NIWA
Taihoro Nukurangi

Review of post-disposal monitoring of the Auckland Marine disposal ground

Prepared in confidence for Maritime New Zealand

July 2011

Authors/Contributors:

Alison MacDiarmid, Helen Neil, John Mitchel, Joanne O'Callaghan and Judi Hewitt

For any information regarding this report please contact:

Neville Ching
Contracts Manager

+64-4-386 0300
n.ching@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd
301 Evans Bay Parade, Greta Point
Wellington 6021
Private Bag 14901, Kilbirnie
Wellington 6241
New Zealand

Phone +64-4-386 0300
Fax +64-4-386 0574

NIWA Client Report No:	WLG2011-31
Report date:	July 2011
NIWA Project:	MNZ11303

© All rights reserved. This publication may not be reproduced or copied in any form without the permission of the copyright owner(s). Such permission is only to be given in accordance with the terms of the client's contract with NIWA. This copyright extends to all forms of copying and any storage of material in any kind of information retrieval system.

Whilst NIWA has used all reasonable endeavours to ensure that the information contained in this document is accurate, NIWA does not give any express or implied warranty as to the completeness of the information contained herein, or that it will be suitable for any purpose(s) other than those specifically contemplated during the Project or agreed by NIWA and the Client.

Contents

Reviewed and Approved for release by

……………………………………

Dr Phil Heath

Executive summary

A review has been undertaken for Maritime New Zealand of a report written as part of the monitoring of effects of a permitted dredge spoil dumping operation at a new dumping site 25 km east of Great Barrier Island. The report actually comprised six individual reports on different aspects of the monitoring programme and a summary report that also made recommendations for future monitoring.

The review examines the content of each report and determined whether the initial objectives (as specified in the Summary Report) had been met by the work undertaken. This entailed reviewing the data collected during the monitoring program and their accuracy and interpretation. Although MNZ specified the review was not to include the methodology used in the collection of the monitoring data, in places it proved impossible to separate the veracity of the findings from the methodology used to collect it.

Lastly the proposals for future monitoring have been reviewed and feedback provided as to its likely effectiveness, including comments on the proposed timeframes for monitoring in relation to dumping volumes, and whether the specific study elements proposed will sufficiently cover potential environmental effects of marine dredge disposal.

Parts of the monitoring programme were found to be deficient to some extent.

Monitoring of the plume during disposal of the dredge spoil (Reports 3 and 6) was adequately done and while we have identified areas where this could be improved we have confidence in the conclusion that it was highly likely that the dispersed material was localised to within 600 m or so of the disposal site.

A detailed error budget for the multibeam survey was not presented. Without this the stated accuracy of the bathymetric survey is impossible to defend.

Cores were taken to determine if the sea floor sediments had altered in response to the deposition of near shore clays and muds. Unfortunately the methodology employed to characterise the sediment was inadequate to draw firm conclusions. Similarly, the way in which core material was analysed for sediment characteristics and contaminants potentially diluted any effect of the recently deposited dredge material.

There were problems with interpreting images of the sea floor in Report 6 (Plume imagery and underwater video). There was a 3-4 month delay between disposal of dredge material and the seafloor image survey, during which time oxidation of previous anoxic black muds took place making it impossible to distinguish between disposed and native sediments. This delay, the poor quality of the imaging, and the lack of scale in the images greatly diminished the usefulness of the seafloor imaging in determining the spatial spread of the dredge material from the disposal location. Little confidence can be placed in the conclusions drawn from these images.

We found serious problems with the biological assessment (Report 5). No statistical analyses of any differences observed between pre and post disposal were given. The discussion relates only to diversity (undefined although probably average number of taxa per core) and abundance. No multivariate community analyses or analyses of dominant taxa, or taxa that could be expected to be sensitive, are discussed. It was not apparent from the design of the monitoring programme whether the intention was for the data to be analysed by ANOVA-

based BACI (Before-After-Control-Impact techniques) or by the more sensitive gradient techniques. The design is not optimal for either but as a result could be analysed by either. The accuracy of the findings, such as they are, are based on a very few, small cores, that are unlikely to sample the area in sufficient detail as to determine any changes related to disposal. The larger scale sampling conducted using the underwater video footage only recorded organisms, rather than details of organism activities.

In conclusion we would suggest that whilst the monitoring programme may be adequate to identify the gross environmental changes resulting from the spoil dumping it is in many areas inadequate to identify the more subtle changes that define the boundaries of the effect. In commenting on the recommendations for future monitoring made in the reports we have taken in to account the weakness identified above and suggest ways in which future monitoring may be substantially improved.

.

1 Background

Maritime New Zealand (MNZ) has contracted NIWA to review a report entitled “Post-Disposal Monitoring of the Auckland Marine Disposal Ground” as completed for Coastal Resources Ltd by the Coastal Marine Group of the University of Waikato in February 2011. The report was written as part of the monitoring of effects of a permitted dredge spoil dumping operation at a new dumping site 25 km east of Great Barrier Island. Maritime New Zealand requires NIWA to undertake the following:

- Review of the report content and confirmation that the initial objectives (as specified in the Executive Summary) have been met by the work undertaken. The review is not to include the methodology used in the collection of the monitoring data.
- Review the data collected during the monitoring program, and provide comments on the results of the monitoring and the accuracy of the findings interpreted from the data. Provide comments on each individual component of the monitoring program and on the program as a whole.
- Review the proposals for future monitoring and provide feedback as to its effectiveness, including comments on the proposed timeframes (dumping volumes), and whether the specific study elements proposed will sufficiently cover potential environmental effects of marine dredge disposal.

In the sections below we comment on each individual component of the monitoring programme and on the programme as a whole, indicate whether the initial objectives of the monitoring programme have been met, and provide feedback on the likely effectiveness of the proposed future monitoring and whether it is likely to cover the potential environmental effects of marine dredge spoil disposal.

2 Review of reports

2.1 Report 1 - Benthic Cores: Initial observations

A weighted gravity corer was used to obtain a sore of sediment from the sampling sites before and after disposal of dredge material. No information is presented regarding the potentially stiffer layer that may have stopped gravity core penetration. Without other supporting evidence, and no documented use of a core catcher or similar, a 12 cm core length may simply be the volume of sediment able to be retained by suction on extraction of the corer.

