Marine Sediment Contaminant Monitoring Programme: Review of Data Quality and Procedures

December 2014 Technical Report 2014/041

Auckland Council
Technical report TR2014/041
ISSN 2230-4525 (Print)
ISSN 2230-4533 (Online)

ISBN 978-1-927216-36-1 (Print)
ISBN 978-1-927216-37-8 (Pdf)

This report has been peer reviewed by the Peer Review Panel.

Submitted for review on 18 December 2013
Review completed on 1 December 2014
Reviewed by two reviewers

Approved for Auckland Council publication by:

Name: Regan Solomon

Position: Manager, Research and Evaluation

Date: 23 December 2014

Recommended citation:

Mills, G and Williamson, B (2014). Marine sediment contaminant monitoring programme: review of data quality and procedures. Prepared by Diffuse Sources Ltd for Auckland Council. Auckland Council technical report, TR2014/041

© 2014 Auckland Council

This publication is provided strictly subject to Auckland Council's copyright and other intellectual property rights (if any) in the publication. Users of the publication may only access, reproduce and use the publication, in a secure digital medium or hard copy, for responsible genuine non-commercial purposes relating to personal, public service or educational purposes, provided that the publication is only ever accurately reproduced and proper attribution of its source, publication date and authorship is attached to any use or reproduction. This publication must not be used in any way for any commercial purpose without the prior written consent of Auckland Council. Auckland Council does not give any warranty whatsoever, including without limitation, as to the availability, accuracy, completeness, currency or reliability of the information or data (including third party data) made available via the publication and expressly disclaim (to the maximum extent permitted in law) all liability for any damage or loss resulting from your use of, or reliance on the publication or the information and data provided via the publication. The publication, information, and data contained within it are provided on an "as is" basis.

Marine Sediment Contaminant Monitoring Programme: Review of Data Quality and Procedures

Dr Geoff Mills

Dr Bruce Williamson

Diffuse Sources Ltd

Executive summary

This review has brought together a number of strands of information to provide a picture of data quality and operational issues affecting Auckland Council's (AC's) sediment contaminant monitoring programmes.

Quality assurance (QA) data reported in the Regional Sediment Contaminant Monitoring Programme (RSCMP), predecessor programmes (the State of the Environment (SoE) and Regional Discharges Project (RDP)), and other programmes in which sediment contaminant monitoring has been undertaken (e.g. the Upper Waitemata Harbour (UWH) benthic ecology programme), have been compiled and analysed to assess the accuracy and variability of sediment contaminant (heavy metals – copper (Cu), lead (Pb), and zinc (Zn) – and polycyclic aromatic hydrocarbons; PAH) and particle size distribution (PSD) data.

Based on these assessments, recommendations for future monitoring and investigation are made. Commentary on sites, sampling approach, analysis methods, and quality assurance protocols are also made.

Key findings:

- A variety of QA approaches have been used since 1998 when monitoring began. This made evaluating and concisely reporting the QA information difficult, and the results were sometimes not clear-cut.
- Since 2009 most QA procedures have been standardized and additional QA measures have been added. This now allows for comprehensive evaluation and reporting of QA data. However, clear objectives for the QA data still need to be defined and a formalised QA plan implemented.
- Within-year data variability for the metals Cu, Pb, and Zn was similar across the RDP, SoE, and UWH programmes, suggesting that differences in sampling, sample processing, and analysis procedures had no major effect on variability. No major changes in data variability have occurred over time, indicating that changes in monitoring and analytical practices have also not greatly affected variability.
- Agreement between blind duplicates for metals and PAH has usually been good (i.e. within acceptability criteria for US sediment monitoring programmes), but large differences between individual results have occasionally been recorded. Analysis of replicates from each site (as has been undertaken for most analytes to date) is therefore required to ensure robust data are obtained.
- For coarser textured sediments with low mud content, within-year variability for metals was generally higher than that for muddier sediments, and agreement between blind duplicates poorer. This higher variability should be taken into account when undertaking trend monitoring at coarser-textured sites (trends may take longer to emerge at these sites).
- The accuracy of the metals data was assessed by regular analysis of Certified Reference Material (CRM). Results for total recoverable metals and PAH have been good, with nearly all individual results being within 20% of the certified reference values. Concentrations of total recoverable metals and PAH are therefore likely to be reasonably accurate.

