle of recovery observations.

Anticipated cost savings in running three consecutive tagging programmes means that the additional cost of the triplet design relative to a single high precision survey will be less than three separate studies, and could be as low as an additional 10%. Furthermore these costs would be spread over four years instead of two, potentially resulting in a lower financial burden on funding stakeholders in each tagging year.

In addition to population estimates, tag recovery data from multiple tagging events over multiple years can be used to derive total mortality estimates for the tagged stock (Seber 1982; Brownie et al. 1985), this being another strong justification for adopting a triplet tagging approach in SNA 1 and SNA 8. Using a number of recognised 'integrated' analytical approaches it may be possible to disentangle the fishing and natural mortality components in the triplet tag survival data (see Polacheck et al. 2010 for a review of integrated tagging models).

We see the main disadvantage of the triplet strategy is that the final high precision estimate would not be available until 3–4 years after the first tagging event as opposed to 1–2 years for the single survey approach. However, under the triplet strategy the biomass estimate from the first survey would likewise be available to managers after 1–2 years which, although of lower precision, is still likely to have some management utility to consider interim management actions.

4 METHODS TO ADDRESS KNOWN SOURCES OF BIAS

An assumption often made in tagging analyses is that tagged and untagged animals behave and act in the same way. However, if this assumption is violated, then biases in resulting estimates may be introduced. Many fish tagging studies have shown that tagging can alter the behaviour of some animals; for example the tagging of snapper, Antarctic toothfish, and Patagonian toothfish has been shown to impact growth rates of individuals.

A key metric derived from tagging data is the ratio of marked to unmarked animals (mark-rate; Section 1.2) in the underlying population. Observations of tagged animals are typically made over a wide time interval after release, it is assumed that the initial mark-rate remains constant through time. Many studies have assumed that the tagged and untagged fish mix equally, or that there is a constant probability of recapturing a tagged fish over space and time. A more complex model was used for the previous snapper assessment. This assumed that while the probability of recapturing a tagged fish was constant within a sub-area, recapture probabilities varied between sub-areas and over time. In addition, individual effects (e.g., size-based tag mortality, growth impacts, or other behavioural changes) may also be important. Most of the analytical complexity in mark-rapture programmes involves estimating and correcting factors that alter the underlying population mark-rate (McKenzie et al. 2011).

4.1 Initial tagging survival

Because the number of tags in the population has a direct and proportional relationship to the tagging population estimate, it is important to know what proportion of fish survive the tagging process, i.e., the effective number of tags in the population at time of release. However, it is not possible to estimate initial tag survival from the tag recovery rates alone; initial survival estimates can only be derived from independent observation or experiment. Furthermore, there is evidence that these initial mortality rates change with size, depth, and capture method (McKenzie & Davies 1996).

There have been three large net holding studies conducted since 1992 specifically to quantify initial mortality of tagged snapper under a range of capture treatments (MPI unpublished data). Each of the three studies made use of similar replicated factorial designs. Treatment classes investigated were combinations of the capture method (trawl; longline), capture depth, fish length, and total catch weight. In each experiment fish were held for observation in large sea-cages for three weeks. As there was little additional mortality observed during the third week in any of the surveys it was likely that the combined initial trauma mortality and mortality through secondary infection and disease were represented in the data. The main difference between the three studies was the type of tags used. In the first study external dart tags were used, the second study used internal coded wire tags (McKenzie & Davies 1996), and the third used plastic coated PIT tags injected into the peritoneal cavity (Gilbert & McKenzie 1999).

The three mortality studies were designed to estimate mortality in snapper specific to the tagging and handling practices used in SNA 1 and SNA 8 tagging programmes conducted after 1990. The combined survival observations from these studies (more than 2000 observations) is likely to adequately describe initial mortality in the post 1990 snapper tagging programmes There would be little additional value in conducting further mortality studies (collecting more observations) to improving a future SNA 1 or SNA 8 tagging programme biomass estimate if the programme used similar tagging and handling practices.

The statistical methods used to derive survival estimates from the snapper mortality data date from the mid-1990s (McKenzie et al. 1996; Gilbert & McKenzie 1999). Data from the most recent of the three mortality studies was reanalysed by Rashid (2014) as a part of a Masters dissertation on survival estimation. Using improved statistical methods, Rashid was able to account for a greater degree of variation in the data. Although the Rashid mortality-rate estimates were not markedly different to those previously derived due the use of superior analytical procedures, the Rashid estimates are likely to be more accurate and precise. It is recommended that for future SNA 1 and SNA 8 tagging programmes

methods used for estimating initial mortality should be reviewed in light of results reported by Rashid (2014).

4.1.1 Options for improving tag release survival in snapper

Previous SNA 1 and SNA 8 tagging programmes have used both longlining and trawling methods to capture and release tagged fish. Barotrauma caused by expanding swim-bladder gasses inside the body cavity has a negative effect on snapper survival from both capture methods. However, longlining was found to inflict markedly lower initial mortality than trawl due to the absence of crushing and herding related trauma (McKenzie et al. 1996; Gilbert & McKenzie 1999; Rashid 2014). Where possible, longlining should be used in preference to trawl for tagging snapper due to the lack of herding and crushing stress. However, over large flat bottom areas, such as the outer Hauraki Gulf and eastern Bay of Plenty, the use of longline can be inefficient and impractical. Although higher numbers of released snapper die from trawl, a higher number of fish can be tagged per fishing day such that the "effective" number of surviving tags in the water achieved per charter day is likely to be similar between the two methods. In the 1994 SNA 1 tagging programme the ratio of longline to trawl release snapper was approximately 4:1. For any future SNA 1 tagging programme the spatial effectiveness of the two methods should be reviewed in the light of current commercial fishing practices. Longlining is currently not practical as a commercial fishing method in SNA 8 due to the presence of strong currents and high numbers of shark predators. Trawl, as a release method, has been used exclusively on all SNA 8 tagging programmes to date.

In June 2014 a new precision trawling technology was successfully trialled by the New Zealand fishing industry. This new technology, termed the "precision seafood harvest" (PSH) device, is intended both to reduce incidental mortality in discarded undersized fish and to deliver fish in better condition for market. As of early 2015, the PSH device was still undergoing trials and hence not available for commercial use. However, if substantive reductions in discard mortality in snapper can be achieved then the PSH device would be ideally suited for use in future SNA 1 and SNA 8 tagging programmes. Use of the PSH device for capturing and tagging snapper would, however, necessitate conducting release mortality studies to derive mortality estimates specific to this technology.

The negative effects of depth related barotrauma on the survival success of released fish are well documented in the literature (Feathers & Knable 1983; Muoneke & Childress 1994; Wild 2009). There is reasonable evidence to suggest that if fish can be quickly returned to depth, thereby alleviating the surface barotrauma effects, survivorship is significantly improved (Muoneke & Childress 1994; Hannah et al. 2008; Sumpton et al. 2010; Rudershausen et al. 2014; Parker et al. 2006). Venting with hypodermic syringes was universally applied in past snapper tagging programmes as a means to allow fish to quickly return to depth. The more recent evidence suggests that returning fish to the seabed using shot lines or cages may be preferable to venting (St John et al. 2009), possibly because the venting itself may induce added trauma and stress through physical injury and the time delay associated with the procedure. In light of the fish survival studies published since the last snapper tagging programme, a review of fish tagging and handling and release methods would need to be undertaken. As with the introduction of the trawl PSH technology, the adoption of new handling and release procedures (e.g., shot descent lines) would require undertaking further experimental release mortality studies to quantify post-tag survival.

4.1.2 Tag-release requirements may limit commercial fishing operations under opportunistic tagging strategies

Previous SNA 1 and SNA 8 tagging programmes adopted a set of fish handling procedures which involved the use of specialised equipment. These were intended to minimise the level of trauma and infection in tagged snapper and to ensure a high level of data accuracy. These procedures included:

- stringent sterilisation and cleaning procedures for tagging, venting and measuring equipment;

- salt-water holding tanks to retain fish prior to tagging;
- water reticulation and oxygenation systems;
- low numbers of fish simultaneously held on board for tagging;
- training courses for tagging and recording staff;
- customised tagging tables, rubberised measuring boards, and gloves;
- electronic measuring boards and data acquisition systems.

As stated above, we recommend snapper tagging equipment, handling and release requirements be reviewed for each new SNA 1 or SNA 8 tagging programme in light of new developments. However, the tagging procedure is likely to take up a considerable amount of time on a vessel, and it may not be feasible to tag fish during and as a part of commercial fishing sets or tows. It is likely that such tagging will require conducting specific fishing events that optimise for post tag survival, and include the use of shorter sets or tows, smaller catches, and additional time delays while fish are tagged.

4.2 Limitations to tagging snapper in deep water

The high incidence of barotrauma and the lower catch-rates from deeper waters mean that it is unlikely to be viable to tag snapper taken from depths greater than 75 m. The amount of bias caused by the lack of tagging beyond 75 m was investigated in the 2001 SNA 8 tagging programme using directed (contract) fishing beyond the 75 m contour and across the length of SNA 8. The results yielded very few recaptured tags but also very low catch. The conclusion from this study was that there was only a small biomass deeper than 75 m. Hence the 2001 estimates that ignored fish deeper than 75 m were likely to be representative of abundance of the SNA8 stock (Davies et al. 2006).

However, it is not likely that the same conclusion would be made for SNA 1. Evidence from the SNA 1 fishery shows that snapper can occur in commercial quantities out to 120 m depth. The inability to tag beyond 75 m means that there are likely to be areas of significant biomass where tagged fish cannot be released alive. Past SNA 1 tagging analyses (Sullivan et al. 1988; McKenzie & Davies 1996) assumed that movement of tagged fish from shallower to deeper waters was homogenous (based on the knowledge that snapper move inshore to spawn in spring/summer i.e., when tagging took place). Although tagged fish have been recovered from beyond 75 m, it has not been possible to estimate the interchange between shallow and deep water areas because of the lack of reciprocal movement observations (i.e., deep to shallow). This means that it was difficult to differentiate between the hypotheses that there was either a high degree of movement of inshore fish to a relatively large off-shore population, or a low degree of movement to a small off-shore population.

In order to determine which of these hypotheses may be correct, an independent measure of relative abundance between on and off-shore strata will be required. It is possible that information from commercial and recreational catch-and-effort indices can provide this. Here, the relative biomass of snapper in a given depth stratum can be assumed to be the product of CPUE (catch rate) and stratum area, i.e., catch-rate × stratum-area = abundance. Thus there are two components to the SNA 1 relative biomass analysis:

1. Derive a CPUE score for the stratum-area (in spring/summer)
2. Determine the area of snapper habitat in the stratum.

A strong assumption required in the use of these data is that the spatial distribution of commercial and recreational catch and effort observations is reflective of the relative distribution and spatial extent of the SNA 1 stocks. A specific important assumption is that the relative stock biomass in areas where little or no snapper catch is reported or observed is effectively close to zero.

Approximately 60% of the annual commercial SNA 1 catch is taken by longline and approximately 30% is taken by trawl (Walsh et al. 2014). Both of these methods have provided event-based catch effort reporting at a fine spatial scale since October 2007. Event based catch reporting by these methods means

that approximately 80–90% of the total SNA 1 commercial catch can be spatially assigned at about 1 nautical mile resolution.

The spatial distribution of SNA 1 trawl and longline catch events in November, December and January of fishing years 2007–08 to 2012–13 is shown in Figure 10 for shallow (less than 75 m) and deep (more than 75 m) water, in East Northland, Hauraki Gulf and Bay of Plenty. Longline and trawl catch and effort data were obtained from the Ministry for Primary Industries commercial catch reporting system for the period October 2007 through to September 2013. The basic data extract criteria was for all effort details and associated catch weights (all species including snapper) from all trips landing SNA 1 catch. The spatial extent of the Hauraki Gulf region deeper than 75 m is small, likewise the proportion of commercial catch (less than 1%; Table 12), and we conclude from this that negligible bias would be introduced by not tagging beyond 75 m in the Hauraki Gulf.

In contrast, significant proportions of the East Northland and Bay of Plenty areas of SNA 1 lie beyond 75 m and the proportion of commercial catch taken from these depths is between 15–20%. (Figure 10 and Table 12). We conclude that a failure to tag deeper than 75 m in the East Northland and Bay of Plenty has the potential to bias the tagging biomass estimates. Therefore some estimate of the relative biomass of the snapper populations beyond 75 m in these areas at the time tagging occurs would be a requirement of a future SNA 1 tagging programme.

Analyses were undertaken using the post 2007 spatially disaggregated trawl and longline data and data from MPI recreational harvest surveys (Hartill et al. 2013) to:

a. Determine the likely proportion of East Northland and Bay of Plenty snapper resident beyond 75 m during spring-summer (i.e., the period when tagging is likely to occur).
b. Determine the utility of these data for "calibrating" the tag observational data from a future SNA 1 tagging programme where no tagging takes place in these areas beyond 75 m.

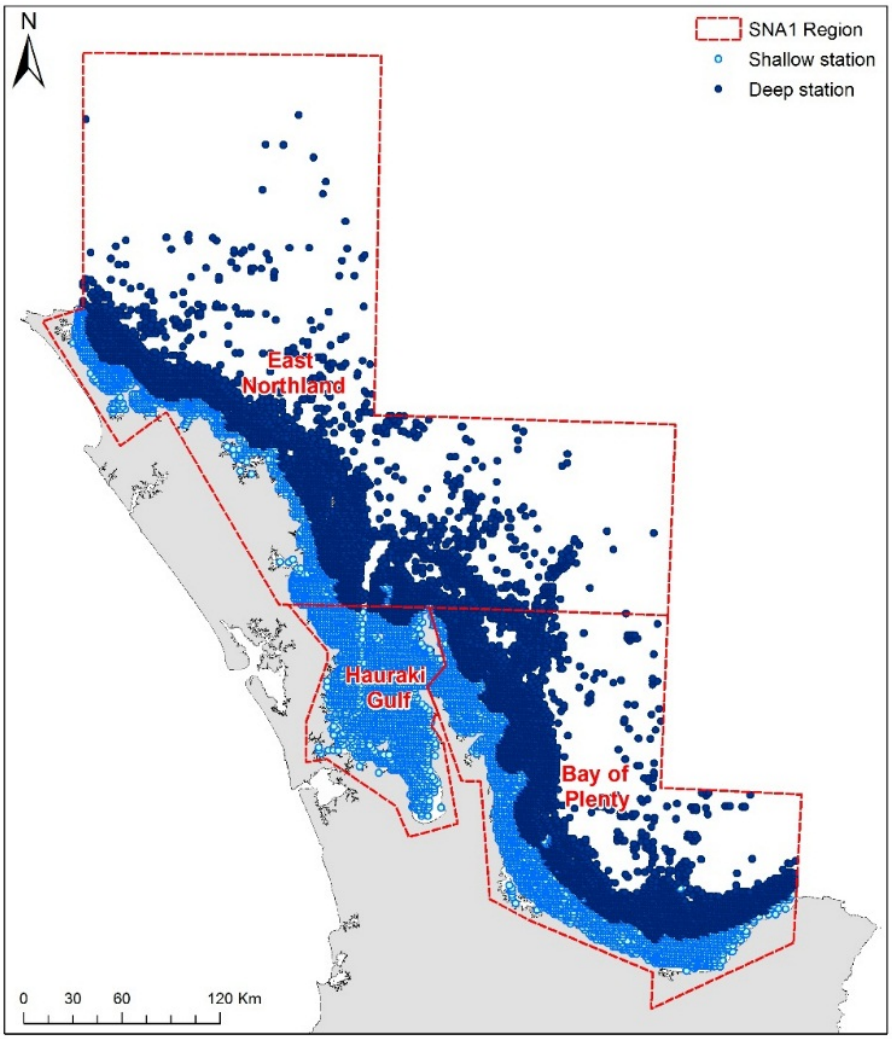


Figure 10: Spatial distribution of SNA 1 catch (t) in 2007–08 to 2012–13 fishing years in months November, December and January in shallow and deep waters, where shallow means depth no more than 75 m and deep means depth greater than 75 m.

Table 12: SNA 1 catch (t) in 2007–08 to 2012–13 fishing years in months November, December and January in shallow and deep waters by fishing method and region.

	East Northland		Hauraki Gulf		Bay of Plenty	
Method code	Shallow	Deep	Shallow	Deep	Shallow	Deep
BLL	782.5	87.4	1833.6	1.9	235.9	36
BPT	36	4.7	14.8	0	0.7	1.4
BT	119.8	102.7	1274.2	29.9	712.5	99.3
DS	0.9	2.2	300.3	1	104.9	22
HL	0	0	0.1	0	0.2	0
PS	0	0	0	0	1.3	0
SN	26.1	1.9	4.4	0	5.2	0.8
Total	965.4	199	3427.3	32.8	1060.6	159.4

4.2.1 Depth stratified CPUE analysis

Methods

The utility of longline CPUE for estimating relative annual abundance in SNA 1 is well-accepted (Francis & McKenzie in press). Longline catch and effort data were used to derive depth-related spatial CPUE indices for East Northland and Bay of Plenty (two indices). Data for the analysis came from the Ministry for Primary Industries commercial catch reporting system and covered the period October 2007 through to September 2013. The basic criterion for the data extract was for all effort details and associated catch weights (all species including snapper) from all trips landing SNA 1 catch. Data analysis was restricted to the months of November, December and January as these were deemed to be the optimum months for tagging in SNA 1 to take place.