The discipline of Marine Sedimentology has over many years used visual observations of sediment characteristics to make qualitative comparisons. However, for these observations to be used in any comparative way, either spatial or temporally, they must be made with appropriate colour charts and comparators (e.g. grain size/clast size, % occurrence). This report uses no such tools and therefore suffers strongly from perception bias. This can lead to errors in judgement between analysts, between samples, between monitoring and control sites, and between prior or future assessments.

Sediment colour analysis is inexact because no two observers have the same colour perception, however this bias can be mitigated by reference to standards. Sediment colour is usually determined visually by comparison to a colour chart, for example the Geological Society of America Rock Colour Chart (Goddard et al., 1948), or other derivatives of the Munsell Colour Chart. Similarly, other field comparators are widely available for use, and again would have provided a level of rigour when assessing such factors as the nature (size and colour) and quantity of ‘clumps’ of apparent dredged disposal material and ‘black’ disposal material.

Sediment colour has long been used as a qualitative measure of redox conditions within marine sediments and the report does use sediment colour to make some statements about this, but references to support these statements are required. However, without rigour in the colour visual assessments, and with no other information about the nature or constituents of the ‘fluffy’ core top material, the usefulness of this information in making comparisons is limited.

The occurrence of ‘fluffy’ material at the core top does indicate that much of the sediment surface has been retained. However, this study used a gravity corer and the bow wave of such corers, as well as the necessary retrieval mechanism, can still vertically disturb the sediment surface during coring. Thus, evidence, or lack thereof, of bed forms at the sediment surface should be viewed with caution. An assessment of the constituents of this flocculated marine sediment (e.g. flocculated clays, organic-matter rich, marine snow) would aid in both the assessment of redox conditions and activity levels of the associated benthic fauna.

Photographs of cores supplied in the appendices, although many are lacking appropriate scale, appear to indicate there is some variability in the depth and nature of boundaries between the layers described in the cores. Variability would be expected in the marine realm and while generalised depth statements are made in the text, the variability should be documented. The boundary between the surface flocculated material and underlying sediment appears in photographs to vary from down core depths of only 2 cm, up to 5cm, and the boundary in some cores is sharp and linear while in others irregular or even indistinct. Without an appropriate scale, statements such as ‘thin golden brown layer” provide no context or understanding of what depth of sediment layers ‘thin’ is ascribed to.

If sufficient photographic records (and image meta-data) were made at the time of collection of the sediment cores, and/or the sediment cores were retained, then more visual observations could be made. Scale should also be applied to all images and depth of the sediment boundary indicated. Also, an assessment of the percentage of ‘clumps’ seen at site H2 would be useful for making comparison with further disposal and monitoring. Comparative visual observations could be easily incorporated in the report as either a table or a simple core log next to each photograph in the appendices.

Until such comparative visual observation data are available or presented, only a record of occurrence can be made in the report. Statements such as “significant amounts of black anoxic mud” (in relation to sites at H2 post-disposal), “black mud was in high concentration” and “black mud concentration decreases” cannot be supported with the evidence presented in this report.

The decision to sample 250m east of H2, as a result of no visual evidence of disposal material 500m distant is a worthwhile one. However, it is again hampered by perception unless some rigour is applied to statements such as “quantity was significantly less”.

New biological information is introduced in the Conclusion. This information would be better placed in the main body of the report or in the biological report.

The conclusions need to be limited to general statements about occurrence or not of features. The lack of supporting data results in the present concluding statements such as ‘small amounts of’ and ‘majority’ being both difficult to defend and of no use in further comparative studies as they appear to be based on an individual’s perception. Hence, the only conclusion that be supported by the data reported is that disposal material was observed at H2 and at 250m east of H2.

2.2 Report 2 - Multibeam echosounder: Seafloor geomorphology

Comparative surveys are the main tool used to determine the bathymetric impact of spoil dumping on the seafloor. As multibeam data collection can be strongly affected by the survey environment, a consistent and carefully thought out work programme and project methodology is required if the data collected are to be directly comparable, survey to survey.

Overall the multibeam surveys conducted to date are a good measure of location, distribution and impact of the dumping of the dredge material, although as the volume of dumped dredged material increases the impact should be more noticeable, and as a result survey technique and repeatability will become more important.

The stated accuracy of the multibeam data in the report of 5 cm RMS for shallow water and increasing with depth is a simplistic view of error budgets. Most error sources for the multibeam bathymetry are predictable and easily accounted for with the exception of speed of sound in water. These predictable errors are normally accounted for by a patch test of the system prior to the start of surveying. The assumption (not stated) is that this was done at the mobilisation stage of the surveys.

Unmeasured changes in sound speed through the water column are unpredictable and can potentially result in significant depth and positioning errors. This error source is mitigated by obtaining a velocity profile prior to the start of data collection and continually monitoring the bathymetry data for evidence of sound speed artefacts during sounding operations. The surface sound speed is continuously measured and used to calculate departure angles at the transducer and are also used as an indicator of sound speed changes throughout the water column. Again this is not stated in the report but the assumption is that this was done.

A typical error budget for the EM3000D as used by NIWA, compared to the LINZ standards, is given in the table below. This emphasises the potential impact of sound velocity errors on the accuracy of soundings.

Table 1: Error Table for EM3000D

	Source of Error	Depth in Metres			
		50	75	100	150
		Error in Metres			
a.	Draught Setting	0.05	0.05	0.05	0.05
b.	Variation in Draught	0.05	0.05	0.05	0.05
c.	Velocity of Sound	0.17	0.23	0.29	0.41
d.	Spatial Variation in SV	0.10	0.10	0.25	0.40
e.	Temporal Variation in SV	0.05	0.05	0.05	0.05
f.	Application of Measured SV	0.05	0.05	0.05	0.05
g.	Depth Measurement	0.50	0.75	1.00	1.50
h.	Heave	0.05	0.05	0.05	0.05
i.	Settlement and Squat	0.15	0.15	0.15	0.15
j.	Roll, Pitch and Seabed Slope	0.05	0.08	0.10	0.15
k.	Co-tidal Readings	0.20	0.20	0.20	0.20
l.	Tide Corrections	0.00	0.00	0.00	0.00
Total Standard Error: $\sqrt{(a^2 + b^2 + \ldots)}$		**0.61**	**0.84**	**1.11**	**1.64**
Required accuracy:					
LINZ Standard MB-1		0.68	0.92	1.19	1.73
LINZ Standard MB-2		0.90	1.23	1.58	2.30
LINZ Standard MB-3		1.13	1.54	1.98	2.88

Accurate multibeam data collection requires the minimising of all these measureable errors. Most of these can be removed by the use of the same equipment, same platform and good work practices. Although the error budget may be 1-1.5m in the survey water depths, the comparative error between surveys can be reduced to a fraction of this by using the same set-up between surveys. Just changing the platform can introduce additional errors such as lever arm errors, draft, squat, and motion mismatch.