- However, the variability in the results for individual CRM sample analyses indicates that analytical variability is probably a major contributor to overall monitoring data variability. Continued investigations into key source(s) of analytical variability, in order to reduce it as much as possible, are worthwhile.
- Trends in CRM results for metals (Cu, Pb, and Zn) monitored since 2002 have been relatively small, up to ±0.6% per year. Only the trend for total recoverable Cu (+0.6% per year) was statistically significant. Based on the CRM results, trends of less than ±1% per year may therefore be affected by analytical variation over time. This approximately ±1% per year could therefore be used as a lower guideline for assessing “meaningful” trends over time.
- Data quality was similar to that required in three United States of America (USA) monitoring programmes. The Auckland Council monitoring data would probably meet the measurement quality objectives set for these programmes, except for agreement between blind duplicates (precision). Reducing the frequency of high variability results (e.g. those found in coarser textured sediments) would be required to meet these US programme measurement quality objectives.
- Comparisons of metals’ variability in the Auckland Council programmes with that measured in a Greater Wellington Regional Council (GWRC) sub-tidal sediment monitoring programme found markedly higher variability in the AC programmes. Further investigations are recommended to establish why this may be, and if appropriate, changes made to the AC programmes in line with GWRC protocols.
- A first look at estimating sample numbers required for trend assessment indicates a need to reduce data variability in order to have a reasonable chance of detecting even moderate sized trends, with a reasonable level of statistical certainty, within useful time frames. Data users need to be involved in defining their needs in terms of key variables: size of trends that are important for meeting resource management objectives, time frames required for trend detection/reporting, and budgets available for monitoring.

Generally, the review found that data quality has been good – mostly meeting data quality objectives prescribed for some major USA monitoring programmes. Aspects requiring improvement include reducing the magnitude and frequency of outlying results to reduce data variability, and the development and implementation of a consistent approach to data quality management.

Key recommendations

The key recommendations of this review are as follows:

1. Formalise QA procedures (based on those developed in the programme to date and currently implemented) in a unified quality assurance plan that clearly defines data quality objectives and protocols for sediment chemistry and PSD analysis.
2. Ensure all future AC sediment contaminant and PSD monitoring complies with the requirements set out in the QA plan.
3. Consult with analytical providers (laboratories) to determine what improvements could be made to current analytical methods to reduce data variability.

4. Consult with GWRC to identify (if possible) reasons for the lower variability obtained their monitoring programme.
5. Further investigate the trend detection capabilities of the monitoring programme in order to provide data end-users with a more quantitative definition of programme capabilities (see 8 below), and to guide future sampling and analysis protocols (e.g. number of replicates taken and analysed each year, sampling frequency etc). A statistical analysis of data variability and its effects on sampling requirements for temporal trend assessment is required to advance this aspect of the programme design. This could be done by extending the initial assessment presented in section 2.7.
6. Update the AC monitoring “blueprint” (ARC 2004) to include new QA protocols, analytical methods, sampling approaches and sample processing procedures as a result of the outcomes of work undertaken in items 1 to 5 above.
7. Review the suitability of all monitoring sites for on-going monitoring, in particular the sites’ physical characteristics (e.g. erosion status, mangrove encroachment) and monitoring data variability. Based on these assessments, prepare a list of core sites that are most suitable for on-going trend monitoring.
8. Investigate further the trend detection capabilities of monitoring at important sites (see item 5. above). Consult with data end-users to better define their needs in terms of site locations and key variables (size of trends that are important for meeting resource management objectives, time frames required for trend detection and reporting, and budgets available for monitoring). Use this information to refine a list of core trend monitoring sites and monitoring approaches (frequency, sample numbers etc). As far as practicable, document the likely time frames and magnitudes of trend detection for these sites.

The review has also highlighted how uncertainty can be created when changes in data providers (such as laboratories) occur. There is a need for a greater number of potential analytical providers that are capable of carrying out the analyses required by AC to run its monitoring programme, and for these providers to be able to clearly demonstrate that their capabilities meet AC’s quality requirements. Involving a wider range of laboratories, provided they can meet AC quality requirements, would be worthwhile for ensuring the long-term security of the monitoring programmes.

Data users should be involved in the future refinement of the monitoring programme, to ensure their key needs are met (as far as possible) and that they are aware of, and comfortable with, the limitations of the data produced. As identified in items 5 and 8 above, further work on better defining the trend detection capabilities of the programme is one area that stands out as being important in this regard.

Table of contents

List of figures

List of tables

1.0 Introduction

1.1 Background

Sediments accumulate many chemical contaminants originating from land-based activities. Sediment contamination therefore provides a useful indicator of land use impacts on aquatic receiving environments.