Core vessels were selected for each of the fishing areas ENLD and BPLE using the method described by Kendrick & Bentley (2011). This method attempts to restrict analysis to vessels that were consistent participants in the fishery. This was done by producing graphics that allowed us to simultaneously assess the number of trips per year and the number of years in the fishery, hence minimising the number of vessels selected and maximising the proportion of the total catch that those vessels accounted for. The proportion of catch explained by the vessels that met the number of trips per year should be above or reasonably close to 60%.

Data grooming was undertaken to address two types of error: missing values and extreme values (outliers). In the case of missing response variables (catch), the whole record was deleted. Where covariates (e.g., effort_num, effort_width and total_hook_num) were missing or erroneous, the covariate value would be replaced with the median of the covariates from the same vessels.

To provide more power to determine how abundance is changing with depth the data was assigned to 6–7 depth bands to an upper depth limit such that the catch from water deeper than that limit was less than 3% of the total catch. The main consideration for the depth band definition was to ensure that there were sufficient catch data for a comparative CPUE analysis.

Catch indices for depth bands (assumed to represent snapper availability) were derived using generalised linear modelling (GLM) procedures (Vignaux 1994; Francis 1999). The GLMs were conducted using the statistical software R. The response variable in the GLM was log catch. Depth band was entered as a categorical covariate (explanatory) term on the right-hand side of the model. Standardised CPUE abundance indices (canonical) were derived from the exponential of the depth band covariate terms as described in Francis (1999).

The approach taken with all the GLMs was to enter fishing ‘effort’ as a covariate (i.e., “right-hand” model term), with logged catch as the regression variable. This is algebraically analogous to subtracting effort from catch in log-space.

In order to accommodate a non-linear relationship with the response variable (log catch) all continuous variables (including effort terms) were offered to the GLMs as third order polynomials. A forward fitting stepwise multiple-regression algorithm was used to fit GLMs to groomed catch and effort data. The stepwise algorithm generates a final regression model iteratively and uses a simple model with a single predictor variable, depth band, as the initial or base model. The reduction in residual deviance relative to the null deviance is calculated for each additional term added to the base model. The term that results in the greatest reduction in residual deviance is added to the base model if this results in an improvement in residual deviance of more than 1%. The algorithm repeats this process, updating the model, until no new terms can be added.

The appropriateness of both the raw data and the validity of the model was assessed by investigating the assumptions of: homogeneity of variance in the response variable across the covariate variable range; normality in the overall response variable distribution; lack of collinearity (correlation) amongst model

covariate parameters; linear relationship between the response variable and continuous covariate model terms; no significant interactions between the fishing year model term and other covariate model terms; independence (lack of autocorrelation) of the observational data series (Zuur et al. 2009). The level of influence that each of the fitted covariate terms had on the fishing year indices were investigated graphically using the Influ R software tools of Bentley et al. (2012).

The total seabed area of each depth-band stratum was derived using GIS software. The effective area of snapper habitat within each depth-band was derived by locating the intersections of depth band polygons and fishing hotspot polygons using GIS. Data used to construct the effective fishing area polygons were: post 2007 MPI commercial trawl and longline data; recreational harvest data obtained in an aerial-access survey analysis (Hartill et al. 2013). Using GIS, a 1 nautical mile circular buffer was drawn around the spatial location of each commercial and recreational data event; these circular areas were combined into a single contiguous area polygon.

With the final GLM model it was possible to generate the expected log-catch weight for each depth stratum and the standard error on this estimate. The account for uncertainty in the depth indices parametric boot straps were undertaken (1000) from a normal distribution defined by the predicted stratum log-catch and its standard error. Each bootstrap estimate was exponentiated to give a relative catch weight and this value was multiplied by the GIS estimate of the snapper habitat area of the stratum.

Results

East Northland

Seven depth bands were defined for east Northland on the basis of the cumulative commercial catch (Figure 11; Figure 12). The assumption is made from these data that the resident snapper biomass beyond 120 m is negligible (Figure 11; Figure 12).

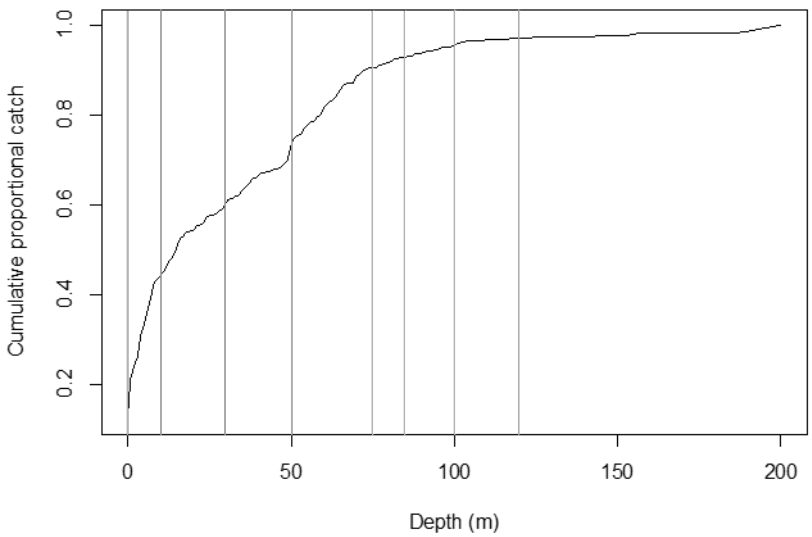


Figure 11: East Northland longline cumulative proportional catch versus depth. Vertical lines indicate the depth bands.

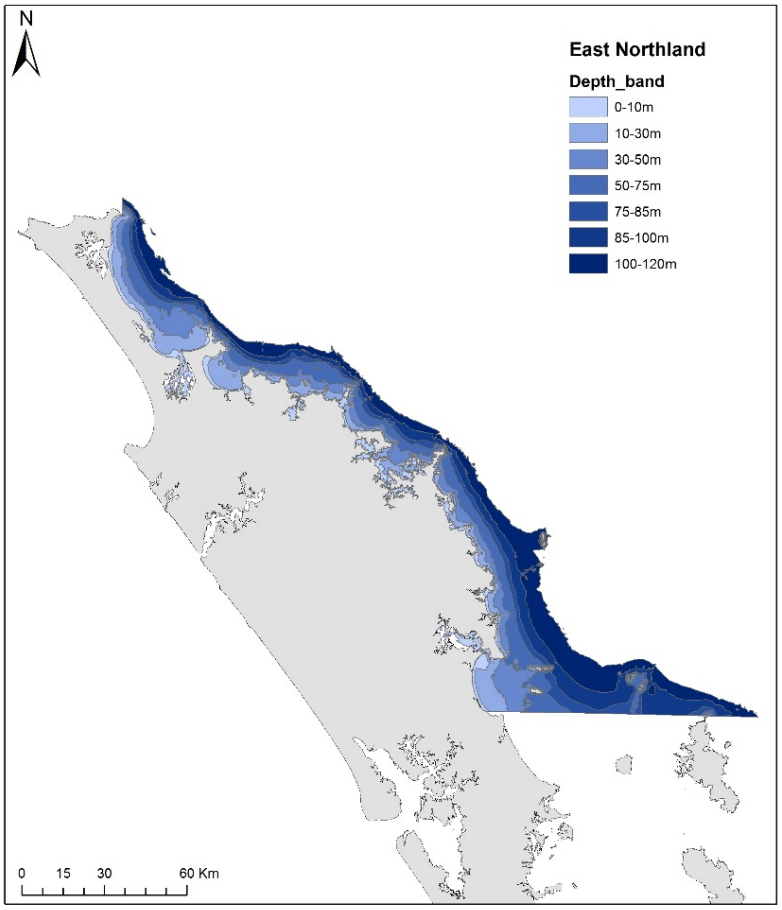


Figure 12: East Northland catch depth strata.

Effective stratum habitat areas as derived from intersecting the estimated snapper habitat area (Figure 13) is given in Table 13.

Table 13: Total and effective snapper habitat area estimates for each east Northland depth stratum.

Depth band (m)	Area (km^2)	Habitat Area (km^2)
(0,10]	537.84	478.98
(10,30]	781.63	774.48
(30,50]	787.55	780.31
(50,75]	956.64	901.7
(75,85]	469.7	412.54
(85,100]	957.96	758.86
(100,120]	1329.08	808.43

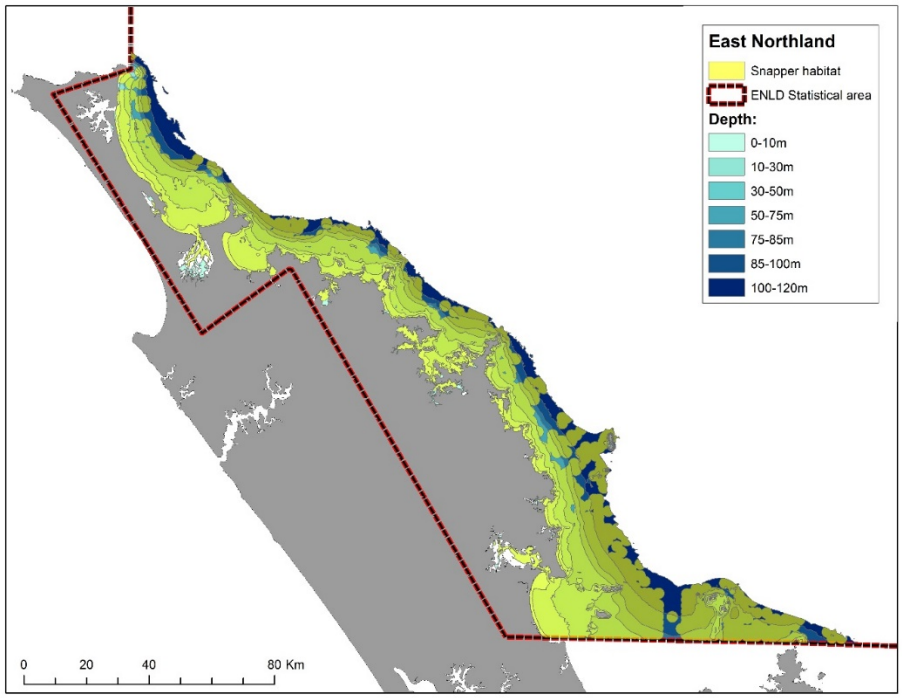


Figure 13: Effective snapper habitat area overlaid on depth strata

The bootstrap stratum abundance indices as derived from the longline catch rate indices (Appendix 8) and stratum areas (Figure 13) shows a general pattern of decreasing snapper abundance with depth (Figure 14), especially for strata deeper than 10m

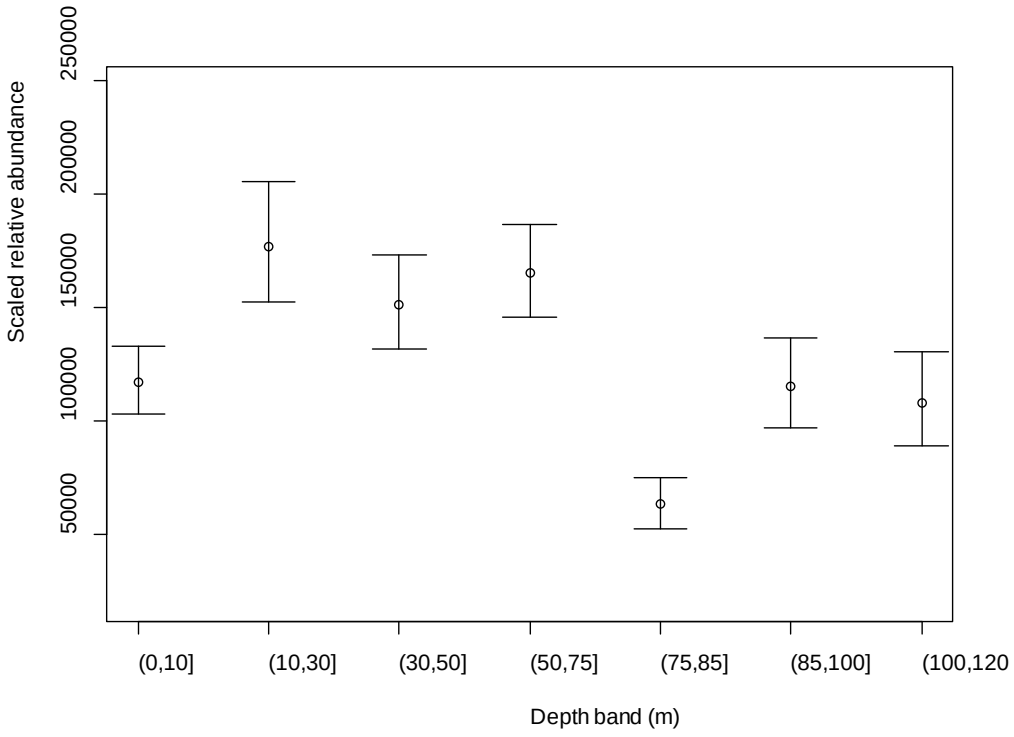


Figure 14: East Northland relative abundance estimates by depth stratum depicting bootstrap mean and associated 95% confidence intervals.

Bay of Plenty

Six depth bands were defined for Bay of Plenty on the basis of the cumulative commercial catch (Figure 15; Figure 16). The assumption is made from these data that the resident snapper biomass beyond 120 m is negligible (Figure 15; Figure 16).

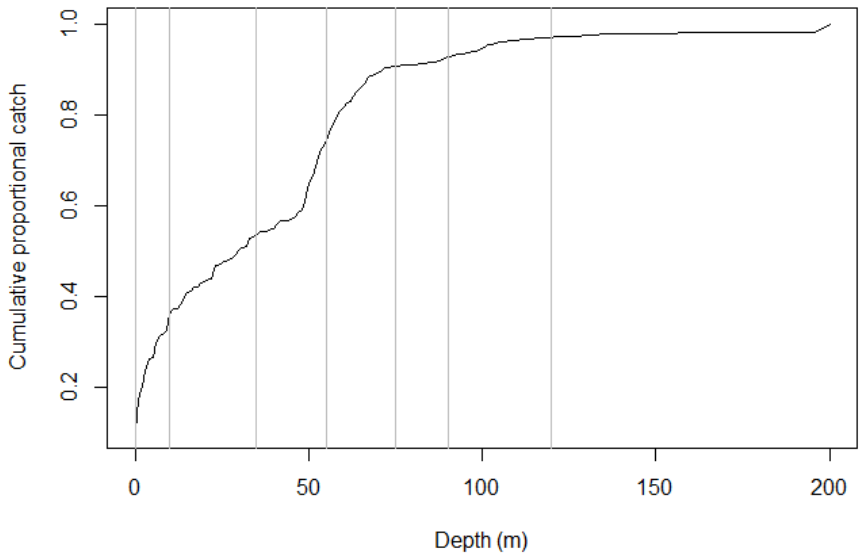


Figure 15: Bay of Plenty longline cumulative proportional catch as a function of depth; vertical lines indicate the depth bands.

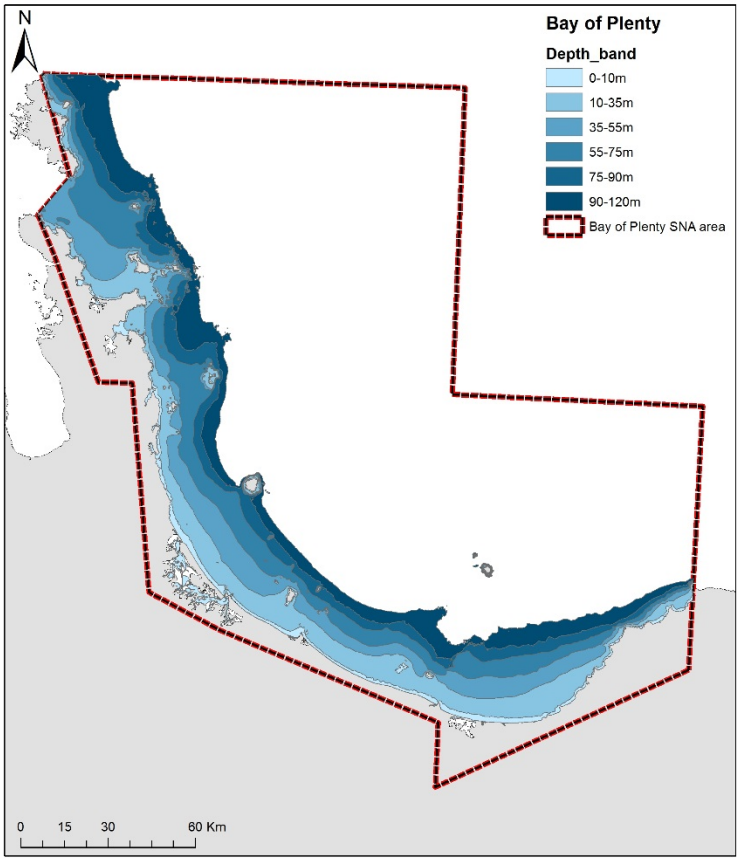


Figure 16: Bay of Plenty catch depth strata.