Depth measurement accuracy is greatest in the inner parts of the swath. Normal survey techniques use this fact to improve accuracy by ensuring sufficient overlap is obtained between lines to minimise the use of the less accurate outer swaths. The surveys to date have not done this resulting in data gaps between lines and increased inaccuracy in the surface comparisons.

Heave measurements supplied by most heave sensors maintain an accuracy of 5% of the measured vertical displacement or ±5 cm (whichever is the larger) for movements that have a period of up to 20 seconds (manufacturer's specifications). As a result, as the maximum swell increases, heave artifacts can be observed in processing of the data. The best method to mitigate this is to survey on relatively calm days only.

The effect of tides was removed from the MBES bathymetric data based on modelled tides. Although modelled tides can be used for this type of survey these are not accurate enough for this application, especially if the surface differences being measured are small and the dumped spoil is expected to spread over the entire surveyed area. If modelled tide data is to be used, calibration against a true water level recorder is recommended as the modelled data can, and often does, have errors in both timing and range values.

Without the provision of a detailed error budget the stated accuracy of the bathymetric survey is impossible to defend.

2.3 Report 3 - Plume monitoring

Comprehensive monitoring of plume dispersal in the coastal zone is non-trivial. Four surveys were completed during Autumn 2010 to determine the "dispersion potential of the suspended sediment". A variety of standard methodologies were employed to meet this objective, including CTD profiles, time series of turbidity, water sampling, drogues and ADCP surveys. Despite the somewhat routine sampling techniques used, the first two surveys should be considered pilot studies. Nevertheless, data from survey 1 and 2 were instructive to improve sampling during the latter two surveys, including vessel positioning and spatio-temporal demands for the rapidly descending plume. While all data were reviewed, there was a focus on observations from survey 4. Methodology from the last survey had been through several iterations and was considered best able to evaluate plume dispersal at the Auckland Marine Disposal Ground.

In general, turbidity sensor data and suspended sediment concentrations (SSC) from water samples did not correlate. The position of turbidity sensors on vessels outside of the affected region for surveys 1 and 2 was confirmed by low turbidity values. Yet, even when sensors were better positioned, higher in situ data did not correspond with higher sensor values (Fig 3.25 at station A after 10:40 am). Insufficient sampling by either technique in the core of the plume makes interpretation of turbidity concentrations difficult in such a transient feature. This shortcoming was identified and was accurately considered impractical for future plume monitoring.

Some concerns exist about water samples collected using Niskin/van Dorn bottle samplers and how these relate to what was observed from ADCP backscatter. In Fig. 3.80 the colour scale shows a relationship between backscatter in dB and concentration in mg/L. In the ADCP transects (Fig 3.29 to 3.71), backscatter was often indicating values of 103-110 dB, which equates to concentrations of around 80-100 mg/L. Yet, for surveys 1 to 4 only a single sample of 76.69 mg/L (11:50, Survey 1, site C) was close to these plume values. Since the plume settling and dispersal was rapid and transient, it is clear that timing water sample collection to high concentrations is complicated. It was surprising that none of the Niskin/van Dorn water samples during survey 4 even approached sediment concentrations > 50 mg/L that were inferred from backscatter. Calibration of ADCP backscatter should be incorporated into future plume monitoring.

ADCP surveys were found to be the most robust method for observing sediment pathways at the sediment disposal site. Transects must be sufficiently long to capture the full spatial extend of the plume. Several transects from survey 3 (i.e.,T4 and T6) were incomplete with high backscatter at either the upstream or downstream boundary. This made it difficult to discern the spatial scale of the plume in these transects. By continuing these beyond the high backscatter signals, the physical area influenced by sediment could be quantified and could be used to identify bed locations where material ultimately settled. By survey 4, transects spanned background-plume-background concentrations (i.e., Fig. 3.65 T7 or Fig.3.69, T6). This meant that vertical and horizontal scales of plume intrusion were far more objective than from earlier shortened transects. Future ADCP transects for monitoring should encompass low-high-low backscatter.

Typically, within 30 minutes of sediment disposal, ADCP backscatter from survey 4 showed the greatest plume concentrations to be below 70 m. After 1 hour, the majority of the backscatter signal is barely above background levels. Of concern, is some remnant material at 100 m depth (at around 60 minutes post disposal). While this was attributed to increased current velocities, drifter speeds were low for this survey (< 0.05 ms^{-1}). Also, vertical current speeds typically decrease away from the surface of the ocean. One recommendation would be to examine current velocities from the ADCP surveys, which would be a step towards quantifying the process responsible for sediment entrainment at 100 m.

The role of stratification cannot be ignored. The well-mixed surface layer that had a thickness of around 80 m which was identified from vertical conductivity/temperature/depth (CTD) profiles obtained during each survey. Below 80 m, temperature decreased linearly to approximately 14^{o} C near the sea bed for all four surveys. This type of profile was considered typical for Autumn (see Fig. 3.18). Vertical changes in temperature and salinity are not negligible as claimed in the summary of Survey 4. The presence of either temperature or salinity stratification can cause layers with unique characteristics to move at different speeds and/or direction. Mixing processes near the pycnocline may also be responsible for the spreading or trapping of sediments in this region. Subsequent repeat transects up to 1.5 hours post-disposal with the same alignment would help address the presence or absence of sediment entrainment below the well-mixed layer.