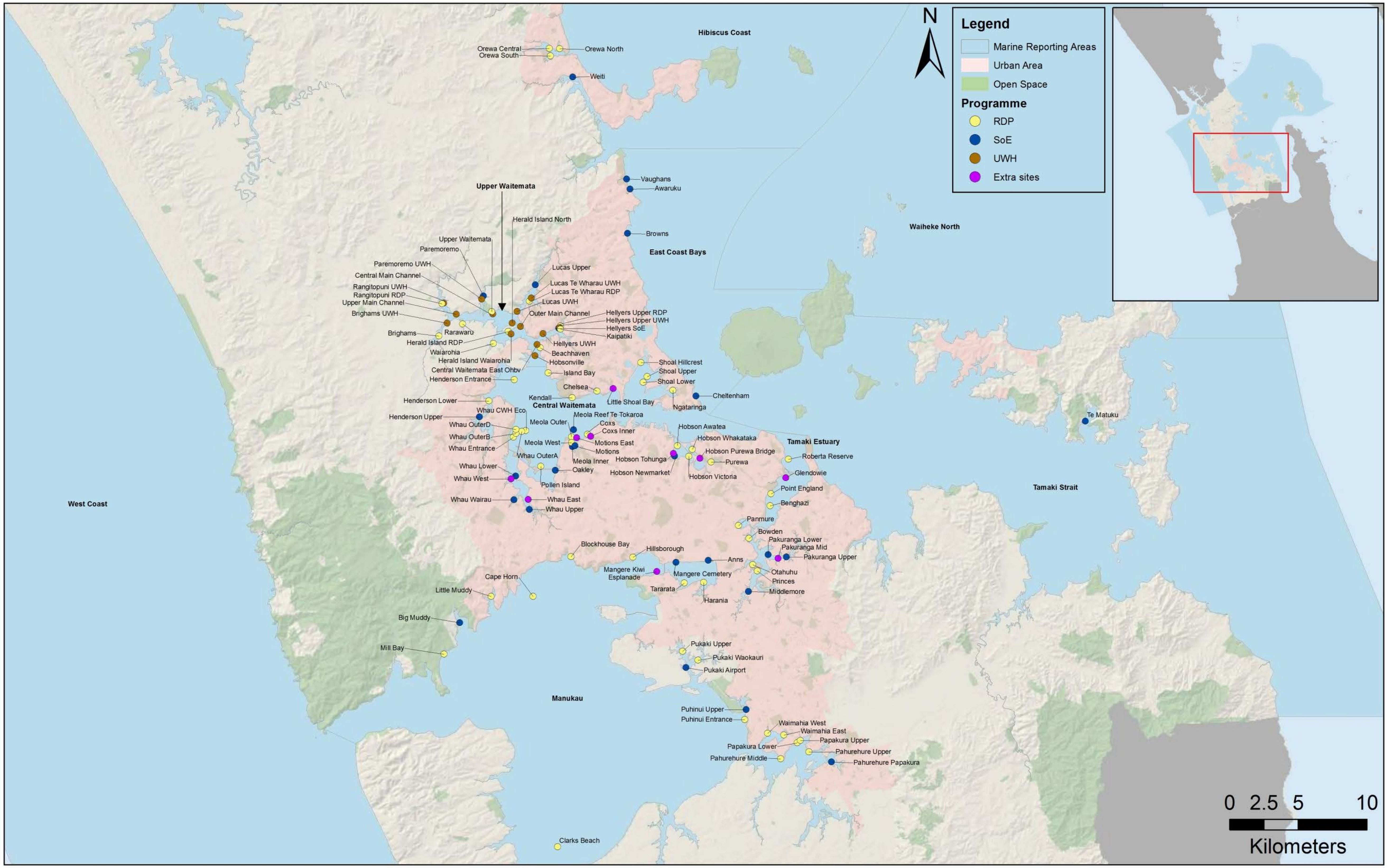
In 1998, the Auckland Regional Council (ARC; now Auckland Council) initiated a sediment contaminant monitoring programme aimed at assessing the spatial distribution and temporal trends in key chemical contaminants across the region's urban estuaries, harbours, and beaches. Key objectives of this State of the Environment (SoE) monitoring programme were to assess the effects of catchment land use, in particular urbanisation, on marine environmental quality, and the effectiveness of resource management initiatives and policies in mitigating adverse effects arising from land use activities.

Subsequently, two additional programmes have been used to acquire sediment contaminant data – the Regional Discharges Project (RDP), and the Upper Waitemata Harbour Benthic Ecology Programme (UWH).