Effective stratum habitat areas as derived from intersecting the estimated snapper habitat area (Figure 17) is given in Table 14.

Table 14: Total and effective snapper habitat area estimates for each Bay of Plenty depth stratum.

Depth band (m)	Area (km^2)	Habitat Area (km2)
0–10	540.29	499.40
10–35	1851.04	1844.61
35–55	1690.88	1652.02
55–75	1745.86	1606.56
75–90	1125.98	1018.17
90–120	1727.91	1293.15

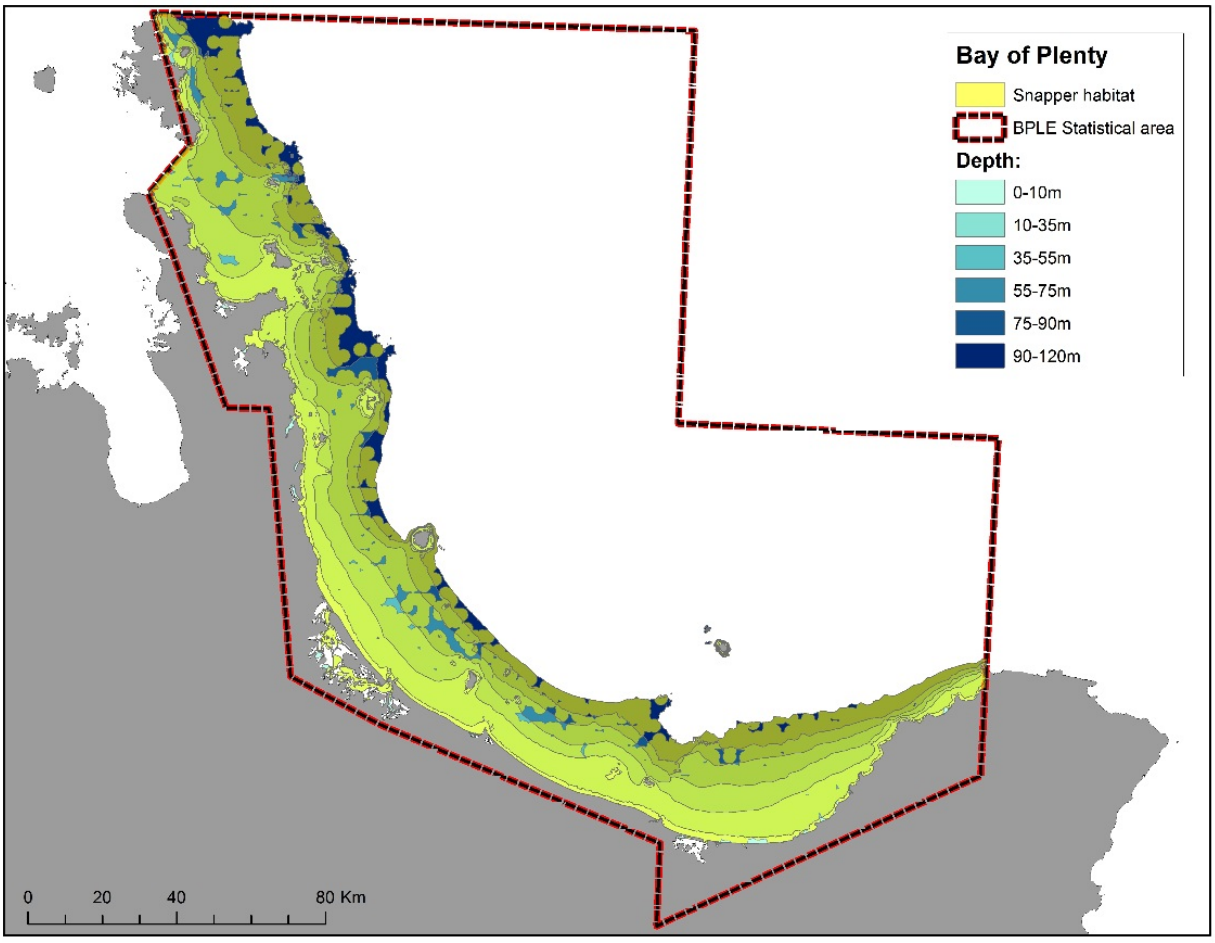


Figure 17: Effective snapper habitat area overlaid on depth strata.

The bootstrap stratum abundance indices as derived from the longline catch rate indices (Appendix 9) and stratum areas (Figure 17) shows a general pattern of decreasing snapper abundance with depth (Figure 18), especially for strata deeper than 10 m.

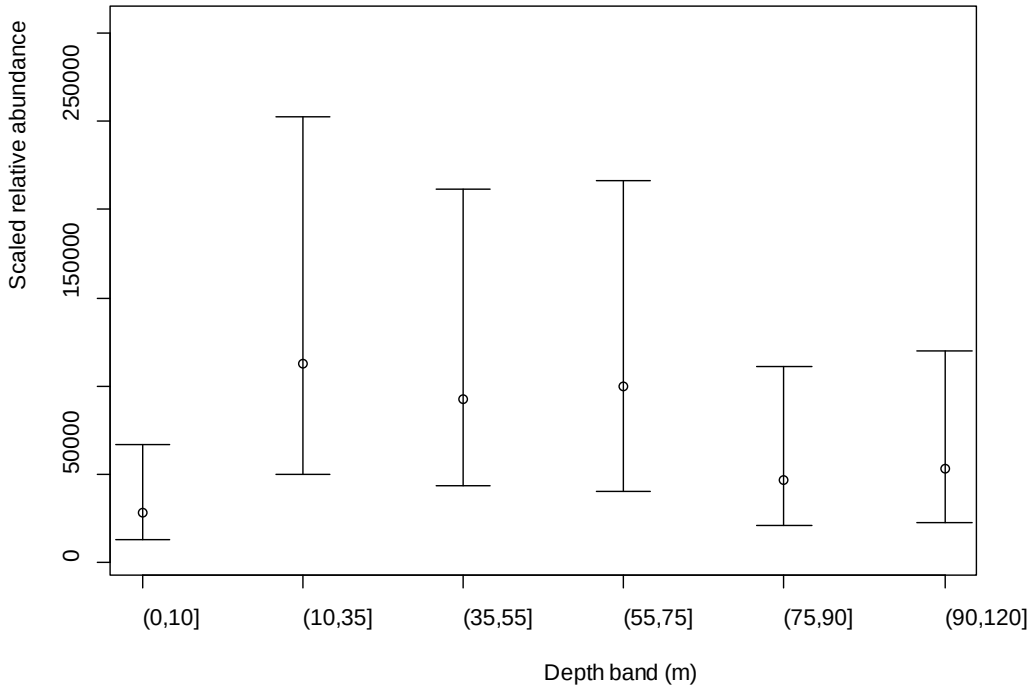


Figure 18: Bay of Plenty relative abundance estimates by depth stratum depicting bootstrap mean and associated 95% confidence intervals.

Conclusions

The analysis suggests that in the order of 20% of the snapper biomass in East Northland and the Bay of Plenty is likely to reside below 75 m during the candidate tagging months (Table 15). Not being able to tag snapper at these depths in East Northland and the Bay of Plenty has the potential to introduce bias into the tagging estimates, if there is no homogeneous mixing of shallow tagged snapper into deeper waters. However, it is feasible to derive absolute estimates of snapper abundance in areas where tagging does not take place given good independent relative abundance estimates from tagged and untagged areas.

The analyses show that it is feasible to derive estimates of relative stock size for the deep water areas of East Northland and the Bay of Plenty using spatially disaggregated catch and effort data from both the commercial and recreational fisheries. This approach could be significantly improved upon, and warrants further investigation, for example, there are likely to be better GIS methods for determining the effective snapper habitat area than the relatively simple approach used here.

Table 15: Estimated relative snapper biomass above and below 75 m during November, December and January.

		Proportion of stock		
Fishing area	Depth	Lower 95% CI	Median	Upper 95% CI
East Northland	< 75 m	0.746	0.774	0.797
East Northland	> 75 m	0.203	0.226	0.254
Bay of Plenty	< 75 m	0.616	0.773	0.878
Bay of Plenty	> 75 m	0.122	0.227	0.384

4.3 Spatial heterogeneity in mark-rates

Spatial differences in mark-rates are usually a result of incomplete mixing or failure to release tags in proportion to abundance. Release designs that tag fish in proportion to abundance are favourable because there is less reliance on fish moving and mixing to distribute tags. But an effective release design can be difficult to achieve in practice. If spatial homogeneity is not achieved, a post-hoc solution can be to subdivide the stock into a large number of spatial strata, and to derive separate population estimates for each. The difficulty in introducing more strata to account for spatial heterogeneity is that the number of observations required to adequately inform stratum movement increases exponentially. Simulations undertaken by McKenzie et al. (2011) showed that when there are too few tags to adequately estimate movement, estimates derived from less spatially complex tagging designs using the same data are generally less biased.

However, it is still important to investigate heterogeneity in the tagging data at least to gain an indication of the likely levels of bias, even if such bias may prove difficult to correct. In past snapper tagging programmes a major impediment to understanding heterogeneity has been the lack of recovery information at a fine spatial-scale (Gilbert & McKenzie 1999; Davies et al. 1999). In the 1994 SNA 1 tagging programme, scanning for tags was only undertaken at the time of landing. This commonly meant that the exact capture location was, at best, known only at the statistical reporting area level. A high proportion of tag recoveries were only assignable to the sub-area level, i.e., east Northland, Hauraki Gulf or Bay of Plenty.

The need for improved spatial detail on tag recoveries was recognised in the 2001 SNA 8 tagging programme. For this programme fishers were required to mark the snapper bins at-sea with colour-coded plastic cards such that any tags, subsequently identified upon landing, could be linked back to an area denoted by the colour of the bin-card. Bin marking compliance was adequate but not ideal with approximately 60–70% of recovered tags assignable by this method. The remaining tags were spatially assigned by other means, however, the fine-scale spatial resolution on these data was not as good.

Fine-scale spatial recording of recovered tags in future SNA 1 and SNA 8 tagging programmes is highly desirable. Options for collecting such spatial information are scanning at-sea and bin marking. Bin-marking is likely to be a more tractable option for most vessels, but the logistical requirements and associated costs for spatial referencing of tags will need to be determined in consultation with the commercial fishing operators. Depending on how this is implemented, the cost of this component in future SNA 1 and SNA 8 tagging programmes could be significant.

4.3.1 Feasible method for achieving a homogeneous release of tags

Tags from past SNA 1 and SNA 8 tagging programmes have been released in accordance to stratified random designs. Release strata boundaries were determined *a priori* on the basis of spatial abundance data from commercial catch and research trawl data (McKenzie & Davies 1996). The basic requirement of the stratified release approach is that the homogeneous distribution of tags within each stratum would be achieved through the natural mixing and movement behaviour of tagged snapper. The stratified tag release method has two components:

1. Definition of stratum boundaries in accordance with the natural mixing and movement requirements.
2. Tag stratum allocation in proportion to the relative snapper population size within each stratum.

Spatially disaggregated catch and effort data from the commercial and recreational fisheries is likely to be useful in developing stratified tag release designs for SNA 1 and SNA 8. Here, we briefly investigate the use of spatially disaggregated commercial longline catch and effort data to define tag release stratum areas for the Hauraki Gulf area of SNA 1.

Methods

Commercial longline data from the Ministry for Primary Industries database were extracted for all fishing years post 2007–08. These data were groomed following similar protocols to those given in Section 4.2.1. The data was constrained to fishing events that were conducted during November, December and January, as these are the months when tagging would most likely take place. For each fishing year we then calculated the mean CPUE within 5 km grid cells. This was used to generate a raster that was coded with a colour ramp. This allowed differentiation of areas with high versus low CPUE, and hence a visual interpretation of the temporal consistency of spatial patterns in CPUE between years was used to decide on spatial polygons (potential tagging strata) that encompassed areas with consistently similar CPUE.

In their analysis of the 1994 SNA 1 tagging data, Gilbert & McKenzie (1999) found that most individual snapper move within a home range of about 10–20 nautical miles in diameter, and that this range did not change over time. In determining release polygons, an attempt was made to balance between having too many (where it would not be logistically feasible to distribute tags with such a fine level of resolution) and too few (where variations in fish abundance would be likely to occur within these larger strata).

Results and conclusions

The spatial patterns in the commercial longline CPUE for the Hauraki Gulf were similar between 2008 and 2013 (Figure 19), and hence we combined the CPUE data from multiple years to derive the pattern in snapper spatial abundance. This was used to determine stratum boundaries (Figure 19). Indications are that a homogeneous release of tags in the Hauraki Gulf could probably be accomplished through the proportional allocation of tags across about 7–8 spatial strata.

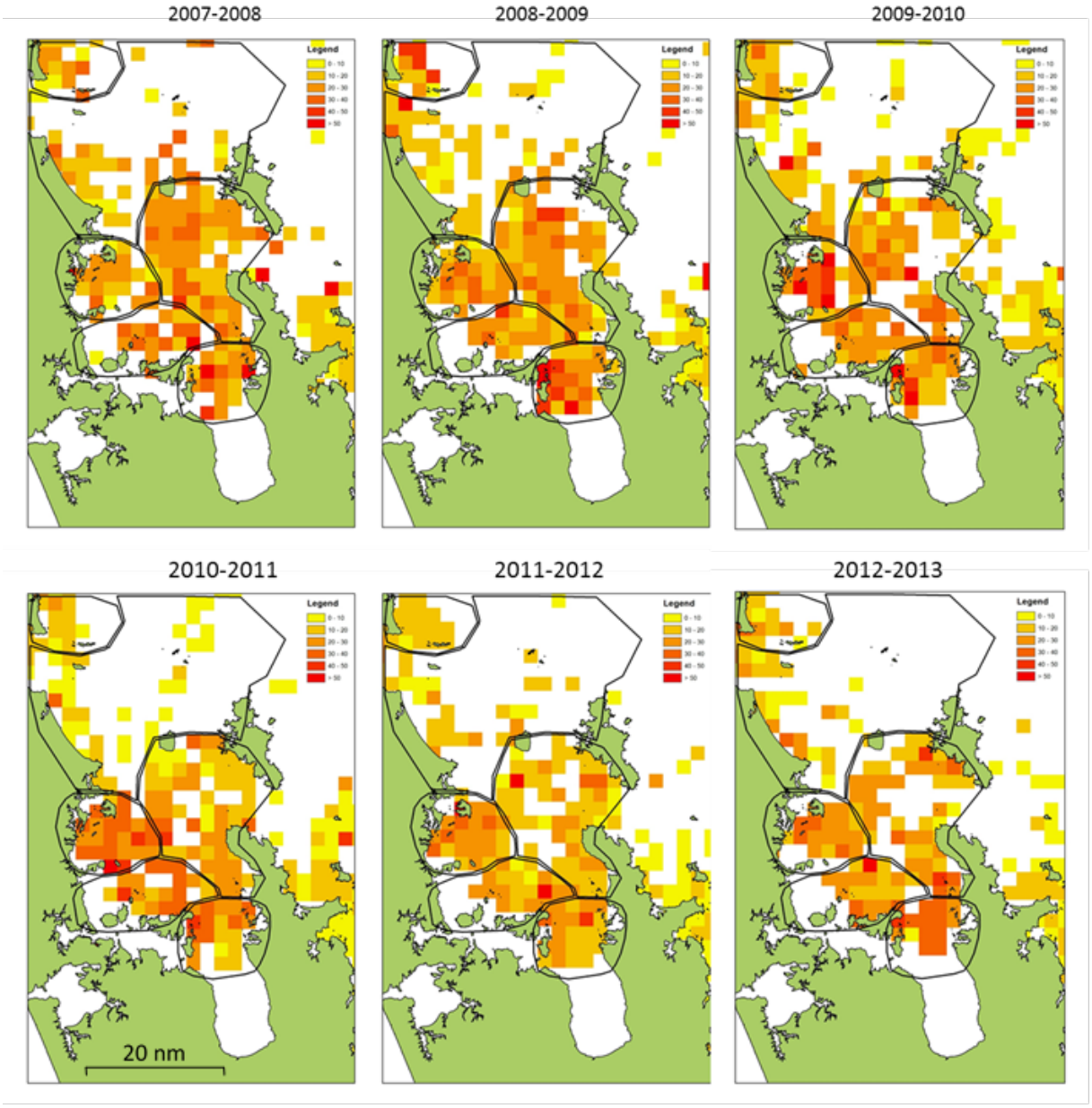


Figure 19: Commercial longline CPUE by 5 nautical mile grid-square for the November – January period from six consecutive fishing-years years. Darker colours denote higher catch rates. Polygons denote the interpolated release stratum boundaries.