An obvious omission, which should be incorporated into future monitoring, is the addition of a turbidity sensor to the Seabird CTD profiler. Observations of the high density sediment plume were not possible due to limitations with the ADCP backscatter. Thus, collecting numerous CTD and turbidity profiles as close as practical to the time of hopper disposal would allow for the high sedimentation concentrations to be measured in situ. Together with calibrated ADCP backscatter (from water samples), this would provide suitable monitoring of plume evolution at the short time scales. ADCP backscatter would not need to be calibrated on each monitoring survey, but would need to be conducted on every third or fourth survey, depending on the frequency of surveys and material being disposed.

In summary, since future sediment disposal will be in similar (or calmer) conditions for wind and sea state to the surveys spanning plume monitoring, it is highly likely that the dispersed material will be localised to within 600 m or so of the disposal site. This lies well within the established boundary of 1500 m at the Auckland Marine Disposal Ground. Thus we broadly agree with the findings of Report 3, but mainly based on the results of survey 4.

2.4 Report 4 - Sediment characteristics

2.4.1 Sediment texture

Particle size data is provided for each sampling site. The material used for grain size analyses was sourced from the sediment cores collected and assessed within the ‘Benthic Cores’ section. The visual assessment of the cores states that there are two distinct sediment layers – a loose fluffy layer that overlies stiffer mud. The report does not indicate which sedimentary layer was used in the analysis, or even if a distinction was made. If no distinction was made then how were sub-samples extracted for analysis? Was homogeneity of the samples ensured, and if so how? Were pseudo random subsamples used i.e. what portion of the sample did the analysist happen to remove? If the latter was the case, was any documentation made about which sedimentary layer may have been analysed? This information needs to be documented in order to put the textural data in context, i.e. is variability seen due to haphazardly random subsampling of potentially different sediment layers or is it due to site location differences or due to dilution?

An analysis of material dredged for disposal was reported previously by a different provider and it is unclear whether there is any further information available from the prior report about how these samples were collected, prepared for analysis and analysed for sediment texture. This information would aid in a discussion of the apparent difference in reported texture of material from the site dredged and post–disposal material analysed here. It is important to determine whether this difference is due to the application of different analytical techniques, differences in sediment sub-sampling, or even dilution of the dredged material as a consequence of the sub-sampling methodology used here.

Grain size class information provided in Table 4.1 shows that at some sampling sites there was up to a 20% difference in the relative percentage of silt and clay e.g Site G3 66% pre and 46% post, Site H3 42% pre and 62% post. Also up to a 15% difference is reported between the control (X) sites and up to 20% between the monitoring sites. Several other sites exhibit ~ a 10% difference either temporally or spatially in relation to their cohorts. These results do not agree with the statements made that “this percentage was fairly consistent across the site .... and pre- and post- disposal” and “generally make up 55%”. The sediment texture variability must be discussed in relation to other reported sediment properties and characteristics. If a mean is to be reported then it needs to be calculated in a useful way e.g. mean of control sites, mean of monitoring sites greater than and closer to a stated number of metres from disposal site etc. A mean of all samples is comparatively meaningless. There is a lack of any appropriate statistics in this section which is unexpected considering the importance of these results.

The final sentence of this section tends to suggest that analyses of a subsample of all material (entire 12cm core) collected at each site were conducted; hence dilution is a viable reason for absence of the hypothesised elevated levels of silt and clay in samples from sites that visually contained disposal material. While the review brief does not cover methodology this does call into question the appropriateness of the inferred subsamples used in defining texture, and therefore the reported data having the power to distinguish change, or absence of change, and therefore impact, or lack thereof, of the disposed material.

The ‘Benthic Core’ report does not provide any indication of background sedimentation rates, nor are sedimentation rates reported in this section. The rate at which sediment naturally accumulates on the seafloor at this site is one metric that can be used to define the depth of material within a sediment core that can be combined for use in pre-, post-, and future monitoring analyses. For example, if only the top 2 cm of a core collected at a site was used in deriving textural information would an elevated level of silt and clay been reported? Alternatively the depth of occurrence of dredged material could be the guiding metric for depth to which sediment can be combined for analysis. The overarching choice of the down core depth to which sediment can be combined should consider the natural rate of accumulation, the disposed material potential accumulation rates, the ability to derive, represent, and to be sensitive to change should it occur, and the potential biases that may be introduced. A metric that does not have the power to detect change does not mean there is no change or negligible impact. Similarly, a metric that does have the power to detect change would not necessarily result in any reported change having a deleterious impact. A robust analysis that has the power to detect change should be used; subsequent to analysis, the significance of any reported results can then be considered with regard to impact.

2.4.2 Sediment Chemistry

The discussion around sub-sampling for textural analysis above is also appropriate to the samples used in this section for heavy metal and total petroleum hydrocarbons, and is not repeated at length here.

The potential dilution of any chemical signature by use of a subsample of all material (entire 12cm core) collected at each site, the potential bias this introduces, and/or the power to detect change needs to be considered despite many of the prior analyses of the dredged material, and analyses of disposed material in this report, falling below ER-L, ISQG-Low levels. Similar comments apply to the levels of TPH reported. As in the section above, the use of appropriate statistical analysis is recommended especially in light of statements such as “concentration increased slightly,...., decreased slightly,... increased significantly”.

The advisability of using only one control site for the chemical analysis given both the natural variability that occurs in the marine realm and the apparently anomalous results from this site is questioned here. Further the comparison of TPH at Site H2 to other sites should be expanded, e.g. which other sites? Are they appropriate for comparison?

2.4.3 Geomechanical Strength

Shear vane measurements were used to verify *Nimrod* shear strength profiles. Consequently shear vane measurements should be reported and an indication of the basis for the statement that these “matched reasonably well” provided. It would aid interpretation if shear strengths were also reported within the descriptors of the four types.

The hypothesis of a variation in geomechanical shear strength profiles as a result of disposal processes cannot be fully tested by the available data, as reported. Without a full post-disposal dataset one cannot discount naturally occurring geomechanical processes. Shear strength profiles were not conducted/reported from the control sites. Measurements at the control sites would aid in determining naturally occurring geomechanical processes in this region.