There is also likely to be good spatial abundance information in the trawl and Danish seine commercial data and the recreational catch survey data. These data are likely to be sufficient for developing tag-release strata across SNA 1, and may also be suitable for determining the relative abundance of snapper within each stratum for purposes of allocating tags.

4.4 Trap-avoidance

The trap-avoidance factor ρ is defined as the proportion of tagged animals that would be expected to be recovered by the same method used to release the fish, i.e., $\rho = 1.0$ if there is no trap-avoidance. ρ can only be estimated when there are two or more release and recovery methods as a proportion difference in recapture rate between the method of interest (g1) and the reference recovery method (g_{ref}) given by the release method ratio (Equation 3).

$$\rho_{g_1} = \frac{m_{2g_1g_1} / n_{2g_1}}{m_{2g_1g_{ref}} / n_{2g_{ref}}}$$ **Equation 4**

An unbiased estimator of ρ ($\hat{\rho}$) is:

$$\hat{\rho}_{g_1} = \frac{m_{g_1g_1} / n_{2g_1}}{(m_{g_1g_{ref}} + 1) / n_{2g_{ref}}}$$ **Equation 5**

Evidence of trap-shyness in longline and trawl tagged fish was found in the 1995 east coast snapper tagging programme recovery data (Gilbert & McKenzie 1999). These results were reviewed during the recent SNA 1 assessment, but the review concluded that the original analysis could not be improved upon without additional data.

4.4.1 Methods

The potential power of the dual method release and recovery design to estimate various levels of trap-avoidance was determined using bootstrap simulation.

The purpose of the described analysis is to determine

1. the number of releases (n_1) by method and the number of fish examined for tags (n_2) that will allow an adequately powerful test for existence of the trap-shyness effect;
2. the bias and precision of the release method ratio (Equation 3) as an estimate of the trap shyness factor.

Simulations were undertaken using the tag population simulation datasets from McKenzie et al. (2011); these data are specific to two release and recovery methods: trawl and longline. The true population length frequency by area was assumed, and set in the McKenzie et al. (2011) simulations at the upper 95% estimates from the 1994 tagging programme (McKenzie & Davies 1996).

The trap-avoidance simulation steps were as follows:

1. Resample with replacement from release events from previous programmes to get a set of release numbers, $\{n_{1Li}\}$ and $\{n_{1Ti}\}$, the numbers released by line events, i = 1,2,… and trawl events, i = 1,2,…. These release numbers are made up of numbers by length class, j, $n_{1Li} = \sum_{j=a}^{h} n_{1Lij}$, etc. Continue until the target number is reached, $\sum_{i=1} n_{1Li} \geq \mu_{1L}$, etc. Aggregate by length class, $n_{1L.j} = \sum_i n_{1Lij}$, etc.

2. Adjust the number of releases by initial mortality as per the logit functions given in Gilbert & McKenzie (1999). Hence get a distribution of tagged fish in the population by release method and size class.

3. Resample with replacement from examined landings from previous programmes to get a set of numbers examined, $\{n_{2Ti}\}$, the numbers examined from trawl landings, i = 1,2,….These numbers are made up of numbers by length class, $n_{2\mathrm{T}i} = \sum_{j=a}^{h} n_{2\mathrm{T}ij}$. Continue until the target number is reached, $\sum_i n_{2\mathrm{T}i} \geq \mu_{2\mathrm{T}}$.

   The simulations were specified using the total catch weight by method and area based on 2012–13 SNA 1 landing data. Scanning targets in the simulations were specified as a percentage of the 2012 landed total, although the relative weight of catch examined from each of the two methods was held constant over the simulations. All the simulations were done using length frequencies, and because the weight of the sampled landing was known, the total weight could be used as a sample stopping criteria.

4. Aggregate across landings to give the number of tags examined, by size class, $n_{2\mathrm{T}.j} = \sum_i n_{2\mathrm{T}ij}$

5. For each value of f, simulate the recapture of trawl released tags with a binomial distribution, $m_{\mathrm{TT}.j} \sim \mathrm{B}(n_{2\mathrm{T}.j}, {}^{(fn_{1\mathrm{T}.j})}\!/_{N_j})$ for each size class, j. Similarly simulate the recapture of line released tags $m_{\mathrm{LT}.j} \sim \mathrm{B}(n_{2\mathrm{T}.j}, {}^{n_{1\mathrm{L}.j}}\!/_{N_j})$.

6. Estimate the release method ratio in each bootstrap from generated tag recoveries for the focus method (Equation 4).

7. From the bootstrap table calculate:
   a. The median of the release method ratios
   b. The %bias in the release method ratios
   c. The RMSE CV (Appendix 2) on the release method ratios

Simulations undertaken

A range of assumed values of ρ were investigated in the simulations.

Simulations assumed one of two tag recovery design options:

1. 500 SNA 1 recoveries;
2. 1500 SNA 1 recoveries.

Release and recovery targets were specified relative to a PIT cost optimised design (see Section 3.4).

Three method release ratios were specified in the simulations:

1. 20:80 Longline:trawl;
2. 50:50 Longline:trawl;

4.4.2 Results

Derived as an average across all length classes and areas, 500 and 1500 tags recovery programmes should have the power to estimate trap avoidance bias $\hat{\rho}$ within a CV = 0.05–0.10 precision range (Table 16; Appendix 7a,b). The level of bias on the simulated $\hat{\rho}$ estimates was less than 1% (Table 16).

Table 16: The median SNA 1 %bias on the trap-avoidance factor estimate ($\hat{\rho}$) derived as an average over all areas and length bins. RMSE CV on $\hat{\rho}$ given in ().

	Longline/trawl tag release ratio	
Tag recoveries	80:20	50:50
500	+0.6% (0.10)	+1.0% (0.10)
1500	+0.2% (0.06)	-0.8 (0.07)

Simulation bias and precision (RSME) estimates were largely independent of the magnitude of the assumed true value of ρ in the simulations.

4.5 Length variant mark-rates

Differences in mark-rates across length categories were evident in both the 1985 and 1994 SNA 1 tag recovery data (Sullivan et al. 1988; McKenzie & Davies 1996); these differences were likely due to the non-uniform selectivity characteristics of the two release methods (single trawl and longline). Size related mark-rate heterogeneity will almost certainly be present in future SNA 1 and SNA 8 tagging data. One method for avoiding bias in the assessment would be to stratify the tagging data by length in the assessment model, as was done in the 2013 SNA 1 assessment. The choice of stratification used in the final assessments will probably need to be determined post hoc as part of the model fitting and weighting process.

Length stratification requires knowledge of the length frequency composition of the three components of the Lincoln-Petersen equation. Whereas the length composition of the released and recovered tagged snapper (n_1, m_2) were relatively easily derived in past snapper tagging programmes, it has been logistically infeasible to measure all fish scanned for tags (n_2). Thus, past SNA 1 and SNA 8 tagging programmes required conducting additional catch sampling for length. The catch sampling proportional length frequency estimates were used to convert the sample scan weights (these being accurately known) into numbers at-length.

It is likely that future SNA 1 and SNA 8 tagging programmes that make use of internal tag markers (i.e., PIT or CWT tags) will similarly require additional catch sampling in order to derive the length composition of the scanned catch. Although this requirement is likely to constitute a significant component of the total programme cost, this cost is determined by the target mean weighted CV (mwCV) of the length sample (0.2 mwCV in past SNA 1 and SNA 8 tagging programmes) and thus largely independent of the target number of fish scanned, i.e., a similar cost under either CV = 0.1 or 0.2 precision programme designs (a fixed cost).

4.6 Growth during time at liberty

In addition to immigration, individual fish growth and recruitment of fish into the population will also result in untagged fish entering the tagged fish population during the recovery period (i.e., the period when tagged fish are at liberty following release and before recapture). If not accounted for, this can introduce bias into population estimates. The effect is further exacerbated if there are differences in the tag rates between length categories, as the mark-rate length pattern will change over the course of the recovery period due to the differential growth of small and large fish. Estimates of the growth rates for tagged and untagged fish over the recovery period will be required to correct for this bias if the tag recovery period is longer than one year (i.e., as in a triplet tagging strategy).

Growth estimates of the tagged fish can be derived from tag growth increment observations using approaches such as GROTAG (Francis 1988). Growth incremental analysis of data collected under the 2001 SNA 8 tagging programme were different from growth of the wider adult population derived from catch sampling (unpublished analysis). That study found that growth in SNA 8 PIT tagged fish was slower, consistent with little or no growth in the first year after tagging. Growth estimates of the

untagged stock will probably be best accomplished with catch at-age sampling during the recovery period. As with length sampling, the cost of age sampling is likely to be largely independent of the number of fish scanned, i.e., it is likely to have a similar cost under either a CV = 0.1 or 0.2 precision design.

4.7 Under-reporting/under-detection of recovered tags

Failure to detect tags in landed catches (scanned fish) can significantly bias the tagging population estimates, with the bias being positive and proportional to the under-reporting rate. Because of difficulties in measuring or obtaining consistent estimates of the detection rate, dedicated tag recovery programmes using CWT, PIT, or other genetic tagging technologies are preferable to voluntary reporting programmes.

For reasons given in Section 2, the PIT tag technology is likely to be the best option for any SNA 1 or SNA 8 tagging programme conducted in the near future. Significant advances in the power and efficiency of PIT tags and scanners have been made during the 15 years since the SNA 8 tagging programme was conducted. Indications are that it is now feasible to construct a PIT tag bin scanner of the type shown in Figure 3 that will be capable of very high scanning success rates which are easily quantifiable through experimentation (e.g., through tag seeding experiments).

4.7.1 Fish-bin PIT scanners

Factory-based fish-bin PIT scanners as described in Section 2.2.1, are likely to be the best tag recovery method in SNA 1 and SNA 8 for use in the short-term. The main limitation of a shed-based tag recovery programme is the approach would require at-sea bin-marking if the precise capture time and location information was to be obtained for recovered tags (see Section 4.3).

Design and construction of the six PIT bin scanners used in the 2001 SNA 8 tagging programme took approximately 6 months. Given similar resourcing and commitment, about 6–10 improved scanners could be developed within a similar time frame for similar cost. Aside from the PIT scanning technology itself, many of the bin scanner functional issues were resolved in 2001, e.g., bin conveyer systems, choice of appropriate food-safe construction materials, collection of real-time data telemetry, and programmable logic controller design. Thus the task of building a new bin scanner is likely to be less technically demanding than it was in 2001.

Although the development of 6–10 PIT bin scanners will be likely to constitute a significant cost for any future SNA 1 tagging programme, this cost would largely be incurred in the set-up year. Hence any set-up costs associated with scanning would be reduced in subsequent years under a triplet tagging programme approach. It is also likely that a similar number of scanners would be required for CV = 0.2 and 0.1 precision surveys; the difference in scanning requirement between the two designs is more likely to depend on scanner catch throughput than the number of scanners.

4.7.2 Other PIT tag recovery options

Scanning catches at-sea after each set or tow avoids the need for bin marking. However, the practicalities of bulk scanning catches at sea may mean that scanning can only take place on larger vessels, which may make it difficult to scan sufficient catch or achieve representative spatial coverage.

Given recent advances in PIT tag technology, the new PHD trawl net and the development of in-trawl biometric identification technologies it may be possible to scan substantial numbers of snapper without the need for kill and capture.

Another non-invasive approach, which does not require the commercial or recreational fishery to recover tags, is to survey areas for tags using baited underwater camera systems and underwater PIT readers.

Like the trawl scanners, baited PIT reader station are technically feasible but the system would need to be developed. Although the high degree of testing and development required is likely to preclude the widespread use of these capture methods in an upcoming SNA 1 tagging programme, such a programme provides a rare opportunity to develop and test potentially less disruptive and highly efficient new stock assessment and management tools.

5 DESIGN SPECIFICATIONS/RECOMMENDATIONS

5.1 Need for tagged-stock definition clarity

Homogeneous releases of tags in SNA 1 and SNA 8 should produce relatively unbiased estimates of the adult population numbers at-length (i.e., length stratified population estimates) within the tagging areas, these specific to the time tags were released. Numbers at-length can be converted to weight at-length and thus provide estimates of total adult biomass (tonnes).

As part of the design phases of future SNA 1 and SNA 8 tagging programmes it will be necessary to define the spatial extent to which the tagging estimates are to apply, and also to define the minimum estimable fish length. It was shown in Section 4.2 that to estimate the stock biomass residing deeper than 75 m in east Northland and the Bay of Plenty an independent relative measure of the snapper abundance deeper and shallower than 75 m will be required.

5.2 Tag release phase requirements (Table 17)

Internal PIT tags are likely to be the best option for a SNA 1 or SNA 8 tagging programme if it is to be conducted within the next five years.

Genetics has high potential as a future marking method for snapper; the main advantages being the ability to tag fish (collect genetic sample) in situ. This removes the need to account for initial mortality and can potentially enable tagging to occur over all depths (i.e., deeper than 75 m). The use of genetic tags in SNA 1 and SNA 8 would be likely to require 2–3 years of development and hence genetic tags are not an option in the short term.

We recommend that spatial allocation of tags is carried out using a stratified design based on an *a priori* understanding of spatial abundance, and that the number of tags released in each stratum be managed. Further, the spatial allocation of tags is probably best made using analyses based on spatially disaggregated catch and effort data from the commercial fishery and recent recreational survey data. Stratum areas should reflect the likely home range of snapper (10–20 nautical miles) with an attempt made also to minimise the number of release strata defined for each stock area (6–12).

To achieve the best balance between the need for spatial coverage across the stock and the need to minimise tag release related mortality two release methods should be considered: commercial longline, and, if available, the new industry PSH trawl device. Two release methods would be required in order to estimate trap-avoidance bias. The relative allocation of tags to each release method would need to be programme specific, with the ratio determined in association with the commercial fishing industry.

Prior to tagging, there will need to be a review of fish on-board tagging and handling practices, which would include an analysis of how the industry PSH technology would be used in snapper tagging programmes. These reviews should also include a review of the 75 m maximum tagging depth criteria. It is very likely that some field work and experimentation will be required to support any conclusions from these reviews.

It should be noted that a new mortality field study will be required if catching and handling procedures of any methods change significantly to those of previous snapper tagging programmes.

Table 17: Tag release phase requirements.

Item/task	Unit cost	Fixed cost	Lead/acquisition time	Specifications/requirements
PIT tag	$2–5	No	2–3 months	• "Food safe" plastic encapsulation • length 12 mm; diameter 3 mm (or smaller) • High read-range (ideally half duplex HDX)
Release phase design and implementation	Unknown	No	10–12 months	• Stratified random • Stratification based on recent commercial and recreational CPUE data • Proportional allocation of tags based on recent commercial and recreational CPUE data • November – January release period • Two release methods used: commercial longline; industry PHD trawl net • Allocation of tags across release methods made in light of consultation with commercial fishers • Electronic data acquisition of tag number fish condition, information, release station details
Tagging and handling practice review	Unknown	Yes	6–8 months	• Revised set of criteria of handling, tagging and releasing snapper • Revise 75 m maximum depth constraint
Possible PHD Mortality study	Unknown	Yes	12 months	• Can be done any time before or after tag release assuming evidence suggests PHD mortality is no worse than trawl

5.3 Tag recovery phase requirements (Table 18)

The use of PIT tags will require building 6–10 bulk bin scanners for use in commercial fish factories. Scanner design and construction will be likely to take about 6–8 months and constitute a significant cost. Scanner construction costs will probably be incurred only in the first scanning year.

PIT technologies have advanced considerably since the last PIT tagging programme in 2001 such that there will need to be a review of PIT tag technologies and scanning options.

For the purposes of testing for and estimating mark-rate heterogeneity there would be a requirement to recover tags over the entire spatial range of the stock. It is likely that a broad spatial coverage of recoveries could only be achieved by scanning across multiple methods in each of the main commercial factories, and this may also require scanning recreational catches in some areas (e.g., inner Hauraki Gulf).

Independent length/age sampling programmes for all tag recovery methods are likely to be required in each stock area for both the estimation of length composition of scanned catches and growth estimation of un-tagged fish in each stock area.