2.4.4 Conclusions

The range of grain size classes did not change pre- and post-disposal, however contrary to the conclusions the data indicate variability in silt and clay percentages across the monitoring and control sites.

Both textual and chemical conclusions need to be considered in light of potential subsample bias and, or, dilution.

2.5 Report 5 - Biological assessment

Potential problems with this assessment, likely to be raised in any Environment Court hearing, relate to both the analyses conducted and the accuracy of the findings.

- No statistical analyses of any differences observed between pre and post disposal are given. While some decreases are noted, the report then goes on to say, firstly, that these decreases may be due to seasonal and local variation, rather than disposal, and secondly, that the decrease is an anticipated effect of deposition of dredged material and therefore not unexpected!
- The discussion relates only to diversity (undefined although probably average number of taxa per core) and abundance. No multivariate community analyses (Anderson 2001, Anderson & Thompson 2004) or analyses of dominant taxa, or taxa that could be expected to be sensitive, are discussed.
- The monitoring programme had 11 replicates within the disposal site, 4 on its outer edges, 5 in the monitoring zone (altogether covering an area of about 3km in diameter). There were 6 control sites located 3-6 km away from the outer perimeter of the monitoring zone. It is not apparent from the design of the monitoring programme whether the intention was for the data to be analysed by ANOVA-based BACI (Before-After-Control-Impact techniques) or by the more sensitive gradient techniques (Ellis et al. 2000, Hewitt, J.E. et al. 2001). The design is not optimal for either but as a result could be analysed by either.
- The accuracy of the findings, such as they are, are based on a very few, small cores, that are unlikely to sample the area in sufficient detail as to determine any changes related to disposal. Unfortunately, no estimates of precision have been made and it is probably not worthwhile doing this given that the samples were frozen rather than preserved by more normal techniques. Freezing generally results in many of the softer bodied organisms (especially smaller polychaetes common in soft-sediment at a range of depths) disintegrating.
- The larger scale sampling conducted using the underwater video footage only recorded organisms, rather than (as is usual, see Auster et al. 1997, Thrush et al. 2002) details of organism activities (e.g., worm casts, crab burrows, mounds and tracks made by organisms moving around on or just under the sediment surface). Indeed from the examples of footage in the report, the tilt of the camera and the uneven lighting may well make it difficult to count these features.

2.6 Report 6 - Plume imagery and underwater video

The aerial photography of the plume was useful in providing understanding of the mechanics of disposal operations and a scale of the plume. It did not provide any quantitative data.

Echo sounder imagery provided a clear visual display of the rapid settling of disposal sediments at ~ 0.2 ms^{-1} during survey 4. It is interesting to note the layer of sediment entrained around and below the 80 m marker of the screen that have turbidity concentrations are greater than background levels. Comparisons to ADCP backscatter from report 3 suggest that the dispersion was still visible at 100 m at 60 minutes after sediment disposal.

The images of the sea floor were of poor quality, with very uneven lighting and had no scaling system so it was difficult to tell the size of the seafloor features. The seafloor image survey was conducted 3-4 months after the dredge material was deposited by which time, as the authors noted, it was likely that the original anoxic black muds had oxidised making them impossible to distinguish from native sea bed by colour alone. This delay, the poor quality of the imaging, and the lack of scale in the images greatly diminish the usefulness of the seafloor imaging in determining the spatial spread of the dredge material from the disposal location. Little confidence can be placed in the conclusions drawn from these images.

3 Do the reports meet the initial objectives?

3.1 Monitoring objectives

The main objectives of the monitoring reports were to:

1. Confirm that the site is not ecologically significant.
2. Determine the fate of dredged material disposed at the site.
3. Assess the dispersion potential of a dredged material plume arising from the disposal process.
4. Evaluate the methodologies used.
5. Make recommendations for future monitoring. This objective is wider than the impact hypothesis with Objectives 1.0, 2.0 and 3.0 covering those matters listed in the impact hypothesis.

Whether each of these objectives was met is addressed in turn below.

3.2 Is the disposal site ecologically significant?

This question has been left largely unanswered, due to lack of any formal analysis in the biological assessment report. We note that in this report:

- No definition of “ecologically significant” was discussed and tested.
- No statistical comparisons between the “control” area (sites X1 – X6) and the disposal or monitoring zones in 2009 appear to have been conducted. In fact there is no

discussion of any difference in faunal composition or diversity between different areas for the 2009 data.

- There is only limited discussion related to the similarity of the organisms collected in this study and those from a similar area in 2002 (Sivaguru and Grace 2002). However, again there is no formal statistical comparison and it is not clear whether the statements as to similarity are based on data from both years (2009 and 2010) and the entire study area (control, monitoring and disposal). In fact it is not clear what the "similarity" they discuss actually means. A statement is made that the disposal site is less diverse and this may be due to its location (further southeast and "somewhat" deeper) but no references are provided to why these factors should result in decreased diversity.
- Different components of biodiversity (i.e., alpha, beta or gamma) are not discussed, although differences between these are important both for the determination of ecological significance and analysing impacts (Hewitt, J. et al. 2010, Loreau et al. 2001, Naeem 2002). No species accumulation curves are presented, a necessary component of assessing biodiversity and determining whether the 21 samples collected (only 13 of which were in the disposal zone) are indeed "sufficient to represent types of organisms" (Arrhenius 1921, Ugland et al. 2003). This is especially important, given the small size of the corer used.

3.3 What is the fate of dredged material disposed at the site?

Several reports (1, 2, 3, 4, and 6) were relevant to this objective and achieved the objective to a greater or lesser degree.

Report 3 on plume monitoring indicated that it was highly likely that the dispersed material was localised to within 600 m or so of the disposal site. This lies well within the established boundary of 1500 m at the Auckland Marine Disposal Ground. However, there were several aspects of plume monitoring which need attention in any future monitoring and these are discussed in section 4.0.

Repeated multibeam echosounder surveys (Report 2) have the potential to define where exactly the material eventually landed. But without the provision of a detailed error budget the stated accuracy of the bathymetric survey is impossible to defend and the ability to distinguish 'before' and 'after' bathymetric profiles may be compromised if the errors were in fact larger than stated.