Table 18: Tag recovery phase requirements

Item/task	Unit cost	Fixed cost	Lead/acquisition Time	Specifications/requirements
PIT scanner	Unknown	Yes	6–8 months	• MPI food safety compliant • OSH compliant • Bin conveyer (12 bins per minute) • Bin counter • Data logging: vessel details; tag number; bin counts • Data telemetry and remote access • At least 95% tag detection success • Quantification of scanner success through independent trials (e.g., tag seeding).
Independent length/age catch sampling programme	Unknown	Yes	2–3 months	• Length sampling of scanned catches • Age-length sampling in each stock area to estimate untagged fish growth rates
Catch scanning programme	Unknown	No	4–5 months	• Need for catch over a wide spatial area including high recreational take areas • Scanning to occur in all major SNA factories from all major methods • Possible need to scan recreational catches from inner Hauraki Gulf and from within the trawl exclusion zone in SNA 8. • 12 month minimum tag recovery period (longer under multiple tagging event designs) • Time allowance for tag population mixing • Electronic data acquisition tag, fish, and scanned catch details. • Fine-scale recapture location data (likely to require at-sea bin marking for commercial catches)

5.4 Analytical requirements (Table 19)

Although the tagging data will be used as an input to a spatial disaggregated stock assessment model (e.g., McKenzie 2012; Francis & McKenzie in press), it will also be necessary to derive estimates of local biomass, movement, and selectivity to better understand how to apply the data (for example, to consider bias adjustments and corrections) in a stock assessment model. Also required will be an investigation of the release and recovery data for evidence of heterogeneity (spatial and temporal) and trap avoidance bias.

An analysis of the length-age sampling data will provide critical parameters for input to the main analysis. And, the results from previous and new mortality studies should be reanalysed in light of recent statistical advances in survival estimation.

Table 19: Analytical requirements

Item/task	Unit cost	Fixed Cost	Lead/acquisition time	Specifications/requirements
Lincoln-Petersen biomass estimation	Unknown	Yes	6–8 months	• Independent Biomass, movement and selectivity estimates • Adjusted (bias corrected) data for input to SA models
Heterogeneity and bias investigation	Unknown	Yes	6–8 months	Investigation of tag recovery data for: • Spatial heterogeneity • Trap avoidance
Growth analysis	Unknown	Yes	4 months	• Growth estimating of untagged fish based on independent age-length sampling • Tagged fish growth based on tag growth increment data
Incidental Mortality analysis	Unknown	Yes	6–8 months	• Analysis of new and old mortality study data using recent statistical approaches
Independent CPUE study to estimate the relative biomass residing deeper than 75 m	Unknown	Yes	6–8 months	• Analysis to determine "effective" snapper area within each depth stratum • commercial and recreational spatial catch and effort analysis

5.5 Future tagging technologies (Table 20)

Our investigations into SNA 1 and SNA 8 mark-recapture options have identified technologies that have some potential to improve the accuracy, precision, and efficiency of snapper tagging programmes in the future. Three include:

1. Genetic tagging;
2. In-net trawl scanners and biometric identification systems;
3. Baited underwater camera-PIT reader

These technologies will require some research and development before they could be used in a snapper mark-recapture programme. However, the potential efficiency and precision gains offered by these technologies (in particular genetic markers) suggests that investment in their development is likely to be worthwhile.

Table 20: Future tagging technologies.

Item/task	Unit Cost ($)	Fixed Cost	Lead/acquisition time	Specifications/requirements
Genetic tagging protocol development	500 – 1000 k	Yes	12–18 months	• Identification of suitable DNS microsatellites for snapper • Development of genetic screening protocols • Development of a suitable biopsy hook for snapper • Development of hook deployment protocols
Development of in-trawl biometric identification technologies	Unknown	Yes	12–24 months	• Develop trawl mounted PIT scanner (required detection success better than 90%) • Develop trawl mounted lowlight camera/acoustic imaging systems for counting fish as they enter the net • Develop post processing and real-time analytical software for deriving counts and logging tags and integrating vessel VMS data.
Baited underwater camera and PIT scanner	Unknown	Yes	12–24 months	• Development and testing of camera-PIT census system • Develop camera mounted PIT scanner (required detection success better than 90%) • Field testing and trialling

5.6 Tagging survey frequency and target precision

5.6.1 Frequency of tagging events

Tagging programmes have the power to estimate and describe the pattern of relative abundance change in the stock over time, if conducted at suitable intervals. However, as tagging is typically high cost relative to other monitoring options for SNA 1 and SNA 8 such as CPUE or trawl surveys, it may not be cost-effective to conduct annual or biennial tagging in SNA 1 and SNA 8 to describe patterns of biomass change.

The power of absolute biomass estimate is that it "anchors" the stock assessment biomass trajectory at a fixed point in time. Hence the previous SNA 1 tag biomass estimates (20 and 30 years old respectively) are still highly informative in the recent (2013) assessment. It is difficult to determine from the limited number of these programmes conducted to date in SNA 1 and SNA 8 (2) the frequency that such biomass estimates should be undertaken. We assume, in this report, that point-in-time biomass estimates in SNA 1 and SNA 8 every 10 years or so would be a reasonable default.

5.6.2 Target precision on the tagging biomass estimate

The scale and scope (i.e., cost) of SNA 1 and SNA 8 tagging programmes is determined by:

- the required target precision on the final biomass estimate;
- the level of analytical "power" required to estimate or account for mark-rate heterogeneity (bias).

Simulation work by McKenzie et al. (2011) has shown that the highest practical survey precision target for SNA 1 was CV = 0.1. McKenzie et al. (2012) showed that this equated to approximately 1500 tag recoveries. Survey designs that have an expected precision of CV = 0.1 represents the highest cost option, and also have the greatest power to estimate movement and to mitigate bias. But a key consideration is what the risks and implications associated with lower precision (and hence cheaper) tagging designs would be?

From McKenzie et al. (2011) the indicative number of tag recoveries required to achieve a precision of CV = 0.2 or 0.3 in the estimates of SNA 1 biomass were 500 and 250 respectively. The 1994 SNA 1 tagging programme had precision CV = 0.2 and achieved 530 tag recoveries. The 1993 SNA 1 tagging data strongly influenced the 2013 assessment, however both the level of spatial bias in this survey and estimates of stock movement were uncertain. In the 2013 assessment, understanding of the bias and corrections required would have been significantly improved with more tag recoveries.

Another consideration is in the additional power provided by two independent similar precision surveys to estimate relative change in stock biomass. Our simulations indicate that the precision on the biomass ratio estimate from two precision CV = 0.1 surveys is CV = 0.15. For two CV = 0.2 precision surveys the ratio reduces to CV = 0.3. The implication being that two CV = 0.1 precision surveys would be 98% successful at detecting a 25% change in biomass whereas the expected success rate of two CV = 0.2 precision surveys drops to 85%.

For reasons given in this report the expected variable cost component of a 1500 tag recovery programme is approximately 1.73 times the cost of a 500 tag recovery programme; as a generality it is more cost effective to conduct fewer high precision tagging studies than many lower precision ones, if the aim is to achieve the same overall precision.

5.6.3 Multiple (triplet) compared to single release event staging strategies

If it can be assumed that the relative change in stock biomass is log-linear over the time interval spanned by two or more independent tagging surveys, then it is possible to derive biomass estimates of higher precision than the individual survey CVs.

The linear-change hypothesis is likely to be a reasonable assumption for SNA 1 and SNA 8 over time intervals up to about two years. And, over a two year period, it would be possible to conduct three independent tagging surveys in either SNA 1 or SNA 8. Hence, there are at least two design options for obtaining a point-in-time biomass estimate:

- one target precision survey;
- three consecutive lower precision surveys the estimates from which can be combined to generate a target precision estimate (i.e., triplet tagging).

There are at least four significant advantages of the triplet tagging strategy:

1. Better understanding of seasonal movement patterns because tagged fish from the first and second release events are observed over two or more annual cycles.

2. More power to correct for and adjust release and recovery designs for spatial heterogeneity, i.e., modify second and third year designs based on first and second year recovery data.

3. Potential to estimate total mortality rates using Brownie type tag decay models.

4. While triplet tagging is likely to be moderately more expensive than a high precision single survey programme, the total cost would be spread over four years instead of two — potentially halving the annual funding burden on fisheries stakeholders in each programme year.

The main disadvantage of the triplet strategy is that the final high precision estimate would not be available until 3–4 years after the first tagging event as opposed to 1–2 years for the single survey approach. However, under the triplet strategy the biomass estimate from the first survey would likewise be available to managers after 1–2 years which would still be able to be used for interim management.

The simulations showed that three CV = 0.2 precision surveys (500 tag recovery), when combined, produce a precision that is the same as that achieved with one 1500 tag recovery programme. These also show that it is about 75% more expensive to conduct three independent 500 tag recovery surveys than one 1500 tag recovery survey, if we do not take into account re-use of scanners or other efficiencies that can be obtained with a multi-year programme. When allowance is made for these efficiencies, the simulations suggest that triplet tagging is only marginally more expensive (about 10%) than a single larger tagging programme.

6 ACKNOWLEDGEMENTS

This work was funded by the Ministry for Primary Industries (project SNA201303).The authors thank the following people for their significant contributions to the conclusions and analyses in the report: David Middleton, Jim Fitzgerald, Alison Undorf-Lay, Paul Starr (Sanford NZ Ltd); Marc Griffiths (Ministry for Primary Industries); Steve Parker, Ian Tuck (NIWA); Paul Pang (Unitec Institute of Technology), David Ashton, Nicholas Tuckey, Alistair Jerrett (Plant and Food Research). Marianne Vignaux for editorial review

David Gilbert is due posthumous recognition for providing much of the method description given in Section 4.4.

7 REFERENCES

Arnason, A.N.; Mills, K.H. (1987). Detection of handling mortality and its effects on Jolly-Seber estimates for mark-recapture experiments. *Canadian Journal of Fisheries and Aquatic Sciences 44 (Suppl. 1)*: 64–73.

Baker, C; Hamner, R.; Cooke, J. Heimeier, D.; Vant, M.; Steel, D.; Constantine, R. (2013). Low abundance and probable decline of the critically endangered Maui's dolphin estimated by genotype capture–recapture. *Animal Conservation 16 (2)*: 224–233.

Banks, S.C.; Hoyle, S.D.; Horsup, A.; Sunnucks, P.; Taylor, A.C. (2003). Demographic monitoring of an entire species (the northern hairy-nosed wombat, *Lasiorhinus krefftii*) by genetic analysis of non-invasively collected material. *Animal Conservation 6*: 101–107.

Bartholomew, A.; Bohnsack, J.A. (2005). A review of catch-and-release angling mortality with implications for no-take reserves. *Reviews in Fish Biology and Fisheries 15(1–2)*: 129–154.

Bellemain, E; Swenson, J.E.; Tallmon, D.; Brunberg, S.; Taberlet, P. (2005). Estimating Population Size of Elusive Animals with DNA from Hunter-Collected Feces: Four Methods for Brown Bears. *Conservation Biology 19 (1)*: 150–161.

Bentley, N., Kendrick, T. H., Starr, P. J., and Breen, P. A. (2012). Influence plots and metrics: tools for better understanding fisheries catch-per-unit-effort standardizations. – ICES Journal of Marine Science, 69: 84–88.

Bravington, M.V.; Grewe, P.M.; Davies, C.R. (2014). Fishery-independent estimate of spawning biomass of Southern Bluefin Tuna through identification of close-kin using genetic markers. FRDC Project No. 2007/034, CSIRO: 147 p.

Brownie, C.; Anderson, D.R.; Burnham, K.P.; Robson, D.S. (1985). Statistical inference from band recovery data: a handbook. U.S. Fish and Wildlife Resource Publications. U.S., Department of the Interior, Washington, D.C.

Buckworth, R.; Ovenden, J.; Broderick, D.; Macbeth, G.; McPherson, G.; Phelan, M. (2012). GENE-TAG: Genetic mark-recapture for real-time harvest rate monitoring: Pilot studies in Northern Australia Spanish Mackerel fisheries. *Northern Territory Government, Queensland, Australia: Fishery Report (107).* 116 p.

Carroll, E.; Patenaude, N.; Childerhouse, S.; Kraus, S.; Fewster, R.; Baker, C. (2011). Abundance of the New Zealand subantarctic southern right whale population estimated from photo-identification and genotype mark-recapture. *Marine biology 158(11)*: 2565–2575.

Chapman, D.G. (1951). Some properties of the hypergeometric distribution with applications to zoological censuses. *University of California Publications in Statistics 1*: 131–160.

Cooch, E.G.; White, G.C. (2013). Program Mark - a gentle introduction (edition 13). University of Colorado electronic library, Denver. 1012 p.

Davies, N.M.; McKenzie, J.R.; Gilbert, D.J. (1999). Monte Carlo estimation of bias in Petersen mark-recapture estimates for snapper (*Pagrus auratus*) New Zealand Fisheries Assessment Research Document 99/20. (Unpublished report held by NIWA library, Greta Point, Wellington). 52 p.

Davies, N.M.; McKenzie, J.R.; Gilbert, D.J. (2006). Assessment of the SNA 8 stock for the 2003–04 fishing year. *New Zealand Fisheries Assessment Report 2006/9*. 58 p.

Feathers, M.G.; Knable, A.E. (1983). Effects of depressurization upon largemouth bass. *North American Journal of Fisheries Management 3*:86–90.
Francis, R.I.C.C. (1988). Maximum likelihood estimation of growth and growth variability from tagging data. *New Zealand Journal of Marine and Freshwater Research 22*: 42–51.
Francis, R.I.C.C. (1999). The impact of correlations in standardised CPUE indices. New Zealand Fisheries Assessment Research Document 99/42. 30 p
Francis, R.I.C.C.; McKenzie, J.R. (in press). Assessment of the SNA 1 stocks in 2013. Draft New Zealand Fisheries Assessment Report 86 p.
Gaskell, T.J.; George, B.J. (1972). A Bayesian Modification of the Lincoln Index. *Journal of Applied Ecology 9(2)*: 377–384.
Gilbert, D.J.; McKenzie, J.R. (1999). Sources of bias in biomass estimates from tagging programmes in the SNA 1 snapper (*Pagrus auratus*) stock. New Zealand Fisheries Assessment Research Document 99/16. (Unpublished report held by NIWA library, Greta Point, Wellington). 47 p.
Gilbert, D.J.; McKenzie, J.R.; Davies, N.M. (2001). Evidence from tag recapture experiments that fish learn to avoid fishing gear. *Journal of agricultural, biological, and environmental statistics 6(2):* 281–291.
Gilbert, D.J.; McKenzie, J.R.; Davies, N.M.; Field, K.D. (2000). Assessment of the SNA 1 stocks for the 1999–2000 fishing year. *New Zealand Fisheries Assessment Report 2000/38.* 52 p.
Hannah, R.W.; Parker, S.J.; Matteson, E.M. (2008). Escaping the surface: the effect of capture depth on submergence success of surface-released Pacific rockfish. *North American Journal of Fisheries Management 28*:694–700.
Hare, M.P.; Nunney, L.; Schwartz, M.K.; Ruzzante, D.E.; Burford, M.; Waples, R.S.; Ruegg, K.; Palstra, F. (2011). Understanding and estimating effective population size for practical application in marine species management. *Conservation Biology 25(3)*: 438–449.
Hartill, B.; Bian, R.; Rush, N.; Armiger, H. (2013). Aerial-access recreational harvest estimates for snapper, kahawai, red gurnard, tarakihi and trevally in FMA 1 in 2011–12. *New Zealand Fisheries Assessment Report 2013/70.* 44 p.
Hoyle, S.D.; Leroy, B.M.; Nicol, S.J.; Hampton, W.J. (2015). Covariates of release mortality and tag loss in large-scale tuna tagging experiments. *Fisheries Research 163*: 106–118.
Kendrick, T.H.; Bentley, N. (2011). Fishery characterisations and catch per unit effort indices for three sub-stocks of John dory in JDO 1, 1989–90 to 2008–09. *New Zealand Fisheries Assessment Report 2011/38.*
Luikart, G.; Ryman, N.; Tallmon, D.A.; Schwartz, M.K.; Allendorf, F.W. (2010). Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conservation Genetics 11(2)*: 355–373.
McKenzie, J.R. (2012). An evaluation of age-structured spatially disaggregated stock assessment models for SNA 1. *New Zealand Fisheries Assessment Report 2012/38*. 76 p.
McKenzie, J.R.; Davies, N.M. (1996). Recruited biomass estimation of the east coast North Island snapper (*Pagrus auratus*) stock SNA 1 by Petersen mark recapture. (Unpublished Fisheries Assessment Working Group Report held by NIWA Library, Greta Point Wellington)
McKenzie, J.R.; Diggles, B.; Tubbs, L.; Poortenaar, C.; Parkinson, D.; Webster, K.; Miller, N.(2006). An evaluation of a new type of plastic-coated PIT tag for tagging snapper (*Pagrus auratus*). *New Zealand Fisheries Assessment Report 2006/8.* 40 p.
McKenzie, J.R.; Gilbert, D.; Bian, R. (2011). A simulation analysis to derive cost-optimised mark-recapture designs for estimating the biomass of east coast North Island snapper (*Pagrus auratus*) sub-stocks (SNA 1). *New Zealand Fisheries Assessment Report 2011/39.*
Mills, L.S.; Citta, J.J.; Lair, K.P.; Schwarz, M.K.; Tallmon, D.A. (2000). Estimating animal abundance using noninvasive DNA sampling: Promise and pitfalls. *Ecological Applications 10(1)*: 283–294.
Ministry for Primary Industries (2013). Fisheries Assessment Plenary, May 2013: stock assessments and yield estimates. Compiled by the Fisheries Science Group, Ministry for Primary Industries, Wellington, New Zealand. 1357 p.
Mormede, S.; Dunn, A. (2013). Quantifying vessel performance in the CCAMLR tagging programme: spatially and temporally controlled measures of tag-detection rates. *CCAMLR Science 20*: 73–80.
Muoneke, Ì.I.; Childress, W.M. (1994). Hooking mortality: a review for recreational fisheries. *Reviews in Fisheries Science 2*:123–156.