The images of the sea floor reported in Report 6 (Plume imagery and underwater video) were of poor quality, with very uneven lighting and had no scaling system so it was difficult to tell the size of the seafloor features. The seafloor image survey was conducted 3-4 months after the dredge material was deposited by which time, as the authors noted, it was likely that the original anoxic black muds had oxidised making them impossible to distinguish from native sea bed by colour alone. This delay, the poor quality of the imaging, and the lack of scale in the images greatly diminish the usefulness of the seafloor imaging in determining the spatial spread of the dredge material from the disposal location. Little confidence can be placed in the conclusions drawn from these images.

Lack of comparisons of sampled sediment colours to standard colour charts in Report 1 (Benthic cores) may have led to errors in judgement of sediment characteristics between analysts, between samples, and between monitoring and control sites. Additionally, a gravity corer and the bow wave of such corers, as well as the necessary retrieval mechanism, can still vertically disturb the sediment surface during coring. Thus, evidence, or lack thereof, of bed forms at the sediment surface should be viewed with caution. Lastly, photographs of cores supplied in the appendices, are mostly lacking an appropriate scale thus statements in the report concerning the amount, concentration or significance of different sediments cannot be supported.

In Report 4 (Sediment characteristics) it appears that the whole 12 cm core was analysed for sediment characteristics and contaminants, thus potentially diluting any effect of the recently deposited dredge material. This may explain why elevated levels of silt and clay, heavy metals and total petroleum hydrocarbons were not observed as expected given the known concentrations in the dredged material. This is a potentially serious problem.

3.4 Assess the dispersion potential of a dredged material plume arising from the disposal process.

The aerial photography of the plume (Report 6) was useful in providing understanding of the mechanics of disposal operations and a scale of the plume. It did not provide any quantitative data. Echo sounder imagery provided a clear visual display of the rapid settling of disposal sediments at ~ 0.2 ms^{-1} during survey 4.

Report 3 on plume monitoring indicated that it was highly likely that the dispersed material was localised to within 600 m or so of the disposal site. This lies well within the established boundary of 1500 m at the Auckland Marine Disposal Ground. However, there were several aspects of plume monitoring which need attention in any future monitoring and these are discussed in section 4.0.

3.5 Evaluate the methodologies used

The reports do a rather poor job of evaluating the methodologies used as evidenced by the detail provided in our reviews of the reports in section 2 above. Evaluating your own methodology is always difficult and this task is best given to someone knowledgeable about the requirements of the study and methodology, but uninvolved in the actual work. This peer review process is important at all stages of research, not just after submission of a report.

3.6 Recommendations for future monitoring

Plume monitoring

The summary report suggests that no further plume monitoring is required if dredge spoil is disposed at a sea state equating to a Beaufort Scale 5 or 6 or less. We disagree and suggest that the role of stratification in keeping a proportion of disposed sediment in the water column for long periods cannot be ignored. One recommendation would be to examine current velocities from the ADCP surveys, which would be a step towards quantifying the process responsible for sediment entrainment at 100 m. Repeat transects up to 1.5 hours post-disposal with the same alignment would help address the presence or absence of sediment entrainment below the well-mixed layer. Future ADCP transects for monitoring should encompass low-high-low backscatter. Future monitoring should ensure that a turbidity sensor is added to the Seabird CTD profiler.

Multibeam echosounder: seafloor geomorphology

The summary document made two recommendations in relation to future multibeam echosounder surveys of the same site:

- That bathymetry change and backscatter variation should be examined and compared to previous datasets.
- If a disposal mound is detected, a volumetric estimate should be derived for comparison to future estimates.

We agree that these are worthwhile recommendations. However, as multibeam data collection can be strongly affected by the survey environment, a consistent and carefully thought out work programme and project methodology is required if the data collected are to be directly comparable, survey to survey. We recommend the compilation of an error budget for the previous survey and for any future multibeam surveys.

Sediment coring and analysis

The summary report makes several recommendations in relation to sediment coring and analysis:

- Collect 6 cores including 1 control core.
- Make visual observations of the core looking at color, layering, and other distinctive physical characteristics.
- Undertake heavy metal and TPH analysis to assess chemical characteristics of the site and if possible trace disposed material by comparing concentrations to those comprising the chemical signature of the material disposed there.
- Compare all data to previous post-disposal core data and look for evidence of the presence of dredge material.

We agree there is a need to collect more core material but suggest that 6 cores is much too few to draw strong inferences about the likely dispersal of dredge material and contamination of the native seabed sediments.

We also agree that visual observations of core colour and layering and other distinctive physical characteristics would be useful but caution against assessing these in the way previously used. Standard colour charts must be used to mitigate observer perception bias in assigning colours and a physical scale must be used to describe layering, layer thickness and other physical characteristics.

We agree that heavy metal and TPH analysis should be undertaken on any core material collected. However, we suggest that first there should be some preliminary study to determine background rates of sediment accumulation on the seafloor at the disposal site. This would then provide the necessary information to decide what thickness of sediment should be analysed to determine contamination in relation to annual background rates. Without this study the decision to analyse contaminants from the entire core or just part of the core is completely arbitrary.

We also agree that comparison of new core material to previous material should be undertaken if the caveats we have raised about previous sampling methodology allow clear conclusions to be drawn from such a comparison.

We further suggest that sediment grain size analysis be undertaken at intervals down the core. This is likely to help identify native sea floor material from disposed material.

Biological Assessment

The summary report makes four recommendations for future biological monitoring and another is made in Report 5. These are:

- Collection of 6 box dredge (or similar) samples including one from a control site. Report 5 explains that the use of a larger sampling device to sample benthic fauna is in order to obtain a more representative sample.
- Improve preservation of biological material
- Sort, identify, and count benthic organisms in each sample
- Assess diversity, abundance, and compositional change and compare to data from previously collected samples
- Report 5 also suggests that if underwater video is to be used in the future improvements are necessary to increase the quality of data relevant to the benthic fauna.

We agree that it is definitely important to increase the size of the sample unit from the 7cm diameter corer initially used. However, a box dredge will, as with all dredges, only provide semi-quantitative information, as it is difficult to determine the area sampled. Instead a box corer or grab designed to eliminate 'washout' on ascent would be appropriate.