Murray, D.L.; Fuller, M.R. (2000). A critical review of the effects of marking on the biology of vertebrates. In: Research techniques in animal ecology: controversies and consequences. Columbia University Press, New York. pp 15–64.

Ovenden, J.R.; Berry, O.; Welch, D.J.; Buckworth, R.C.; Dichmont, C.M. (2015). Ocean's eleven: a critical evaluation of the role of population, evolutionary and molecular genetics in the management of wild fisheries. *Fish and Fisheries 16*: 125–159.

Parker, S.J.; McElderry, H.I.; Rankin, P.S.; Hannah, R.W. (2006). Buoyancy regulation and barotrauma in two species of nearshore rockfish. *Transactions of the American Fisheries Society 135*:1213–1223.

Parker, S.J.; Mormede, S. (2012). Drawing on international experience to improve performance of CCAMLR tagging programs. *CCAMLR paper WG-SAM-12/26.* 12 p.

Polacheck, T.; Eveson, J.P.; Laslett, G.M. (2010). Classifying tagging experiments for commercial fisheries into three fundamental types based on design, data requirements and estimable population parameters. *Fish and Fisheries 11*: 133–148.

Quinn, T.J.; Deriso, R.B. (1999). Quantitative fish dynamics. Oxford University Press, New York.

Raby, G.D.; Packer, J.R.; Danylchuk, A.J.; Cooke, S.J. (2014). The understudied and underappreciated role of predation in the mortality of fish released from fishing gears. *Fish and Fisheries 15(3)*: 489–505.

Rashid, N.A.M. (2014). Factors affecting tagging mortality in snapper (*Pagrus auratus*). Unpublished MSc Thesis held University of Auckland Library.

Robson, D.S.; Regier, H.A. (1964). Sample size in Petersen mark-recapture experiments. *Transactions of the American Fisheries Society 93*:215–226.

Rudershausen, P.J., Buckel, J.A.; Hightower, J.E. (2014). Estimating reef fish discard mortality using surface and bottom tagging; effects of hook injury and barotrauma. *Canadian Journal of Fisheries and Aquatic Sciences 71:*514–520.

St John, J.; Keay, I.; Wright, I. (2009) Effects of onboard handling techniques and methods of release on recapture rates of temperate demersal species in Western Australia. In: Maximising survival of released undersize west coast reef fish. FRDC Report No. 191 Western Australia Department of Fisheries.

Seber (1982). The estimation of animal abundance and related parameters. Charles Griffen & Company, London.

Steele, C.A.; Anderson, E.C.; Ackerman, M.W.; Hess, M.A.; Campbell, N.R.; Narum, S.R.; Campbell, M.R.; Grant, J. (2013). A validation of parentage-based tagging using hatchery steelhead in the Snake River basin. *Canadian Journal of Fisheries and Aquatic Sciences 70(7)*: 1046–1054.

Sullivan K.J.; Hore A.J.; Wilkinson, V.H. (1988). Snapper: In: Baird, G.G.; McKoy, J.L. Papers from the workshop to review fish stock assessments for the 1987–88 New Zealand Fishing Year. 300 p. (Preliminary discussion paper, held in NIWA library, Wellington).

Sumpton, W.D.; Brown, I.W.; Mayer, D.G.; McLennan, M.F.; Mapleston, A.; Butcher, A.R.; Welsh, D.J.; Kirkwood, J.M.; Sawynok, B.; Begg, G.A. (2010). Assessing the effects of line capture and barotrauma relief procedures on post-release survival of key tropical reef fish species in Australia using recreational tagging clubs. *Fisheries Management and Ecology (17*): 77–88.

Thorsteinsson, V. (2002). Tagging methods for stock assessment and research in fisheries. Report of Concerted Action FAIR CT.96.1394 (CATAG). *Marine Research Institute Technical Report No. 79*, Marine Research Institute, Reykjavik.

Tomich, S.D.; Clarke, M.E.; Harms, J.H.; Hempelmann, J.A. (2010). DNA sampling hook. US Patent US7823321 B2.

Vignaux, M. (1994) Catch per unit effort (CPUE) analysis of west coast South Island and Cook Strait spawning hoki 1987-93 New Zealand Fisheries Assessment Research Document 94/11. 29p

Waits, L.P.; Luikart, G.; Taberlet, P. (2001). Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Molecular Ecology (10):* 249–56.

Walsh, C.; McKenzie, J.; Bian, R.; Armiger, H.; Rush, N.; Smith, M.; Spong, K.; Buckthought, D. (2014). Age composition of commercial snapper landings in SNA 1, 2012–13. New Zealand Fisheries Assessment Report 2014/55 62 p.

Welsford, D.C.; Ziegler, P.E. (2013). Measures to avoid bias in abundance estimates of Dissostichus spp. based on tag-recapture data. CCAMLR Science (20): 63–72.

Wilde, G.R. (2009). Does venting promote survival of released fish? *Fisheries 34*:20–28.
Wydoski, R.; Emery, L. (1983) Tagging and Marking. In: Nielsen, L.A.; Johnson, D.L. (eds), Fisheries Techniques, American Fisheries Society, pp 215–237.
Zuur, A.F.; Ieno, E.H.; Elphick, C.S. (2009). A protocol for data exploration to avoid common statistical problems *Methods in Ecology and Evolution 1*: 3–14

8 APPENDICES

Appendix 1: Basic (Lincoln-Petersen) tagging estimator for closed populations.

The simplest tagging experiment involves a single tag release event followed by a second sampling event to recover tagged animals. The second event provides an estimate of the initial population tag-ratio. From this a population estimate can be derived using the Lincoln-Petersen estimator:

$$\hat{N}_{t1} = \frac{n_{1t1} n_{2t2}}{m_{2t2}}$$

Where: $\hat{N}_{t1}$ = estimated number of animals in population at time *t1*

n_{1t1} = number of marked animals in population *N* at time *t1*
n_{2t2} = number of animals examined for marks in population *N* at time *t2*
m_{2t2} = number of marked animals recovered from population *N* at time *t2*

The Lincoln-Petersen estimator has two implicit assumptions: that the population is closed over the time interval *t1* to *t2* and that tagged animals are homogenously mixed in the population at time *t2*. The closure assumption only needs to hold for additive processes i.e., immigration and recruitment. In general the initial tag ratio ($\frac{n_{1t2}}{N_{t2}}$) is unchanged by removal processes (emigration, mortality) occurring over the interval *t2-t1;* if these processes act equally on tagged and untagged animals then.

$$\frac{n_{1t1}}{N_{t1}} = \frac{n_{1t2}}{N_{t2}} = \frac{m_{2t2}}{n_{2t2}}$$

It is important to realise that the population mark-rate represents $\left(\frac{n_{1t2}}{N_{t2}}\right)$ the expectation or probability of observing a tag in the recovery sample $\left(n_{2t2}\right)$ at *t2*. Because tag recoveries are only observed as integers, the mark rate estimate $\left(\frac{m_{2t2}}{n_{2t2}}\right)$ and therefore $\hat{N}_{t1}$ will be biased at low sample sizes. Bias corrections and tagging estimator likelihood functions are described by a number of authors (Chapman 1951; Robson & Regier 1964; Gaskell & George 1972); see Seber (1982) for an overview. In general, recovery sample (m_{2t2}) sizes of less than 4 at the stratified design level should be avoided, recovery sample sizes greater than 10 have a better than 95% probability of producing unbiased estimates (Seber 1982).

Appendix 2: Root mean square error (RMSE) and bias calculation.

Scenarios were compared on the basis of root mean square error (RMSE) on the population biomass and other parameters of interest given by:

$$RMSE_{F_m(x)} = \sqrt{\frac{1}{N}\sum_{i=0}^{N}[\hat{F}_m(x_i) - F_m(x_i)]^2}$$

Where:

N	=	number of bootstraps
m	=	scenario
$F_m(x)$	=	a function of parameter *x*
$\hat{F}_m(x_i)$	=	estimator value for the function $F_m(x)$ from the *i*th bootstrap
$F_m(x_i)$	=	operating model (true) value for the function $F_m(x)$ from the *i*th bootstrap

The coefficient of variation (CV) is $RMSE_m$ as a proportion of the bootstrap mean:

$$cv = \frac{RMSE_m}{E[F_m(x)]}$$

The RMSE of the summation of, for example, the individual stratum biomass estimates B_n is given by:

$$RMSE_{\sum B_n} = \sqrt{\sum_{i=1}^{n} RMSE_{B_i}{}^2}$$

$$cv = \frac{RMSE_{\sum B_n}}{\sum B_n}$$

Bias was expressed as a percentage of the expected value of the parameter or statistic of interest to the (operating model) value:

$$\%Bias_{\hat{F}_m(x)} = 100 \text{ x } \frac{\hat{F}_m(x)}{F_m(x)}$$

Appendix 3: Table of simulated release and recovery numbers from McKenzie et al. (2011) used by the design cost optimisation tool.

Stock area	Release tags (n_1)	Scanned fish (n_2)	Recovered tags (m_2)
ENLD	4 278	493 827	67
ENLD	4 284	905 894	124
ENLD	4 278	1 318 524	180
ENLD	8 243	493 383	129
ENLD	8 238	905 382	237
ENLD	8 238	1 318 844	345
ENLD	16 154	493 858	254
ENLD	16 152	906 971	466
ENLD	16 153	1 319 108	678
HAGU	7 338	778 601	117
HAGU	7 344	1 403 398	211
HAGU	7 349	2 031 006	306
HAGU	14 126	778 261	224
HAGU	14 126	1 404 922	406
HAGU	14 132	2 032 516	589
HAGU	27 685	777 208	438
HAGU	27 675	1 404 838	796
HAGU	27 690	2 031 441	1 153
BPLE	4 787	569 913	83
BPLE	4 785	1 010 792	148
BPLE	4 790	1 452 263	213
BPLE	9 054	568 245	157
BPLE	9 052	1 010 760	281
BPLE	9 039	1 453 854	403
BPLE	17 573	568 618	307
BPLE	17 575	1 010 505	544
BPLE	17 586	1 453 174	784
SNA1	16 403	1 842 341	266
SNA1	16 413	3 320 084	483
SNA1	16 417	4 801 793	699
SNA1	31 423	1 839 889	509
SNA1	31 416	3 321 063	924
SNA1	31 408	4 805 214	1 337
SNA1	61 412	1 839 683	998
SNA1	61 402	3 322 313	1 806
SNA1	61 429	4 803 723	2 616

Appendix 4: Design and cost specification input section of the SNA 1 tag design optimisation tool (R-code).

```
# Enter target number of tag recoveries by area
tarn<- c("ENLD", "HAGU","BPLE")
#tag_tar <-c(400, 600, 400) #0.10
#tag_tar <-c(200, 300, 200) #0.15
#tag_tar <-c(150, 200,     150) #0.20
#tag_tar <-c(100, 100, 100) #0.30

###############
## COSTS  ##
###############

# VARIABLE COSTS
# cost for scanning each fish in first survey year
scost = 0.02 #PIT
#scost = 0.28 #genetic

#cost for tagging each fish in first survey year
tcost = 1.0 #PIT
#tcost = 1.0 #genetic

#proportional reduction in scan cost in subsequent years in multi-year programme
scost_offset = 0.8

#proportional reduction in tag cost in subsequent years in multi-year programme
tcost_offset = 0.9

# FIXED cost {note: these are prorated across areas in proportion the tag and
scanning totals}

# Scan fixed cost in first year
fixscan_init = 100

# Tag fixed cost in first year
fixtag_init = 0

# Scan fixed cost in subsequent years in multi-year programme
fixscan_sub = 50

# Tag fixed cost in subsequent years in multi-year programme
fixtag_sub = 0


#annual tagloss rate
tagloss = 0.3

# tagging sequence interval years
tint <- 2

# Number of surveys to run
numsurv = 3



# regularly spaced programme events (TRUE/FALSE)

#event_order = TRUE
event_order = FALSE
```

The cost-benefit simulation tool fits a regression model to the Appendix 4 input table to derive regression relationships between: numbers tagged; numbers scanned; numbers recovered such that given any two of these three values the simulator can predict the third. Note: as the tag recovery target is always fixed in the cost optimisation runs the regression predictions are numbers tagged given scanned or numbers scanned given tagged.

Appendix 5: The effect of changing tag and catch examination numbers on the number of tags recovered

The Lincoln-Petersen equation (Appendix 1) re-expressed for m_2 is

$$m_2 = \frac{n_1 n_2}{N}$$

The total number of fish tagged and examined for marks in a mark-rapture programme is:

$$n_{1+} n_2$$

Increasing the number of tag recoveries by a factor of *x* gives:

$$xm_2 = \frac{x(n_1 n_2)}{N}$$

$$xm_2 = \frac{\sqrt{x} n_1 \sqrt{x} n_2}{N}$$

The required increase in the total number of fish tagged and examined for marks necessary to increase the predicted number of recoveries by *x* is $\sqrt{x}$ i.e.:

$$\sqrt{x}(n_{1+} n_2)$$

Appendix 6: Minimum total number of tagged and scanned fish (cost) required to achieve a given number of recoveries

The Lincoln-Petersen equation (Appendix 1) re-expressed for m_2 is

$$\frac{1}{N} n_1 n_2 = m_2$$

Expressing this in terms of variables *x* and *y* gives:

$$\boldsymbol{cxy = z}$$

Where *c* and *z* are constants

Let *r* be the cost of releasing a tag
s be the cost of scanning a fish

Then z (the expected number of tags recovered) can be expressed as a cost function

$$crxsy = z$$

We need to find solutions for *rx* and *sy* where the total cost (i.e., *rx* + *sy*) is a minimum.

Minimum x
Expressing y in terms of x gives:

$$y = \frac{z}{crsx}$$

Such that:

$$rx + \frac{sz}{crsx} = \min(rx + sy)$$

$$\Rightarrow\ rx + \frac{z}{crx} = \min(rx + sy)$$

Minimum x where:

$$\frac{\partial}{\partial x}\left(rx + \frac{z}{crx}\right) = 0 \ \Rightarrow\ r - \frac{z}{crx^2} = 0$$

$$\min(rx) = \sqrt{\frac{z}{c}}$$

Minimum y
Rearranging for y, differentiating, and solving as above gives:

$$\min(\text{sy}) = \sqrt{\frac{z}{c}}$$

Thus the minimum cost tagging design will always be achieved when *rx* = *sy* i.e., the cost of tagging equals the cost of scanning.

Appendix 7a: The median %bias and RMSE CV on $\hat{\rho}$ as derived through bootstrap simulation relative to 500 and 1500 tag recovery designs and 80:20 trawl tag release ratios.

Longline/trawl tag release ratio 80:20							
500 tag recovery				1500 tag recovery			
Area	Length bin	%bias	RMSE CV	Area	Length bin	%bias	RMSE CV
ENLD	25	36.20%	1.04	ENLD	25	13.70%	0.75
ENLD	27	3.50%	0.36	ENLD	27	0.60%	0.24
ENLD	30	0.40%	0.55	ENLD	30	-6.10%	0.32
ENLD	32	-3.30%	0.50	ENLD	32	-2.30%	0.23
ENLD	35	3.60%	0.81	ENLD	35	-1.80%	0.33
ENLD	40	-3.90%	0.71	ENLD	40	-2.40%	0.29
HAGU	25	15.30%	0.96	HAGU	25	0.70%	0.39
HAGU	27	-3.10%	0.28	HAGU	27	0.50%	0.16
HAGU	30	-4.10%	0.36	HAGU	30	-0.10%	0.20
HAGU	32	-3.70%	0.33	HAGU	32	1.50%	0.24
HAGU	35	-4.80%	0.40	HAGU	35	2.00%	0.23
HAGU	40	1.50%	0.43	HAGU	40	-1.00%	0.21
BPLE	25	34.10%	0.82	BPLE	25	5.60%	0.56
BPLE	27	-6.20%	0.37	BPLE	27	-0.90%	0.20
BPLE	30	-3.10%	0.46	BPLE	30	4.60%	0.32
BPLE	32	2.30%	0.50	BPLE	32	-3.40%	0.32
BPLE	35	-4.00%	0.59	BPLE	35	1.60%	0.33
BPLE	40	3.00%	0.54	BPLE	40	-1.90%	0.39
ENLD	all	1.40%	0.25	ENLD	all	-1.20%	0.14
HAGU	all	-1.60%	0.17	HAGU	all	0.70%	0.08
BPLE	all	-0.40%	0.21	BPLE	all	0.30%	0.12
SNA1	all	-0.60%	0.11	SNA1	all	0.20%	0.06

Appendix 7b: The median %bias and RMSE CV on $\hat{\rho}$ as derived through bootstrap simulation relative to 500 and 1500 tag recovery designs and 50:50 trawl tag release ratios.