We completely agree that immediate preservation by a standard technique is also necessary. Preservation could either be by 70% IPA or, as recommended by 10% formalin for at least 24 hours before transferring to 70% IPA. If the latter technique is to be used it is important that the formalin mix be buffered otherwise calcium material may begin to dissolve.

Six samples, with just one from a 'control' site are completely inadequate to draw any conclusions given the expected variance in faunal numbers and species composition among samples. More importantly, if the monitoring is to be able to determine whether any environmental effects of disposal occur or not, the sampling programme and analysis needs to be designed to achieve this. This analysis will have to be of the Control-Impact type, rather than a BACI, due to the changes in sampling required. The number of samples needed to detect the size of change deemed significant will need to be determined based on the new sampling method. Given these changes it would also be appropriate to change some of the sample locations in order to design an optimal gradient study.

The suggestions for the future monitoring should include what aspects of the biology are to be assessed. Simple diversity and abundance is not sufficient for detection of environmental impacts. Instead aspects of diversity based on both number and abundance of taxa should be analysed, together with community composition, functional attributes, dominant taxa and taxa predicted from other work to be sensitive to dredge disposal.

We agree that a different imaging system would be needed, preferably involving some scaling lights so that actual area sampled could be determined. However, it is also important to record sedimentary features resulting from organisms' activity as well as actual organisms.

Finally, the recommendations for monitoring should include a time period. How soon after the disposal will the monitoring take place? If this is too long afterwards, then recovery may already be taking place. If too soon, the full implications of disposal may not be seen.

General recommendations

The summary report recommends that:

- If no evidence of loss of material or migration of the disposed material beyond the boundaries of the site is detected, then disposal operations can resume.
- The monitoring should be repeated after disposal of 150,000 m^3 and 300,000 m^3 of dredged material.
- If evidence of loss of material or migration of the disposed material beyond the boundaries of the site is detected, then disposal operations should be suspended pending further assessment.

We agree with these recommendations but caution that the monitoring programme be substantially improved along the lines suggested in order to increase confidence in the spatial extent of dredged material on the sea floor and its impact on the benthos.

4 Discussion and Conclusions

4.1 Context

A review has been undertaken for Maritime New Zealand of a report written as part of the monitoring of effects of a permitted dredge spoil dumping operation at a new dumping site 25 km east of Great Barrier Island. The report actually comprised six individual reports on different aspects of the monitoring programme and a summary report that also made recommendations for future monitoring.

The review examines the content of each report and determined whether the initial objectives (as specified in the Summary Report) had been met by the work undertaken. This entailed reviewing the data collected during the monitoring program and their accuracy and interpretation. Although MNZ specified the review was not to include the methodology used in the collection of the monitoring data, in places it proved impossible to separate the veracity of the findings from the methodology used to collect it.

Lastly the proposals for future monitoring have been reviewed and feedback provided as to its likely effectiveness, including comments on the proposed timeframes for monitoring in relation to dumping volumes, and whether the specific study elements proposed will sufficiently cover potential environmental effects of marine dredge disposal.

4.2 Reviews of individual reports

Parts of the monitoring programme was found to be deficient to some extent.

Monitoring of the plume during disposal of the dredge spoil (Reports 3 and 6) was adequately done and while we have identified areas where this could be improved we have confidence in the conclusion that it was highly likely that the dispersed material was localised to within 600 m or so of the disposal site. This lies well within the established boundary of 1500 m at the Auckland Marine Disposal Ground.

Comparative surveys are the main tool used to determine the bathymetric impact of spoil dumping on the seafloor. As multibeam data collection can be strongly affected by the survey environment, a consistent and carefully thought out work programme and project methodology is required if the data collected are to be directly comparable, survey to survey. Unfortunately a detailed error budget for the multibeam survey was not presented. Without this the stated accuracy of the bathymetric survey is impossible to defend.

Cores were taken to determine if the sea floor sediments had altered in response to the deposition of near shore clays and muds. Unfortunately the methodology employed was inadequate to draw firm conclusions. We concluded that a lack of comparisons of sampled sediment colours to standard colour charts in Report 1 (Benthic cores) may have led to errors in judgement of sediment characteristics between analysts, between samples, and between monitoring and control sites. We also noted that photographs of cores supplied in the appendices, are mostly lacking an appropriate scale thus statements in the report concerning the amount, concentration or significance of different sediments cannot be supported.

The sediment characteristics were also examined in Report 4. However, it appears that the whole 12 cm core was analysed for sediment characteristics and contaminants, thus potentially diluting any effect of the recently deposited dredge material. This may explain why elevated levels of silt and clay, heavy metals and total petroleum hydrocarbons were not observed as expected given the known concentrations in the dredged material. This is a potentially serious problem.

There were problems with interpreting images of the sea floor in Report 6 (Plume imagery and underwater video). There was a 3-4 month delay between disposal of dredge material and the seafloor image survey, during which time oxidation of previous anoxic black muds took place making it impossible to distinguish between disposed and native sediments. This delay, the poor quality of the imaging, and the lack of scale in the images greatly diminished the usefulness of the seafloor imaging in determining the spatial spread of the dredge material from the disposal location. Little confidence can be placed in the conclusions drawn from these images.

We found serious problems with the biological assessment (Report 5). No statistical analyses of any differences observed between pre and post disposal were given. The discussion relates only to diversity (undefined although probably average number of taxa per core) and abundance. No multivariate community analyses or analyses of dominant taxa, or taxa that could be expected to be sensitive, are discussed. It was not apparent from the design of the monitoring programme whether the intention was for the data to be analysed by ANOVA-based BACI (Before-After-Control-Impact techniques) or by the more sensitive gradient techniques. The design is not optimal for either but as a result could be analysed by either. The accuracy of the findings, such as they are, are based on a very few, small cores, that are unlikely to sample the area in sufficient detail as to determine any changes related to disposal. The larger scale sampling conducted using the underwater video footage only recorded organisms, rather than details of organism activities.

4.3 Recommendations for future monitoring

The summary report suggests that no further plume monitoring is required if dredge spoil is disposed at a sea state equating to a Beaufort Scale 5 or 6 or less. We disagree and suggest that the role of stratification in keeping a proportion of disposed sediment in the water column for long periods cannot be ignored.

The summary report recommended that multibeam echosounder surveys be used to monitor the accumulation of disposal sediments on the seafloor. We agree and recommend the compilation of an error budget for the previous survey and for any future multibeam surveys.

The summary report recommended that in future monitoring 6 sediment cores be collected, that these be examined for color, layering, and other distinctive physical characteristics and analysed for heavy metal and TPH contamination and comparisons made with previous core data. We agree but suggest that 6 cores is too few to draw strong inferences about the likely dispersal of dredge material and contamination of the native seabed sediments.

We also agree that visual observations of core colour and layering and other distinctive physical characteristics would be useful but caution against assessing these in the way previously used. Standard colour charts must be used to mitigate observer perception bias in assigning colours and a physical scale must be used to describe layering, layer thickness and other physical characteristics.

We agree that heavy metal and TPH analysis should be undertaken on any core material collected. However, we suggest that first there should be some preliminary study to determine background rates of sediment accumulation on the seafloor at the disposal site. This would then provide the necessary information to decide what thickness of sediment should be analysed to determine contamination in relation to annual background rates. Without this study the decision to analyses contaminants from the entire core or just part of the core is completely arbitrary.

We further suggest that sediment grain size analysis be undertaken at intervals down the core. This is likely to help identify native sea floor material from disposed material.

The summary report makes several recommendations with regard to future biological monitoring. We agree that it is definitely important to increase the size of the sample unit from the 7cm diameter corer initially used. However, a box dredge will, as with all dredges, only provide semi-quantitative information, as it is difficult to determine the area sampled. Instead a box corer or grab designed to eliminate 'washout' on ascent would be appropriate.

We completely agree that immediate preservation by a standard technique is also necessary. Preservation could either be by 70% IPA or, as recommended by 10% formalin for at least 24 hours before transferring to 70% IPA. If the latter technique is to be used it is important that the formalin mix be buffered otherwise calcium material may begin to dissolve.

Six samples, with just one from a 'control' site are completely inadequate to draw any conclusions given the expected variance in faunal numbers and species composition among samples. More importantly, if the monitoring is to be able to determine whether any environmental effects of disposal occur or not, the sampling programme and analysis needs to be designed to achieve this. This analysis will have to be of the Control-Impact type, rather than a BACI, due to the changes in sampling required. The number of samples needed to detect the size of change deemed significant will need to be determined based on the new sampling method. Given these changes it would also be appropriate to change some of the sample locations in order to design an optimal gradient study.

The suggestions for the future monitoring should include what aspects of the biology are to be assessed. Simple diversity and abundance is not sufficient for detection of environmental impacts. Instead aspects of diversity based on both number and abundance of taxa should be analysed, together with community composition, functional attributes, dominant taxa and taxa predicted from other work to be sensitive to dredge disposal.

We agree that a different imaging system would be needed, preferably involving some scaling lights so that actual area sampled could be determined. However, it is also important to record sedimentary features resulting from organisms' activity as well as actual organisms.

Finally, the recommendations for monitoring should include a time period. How soon after the disposal will the monitoring take place? If this is too long afterwards, then recovery may already be taking place. If too soon, the full implications of disposal may not be seen.

The summary report recommends that:

- If no evidence of loss of material or migration of the disposed material beyond the boundaries of the site is detected, then disposal operations can resume.
- The monitoring should be repeated after disposal of 150,000 m^3 and 300,000 m^3 of dredged material.
- If evidence of loss of material or migration of the disposed material beyond the boundaries of the site is detected, then disposal operations should be suspended pending further assessment.

We agree with these recommendations but caution that the monitoring programme be substantially improved along the lines suggested in order to increase confidence in the spatial extent of dredged material on the sea floor and its impact on the benthos.

5 Acknowledgements

We thank Neville Ching for proposal and project administration and Phil Heath for reviewing the report.

6 References

Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. *Australian Ecology 26*: 32-46.

Anderson, M.J.; Thompson, A.A. (2004). Multivariate control charts for ecological and environmental monitoring *Ecological Applications 14*: 1921-1935.

Arrhenius, O. (1921). Species and area. *Journal of Ecology 9*: 95-99.

Auster, P.J.; Malatesta, R.J.; Donaldson, C.L.S. (1997). Distributional responses to small-scale habitat variability by early juvenile silver hake, *Merlucius bilinearis*. *Environmental Biology of Fishes 50*: 195-200.

Ellis, J.I.; Schneider, D.C.; Thrush, S.F. (2000). Detecting anthropogenic disturbance in an environment with multiple gradients of physical disturbance, Manukau Harbour, New Zealand. *Hydrobiologia 440(1-3)*: 379-391.

Hewitt, J.; Thrush, S.; Lohrer, A.; Townsend, M. (2010). A latent threat to biodiversity: consequences of small-scale heterogeneity loss. *Biodiversity and Conservation 19*: 1315-1323. <http://dx.doi.org/10.1007/s10531-009-9763-7>

Hewitt, J.E.; Thrush, S.F.; Cummings, V.J. (2001). Assessing environmental impacts: effects of spatial and temporal variability at the scale of likely impacts. *Ecological Applications 11*: 1502-1516.

Loreau, M.; Naeem, S.; Inchausti, P.; Bengtsson, J.; Grime, J.P.; Hector, A.; Hooper, A.; Huston, D.U.; Raffaelli, D.; Schmid, B.; Tilman, D.; Wardle, D.A. (2001). Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science 294*: 804-808.

Naeem, S. (2002). Ecosystem consequences of biodiversity loss: The evolution of a paradigm. *Ecology 83(6)*: 1537-1552.

Thrush, S.F.; Schultz, D.; Hewitt, J.E.; Talley, D. (2002). Habitat structure in soft-sediment environments and the abundance of juvenile snapper (*Pagrus auratus* Sparidae): Developing positive links between sustainable fisheries and seafloor habitats. *Marine Ecology Progress Series 245*: 273-280.

Ugland, K.I.; Gray, J.S.; Ellingsen, K.E. (2003). The species-accumulation curve and estimation of species richness. *Journal of Animal Ecology 72*: 888-897.