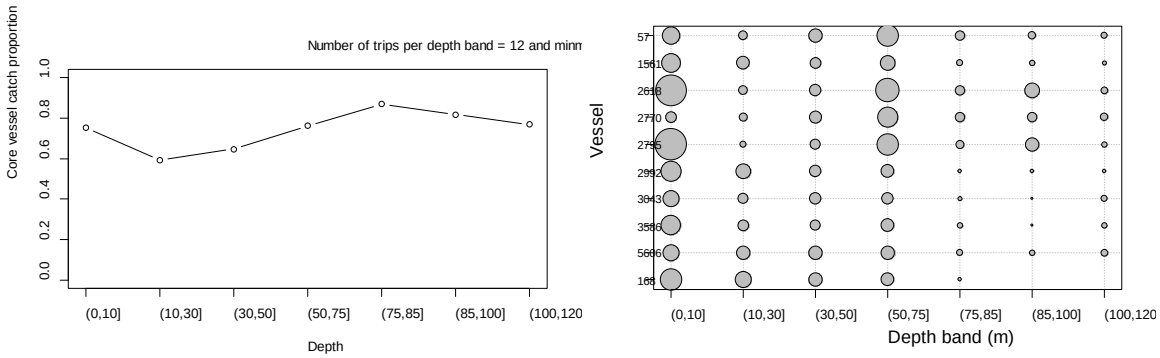
Longline/trawl tag release ratio 50:50

500 tag recovery				1500 tag recovery			
Area	Length bin	%bias	RMSE CV	Area	Length bin	%bias	RMSE CV
ENLD	25	31.60%	1.16	ENLD	25	-3.80%	0.50
ENLD	27	0.60%	0.37	ENLD	27	0.60%	0.21
ENLD	30	0.60%	0.65	ENLD	30	-6.00%	0.26
ENLD	32	15.00%	0.79	ENLD	32	3.40%	0.39
ENLD	35	0.30%	0.46	ENLD	35	0.60%	0.31
ENLD	40	4.60%	0.65	ENLD	40	-6.90%	0.32
HAGU	25	8.30%	0.52	HAGU	25	-2.80%	0.38
HAGU	27	-4.80%	0.23	HAGU	27	2.20%	0.17
HAGU	30	2.20%	0.34	HAGU	30	-0.90%	0.18
HAGU	32	-3.40%	0.32	HAGU	32	-2.30%	0.24
HAGU	35	3.60%	0.42	HAGU	35	1.40%	0.24
HAGU	40	5.00%	0.47	HAGU	40	1.70%	0.26
BPLE	25	39.50%	0.90	BPLE	25	-2.40%	0.55
BPLE	27	-0.20%	0.39	BPLE	27	-5.20%	0.21
BPLE	30	-5.40%	0.46	BPLE	30	0.20%	0.31
BPLE	32	-1.30%	0.49	BPLE	32	-0.40%	0.32
BPLE	35	3.30%	0.54	BPLE	35	1.70%	0.34
BPLE	40	-2.90%	0.48	BPLE	40	0.80%	0.35
ENLD	all	5.00%	0.23	ENLD	all	-1.40%	0.13
HAGU	all	-0.50%	0.13	HAGU	all	0.40%	0.09
BPLE	all	1.20%	0.21	BPLE	all	-1.90%	0.13
SNA1	all	1.00%	0.10	SNA1	all	-0.80%	0.07

Appendix 8: East Northland CPUE depth stratum indices

Core vessel selection

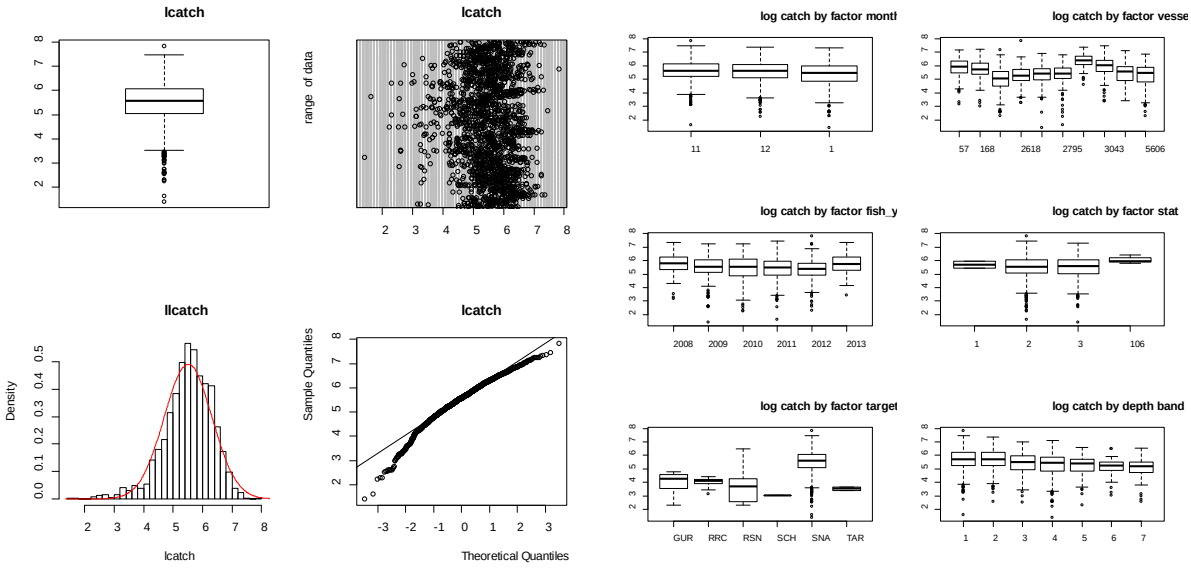
Core vessels were selected by setting the minimum number of trips in each depth band to twelve and minimum depth bands to four. This resulted in the data from 10 out of 50 vessels being used in the analysis. The core vessels account for more than 60% of the total catch for most depth bands, except depth band (10, 30], which is also very close to 60% (Appendix 9 Figure 1). Although the number of trips from the core vessels are lower in deeper than shallow waters, most vessels have fished at least twelve trips in both shallow and deep waters (Appendix 9 Figure 1).



Appendix 9 Figure 1: The proportion of core vessels' catch in that of all vessels on the seven depth bands (left panel) and each of the core vessels' catch on the seven depth bands (right panel) in East Northland in fishing years 2007–08 to 2012–13.

Data exploration

The raw log-transformed catch (per set) data did not show any major departures from that of a normal distribution (Appendix 9 Figure 2). The spread of catch data, across the different levels of the major categorical explanatory variables, appeared to be reasonably even (Appendix 9 Figure 2), therefore conforming to the assumption of homogeneous variances, except for target species and statistical area (Appendix 9 Figure 2). Statistical areas 001 (2 records) and 106 (4 records) were removed from the data because all records in these two statistical areas had no depth bands defined. Target species tends to be correlated with depth band (Appendix 9 Figure 2), e.g., fishing target at SCH only happen in depth band 6 and target TAR only happen in depth band 5 (75–85 m). Therefore target was excluded from the model covariates.



Appendix 9 Figure 2: The distribution of number of events for target species in depth bands.

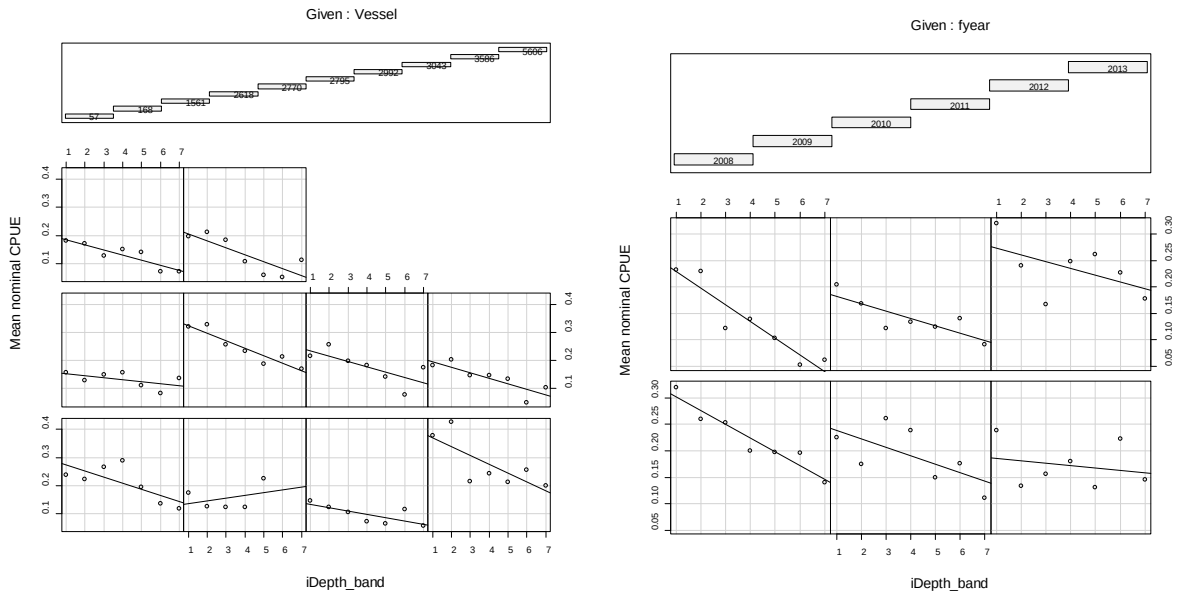
Model selection and CPUE standardisation

The full stepwise regression analysis of the East Northland longline catches (i.e., a model that did not force depth band and allowed interactions) explained about 48% of variation in the data using six model terms (Appendix 9 Table 1). Depth band was selected as a significant term in this unforced model, explaining 5.3% of the variation in the data. The addition of a vessel and fishing year interaction term explained an additional 5.5% of variation (Appendix 9 Table 1). Although the two interaction terms are selected in the model, the variation explained by the terms is relatively small. The differences between the regression lines of nominal CPUE vs depth band for vessel and fishing year are subtle (Appendix 9 Table 1), therefore, it is reasonable to not allow the interaction in the modelling.

Appendix 9 Table 1: R-squared for model in which depth band is not forced and interactions allowed

	1	2	3	4	5	6
poly(lhooks, 3)	0.012					
vessel	0.011	0.341				
iDepth_band	0.004	0.259	0.394			
iDepth_band:vessel				0.432		
fish_year	0.006	0.253	0.368	0.423	0.458	
iDepth_band:fish_year						0.475

The relationship between depth band and mean nominal CPUE for vessels and fishing years shows subtle differences, which can be seen as an indication of an interaction between depth band and vessel and fishing year (Appendix 9 Figure 3).



Appendix 9 Figure 3: Mean nominal CPUE vs depth band for vessels (left panel) and fishing years (right panel).

When interactions are excluded and depth band is forced in, the final model obtained from the stepwise regression process explained about 43.5% of variation in the data using five model terms (Appendix 9 Table 2).

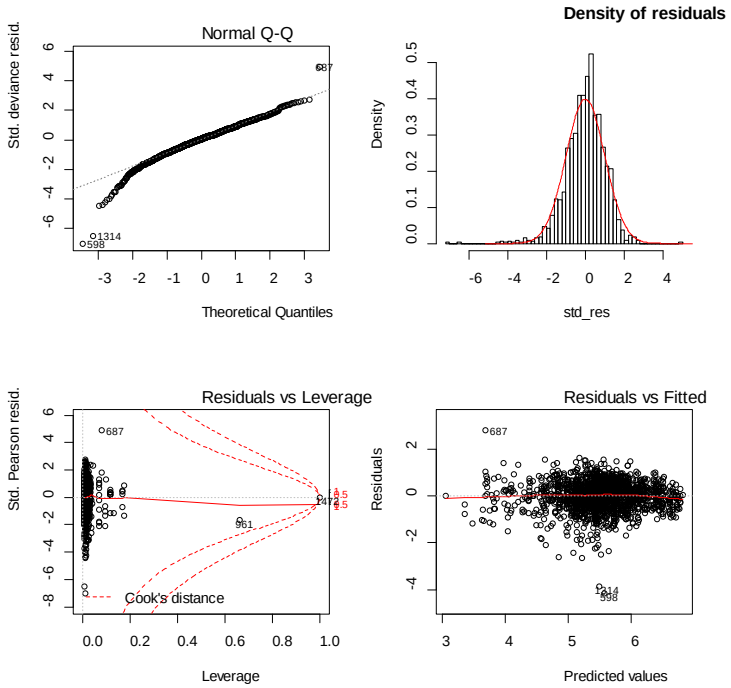
```
Final Model:

    lcatch ~ iDepth_band + vessel + poly(lhooks, 3) + fish_year + month
```

Appendix 9 Table 2: R-squared for model without interactions.

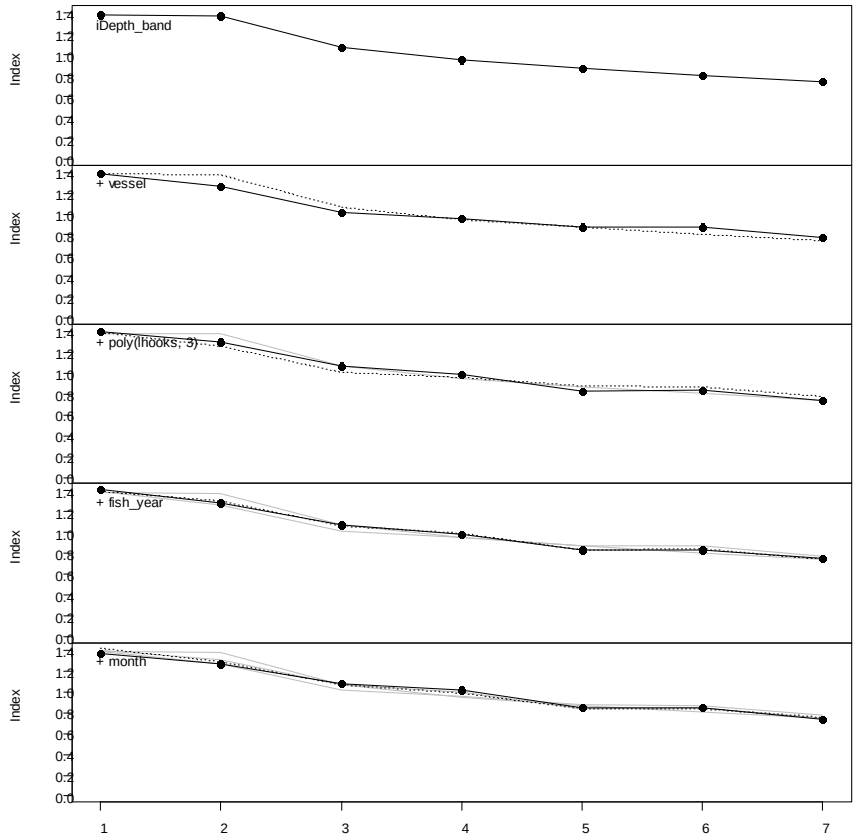
	1	2	3	4
vessel	0.263			
poly(lhooks, 3)	0.259	0.394		
fish_year	0.108	0.291	0.423	
month	0.071	0.270	0.406	0.435

The residuals in the final model were normally distributed across ±2 standard deviations (95%) of the range of the data (Appendix9 Figure 4). The plot of Cook's distance scores shows that no observations had undue influence on the model fit (Appendix9 Figure 4).



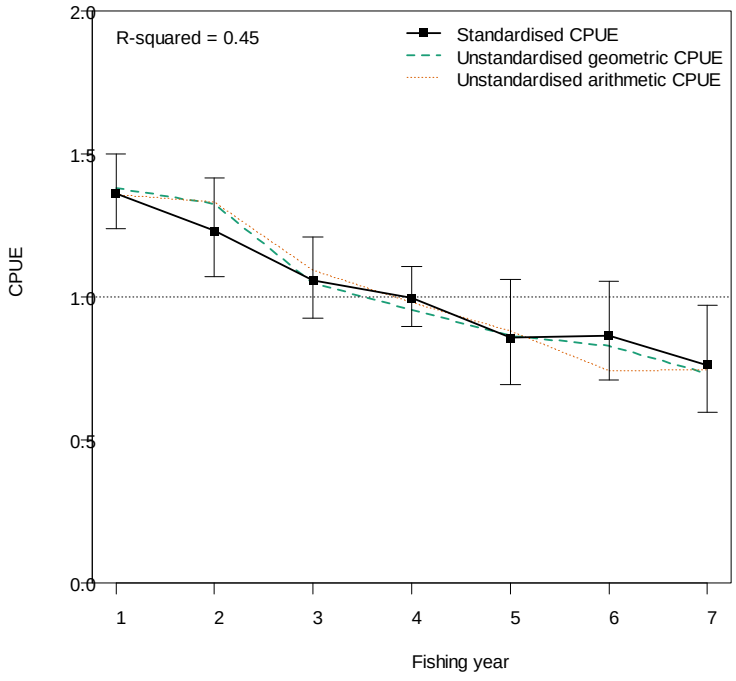
Appendix9 Figure 4: Diagnostics for the final east Northland model.

The variables most influential on the CPUE index were the number of hooks set and vessel; the addition of the other model terms had little impact on the pattern in the index (Appendix 9 Figure 5).



Appendix 9 Figure 5: Influences of model covariates sequentially added in the final model.

The final East Northland index shows a clear trend of decreasing with increasing depth (Appendix Figure 6).

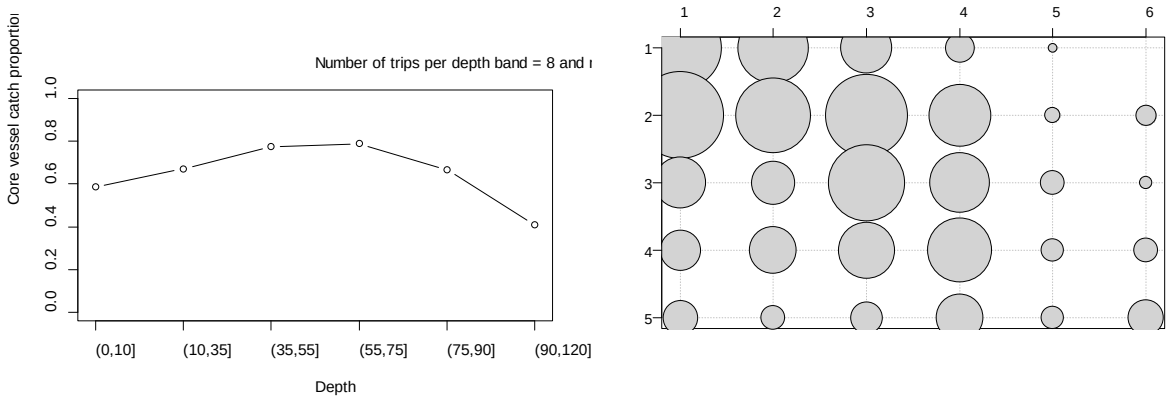


Appendix Figure 6: East Northland Standardised and unstandardised depth band CPUE indices.

Appendix 9: Bay of Plenty CPUE depth stratum indices

Core vessel selection

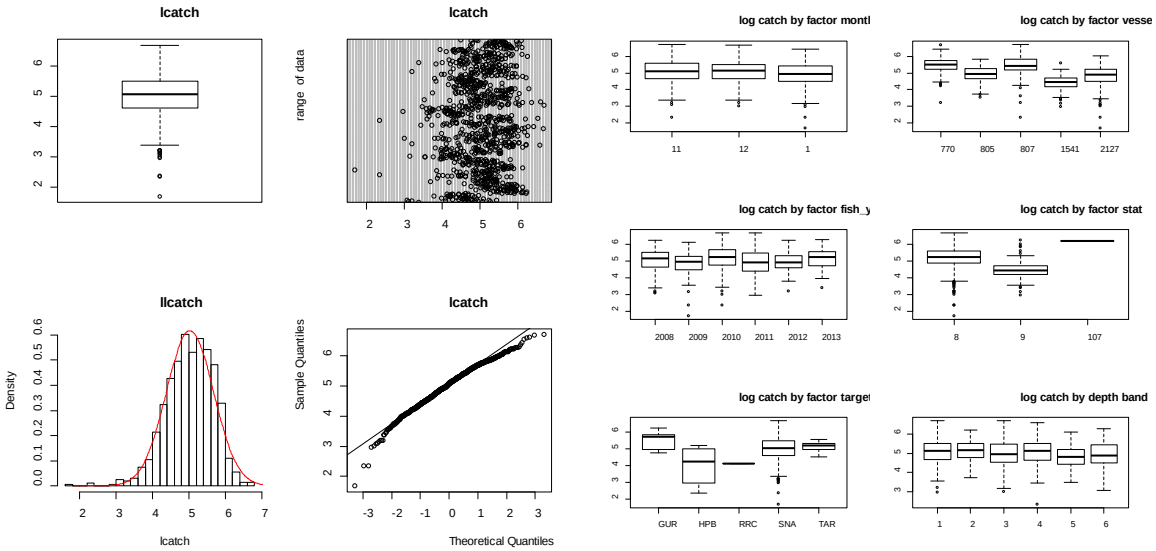
Core vessels were selected subject to a minimum number of trips in each depth band of 8 and a minimum number of depth bands of 4. This resulted in the data from 5 out of 36 vessels being used in the analysis. The core vessels account for more than 60% of the total catch for most depth bands, except depth band (10, 30], which is also very close to 60% (Appendix 10 Figure 1). Although the number of trips from the core vessels are lower in deeper than shallow waters, most vessels have fished at least 12 trips in both shallow and deep waters (Appendix 10 Figure 1).



Appendix 10 Figure 1: The proportion of core vessels' catch in that of all vessels on the six depth bands (left panel) and each of the core vessels' catch on the seven depth bands (right panel) in Bay of Plenty in fishing years 2007–08 to 2012–13.

Data exploration

The raw log-transformed catch (per set) data did not show any major departures from that of a normal distribution (Appendix 10 Figure 2). The spread of catch data, across the different levels of the major categorical explanatory variables, appeared to be reasonably even (Appendix 10 Figure 2), therefore conforming to the assumption of homogeneous variances, except for target species (Appendix 10 Figure 2). Target species tends to be correlated with depth band (Appendix 10 Figure 2), e.g., fishing target at RRC only happen in depth band 2.



Appendix 10 Figure 2: The distribution of number of events for target species in depth bands.

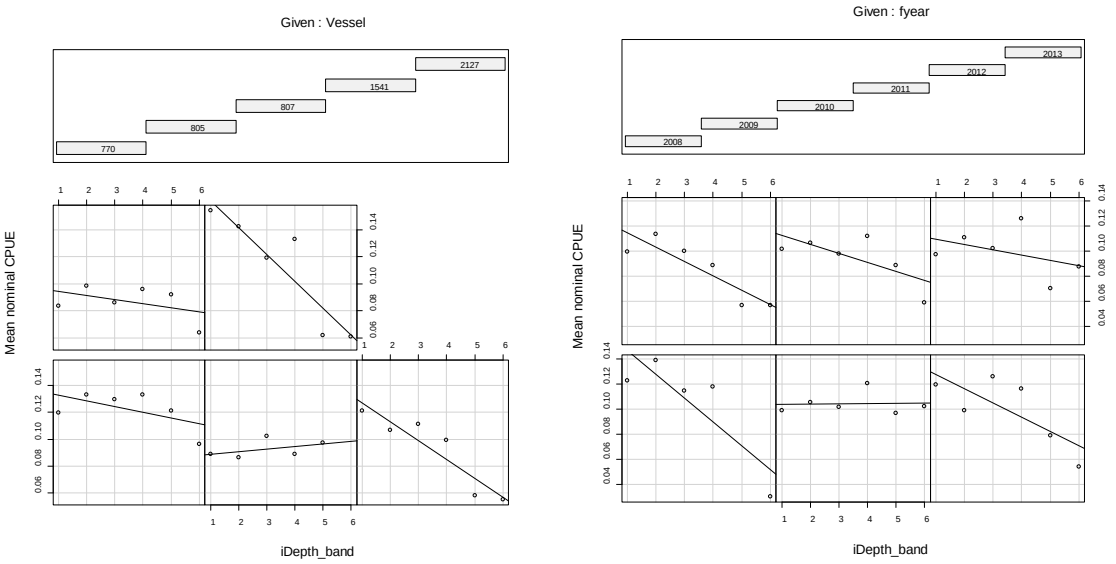
Model selection

The full stepwise regression analysis of the Bay of Plenty longline catches (i.e., a model that did not force depth band and allowed interactions) explained about 64% of variation in the data using seven model terms (Appendix 10 Table 1). Depth band was selected as a significant term in this unforced model, explaining 1.3% of the variation in the data. The addition of a vessel and fishing year interaction term explained an additional 3.7% of variation (Appendix 10 Table 1). Although the two interaction terms are selected in the model, the variation explained by the terms is relatively small. The differences between the regression lines of nominal CPUE vs depth band for vessel and fishing year are subtle (Appendix 10 Table 1), therefore, it is reasonable to not allow the interaction in the modelling.

Appendix 10 Table 1: R-squared for model in which depth band is not forced and interactions allowed.

	1	2	3	4	5	6	7
poly(lhooks, 3)	0.480						
vessel	0.428	0.534					
target	0.021	0.520	0.571				
fish_year	0.022	0.497	0.555	0.591			
iDepth_band	0.017	0.505	0.561	0.586	0.604		
iDepth_band:fish_year						0.629	
iDepth_band:vessel						0.624	0.641

The relationship between depth band and mean nominal CPUE for vessels and fishing years shows differences, which can be seen as an indication of an interaction between depth band and vessel and fishing year (Appendix 10 Figure 3).



Appendix 10 Figure 3: Mean nominal CPUE vs depth band for vessels (left panel) and fishing years (right panel).

When interactions are excluded and depth band is forced in, the final model obtained from the stepwise regression process explained about 60.4% of variation in the data using five model terms (Appendix 10 Table 2).

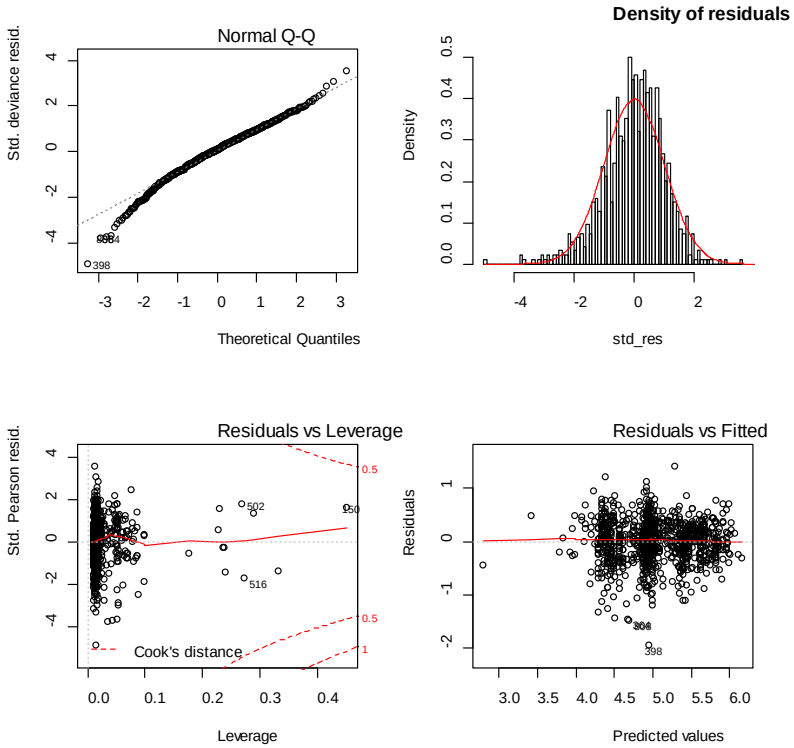
```
Final Model:

lcatch ~ iDepth_band + poly(lhooks, 3) + vessel + target + fish_year
```

Appendix 10 Table 2: R-squared for model excluding interaction and forcing depth band in.

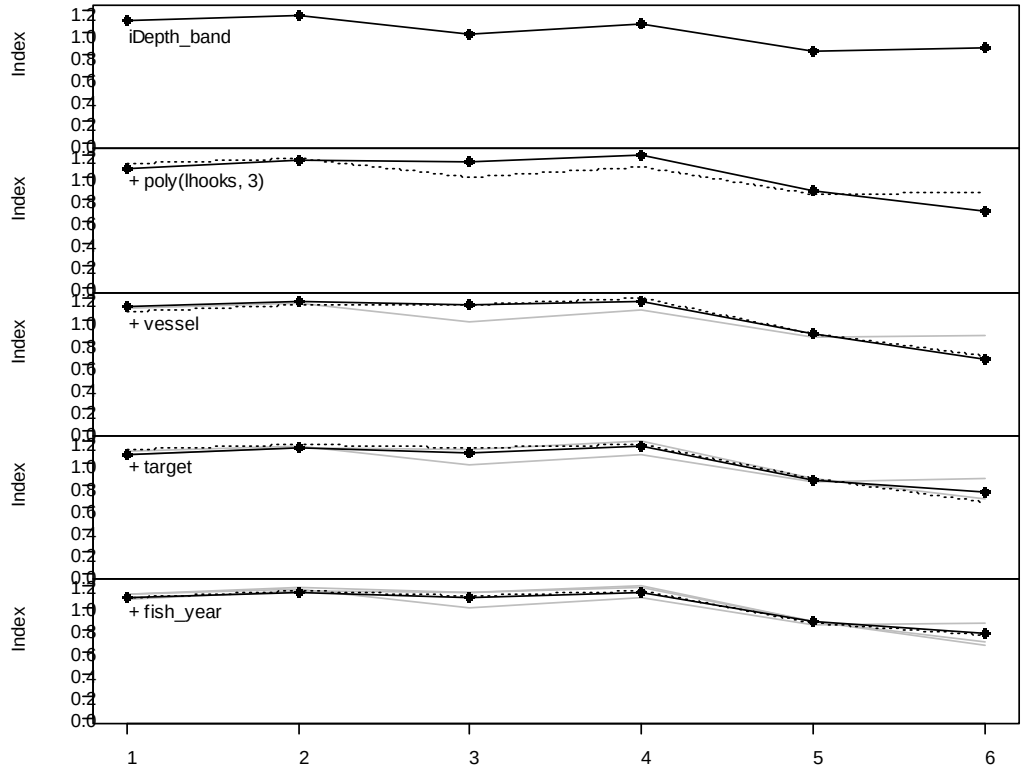
	1	2	3	4
poly(lhooks, 3)	0.505			
vessel	0.448	0.561		
target	0.042	0.535	0.586	
fish_year	0.035	0.519	0.578	0.604

The residuals in the final model were normally distributed across ± 2 standard deviations (95%) of the range of the data (Appendix 10 Figure 4). The plot of Cook's distance scores shows that no observations had undue influence on the model fit (Appendix 10 Figure 4).



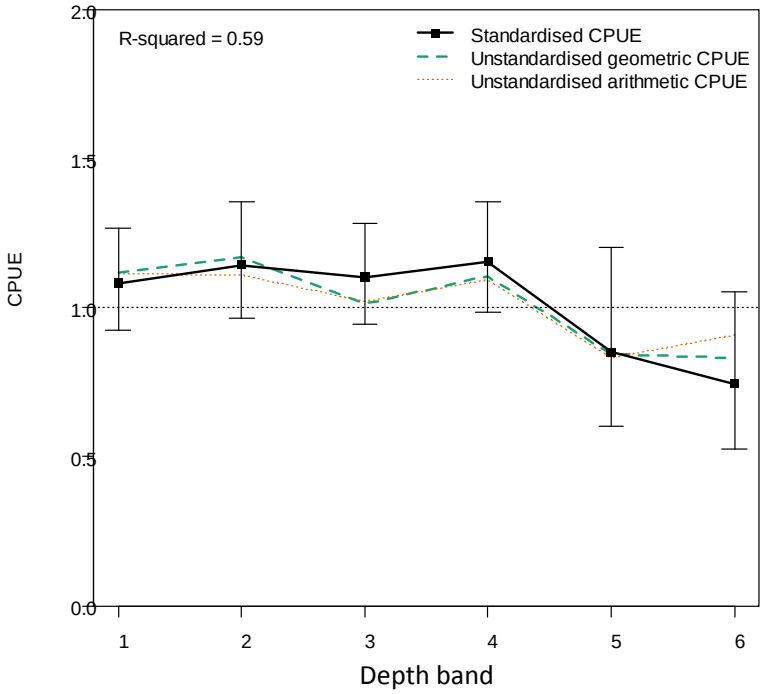
Appendix 10 Figure 4: Diagnostics for the final east Northland model.

The variables most influential on the CPUE index were the number of hooks set and vessel; the addition of the other model terms had little impact on the pattern in the index (Appendix 10 Figure 5).



Appendix 10 Figure 5: Influences of model covariates sequentially added in the final model.

The final Bay of Plenty index shows a clear trend of decreasing with increasing depth (Appendix 10 Figure 6).



Appendix 10 Figure 6: East Northland Standardised and unstandardised depth band CPUE indices.