Briefly, these complementary programmes were as follows:

1. State of the Environment (SoE) marine sediment monitoring programme, with 27 sites, monitored every two years since 1998. This programme aimed to provide long-term broad-scale information on contaminant status and trends across the region;
2. Regional Discharges Project (RDP), which monitored an additional 51 sites, at 2–5 yearly intervals (depending on their contamination status; see Kelly 2007). Monitoring in the RDP began in 2002, and was administered by the ARC on behalf of the region's Territorial Local Authorities (TLAs). This programme was aimed primarily at monitoring the effects of stormwater discharges, as part of the TLA stormwater network discharge consenting programme; and
3. The UWH programme, which has monitored up to 14 Upper Waitemata Harbour sites annually from 2005. This programme provides specific information on the effects of urban development on the Upper Waitemata Harbour.

The locations of the sites monitored in these programmes are shown in Figure 1-1.



Date: 6 June 2012

Location of sediment sampling sites and associated programme

Map Produced by
Research, Investigations
& Monitoring Unit.
Auckland Council

Figure 1-1 Site locations for the sediment contaminant monitoring programmes

Auckland Council (AC) has continued the former ARC sediment contaminant monitoring programmes, and, in order to achieve efficiencies and cost savings, has integrated the contaminant monitoring components of the SoE, RDP, and UWH programmes into a single programme, the Regional Sediment Contaminant Monitoring Programme (RSCMP).

Additional sediment contaminant sampling has also been carried out in conjunction with benthic ecology monitoring in a number of other estuaries and programmes around Auckland, including the Kaipara and Mahurangi Harbours, and Whangateau, Wairoa, Waiwera, Puhoi, Mangemangeroa, Waikopua, Okura, Turanga, and Orewa estuaries. These programmes are not currently part of the on-going sediment contaminant monitoring programme. Data for these sites can be found in reports appearing on the Auckland Council website – to date these include: Hailes et al. (2010) for the Kaipara Harbour; Townsend et al. (2010) for the Whangateau Harbour; Halliday and Cummings (2012) for the Mahurangi Estuary; Hewitt and Simpson (2012) for Waiwera, Puhoi, Mangemangeroa, Waikopua, Okura, Turanga, and Orewa estuaries and Lohrer et al. (2012) for the Wairoa embayment.

Details of the monitoring programme design, operation, and results are given in a number of reports; e.g. ARC (2004), Kelly (2007), Lundquist et al. (2010), Mills et al. (2012) and Mills (2014b). Data from the former SoE and RDP monitoring programmes have been compiled and reported annually; e.g. Reed and Gadd (2009) for the SoE, and Diffuse Sources (2007 to 2011) for the RDP. More recently, investigations of emerging contaminants or chemicals of potential environmental concern (CPEC) have been reported (Stewart et al. 2009; Stewart 2013).

Every few years, the monitoring data and programme operation have been reviewed to assess longer-term patterns and trends in contamination, and to assess operational changes that may be required to improve programme performance. Spatial patterns and temporal trends in sediment contaminants (metals and polycyclic aromatic hydrocarbons; PAH) were most recently reported by Mills et al (2012), covering monitoring data collected between 1998 and 2010 (inclusive). Organochlorines – organochlorine pesticides (OCPs) and polychlorinated biphenyls (PCBs) – and emerging contaminants were reviewed in Mills (2014a) and Mills (2014c). Operational aspects were last reviewed in 2002 (Timperley and Mathieson 2002), covering the initial three-year phase of the SoE programme, from 1998 to 2001.

1.2 Purpose of this report

This document provides a review of data quality and operational aspects of the RSCMP, updating the review undertaken by Timperley and Matheson (2002).

The focus is on assessing the reliability (principally the variability but, where possible, accuracy as well) of the monitoring data, and consequent implications for meeting the primary monitoring objectives, namely measurement of spatial patterns and temporal trends in sediment contamination.

Based on this assessment, areas with the potential to improve programme performance are identified.

1.3 Approach

As will become apparent to the reader of this review, assessment of data quality in the various sediment contaminant monitoring programmes undertaken by AC has not been straightforward. A range of information strands have been brought together in an attempt to provide a useful picture of data quality, and associated implications for data use.

Quality assurance data reported in the SoE, RDP, and UWH programmes have been compiled and analysed to assess the accuracy and variability of the sediment contaminant and particle size distribution (PSD) data.

Having a quantitative understanding of these factors is important, so that data users can better interpret the real world significance of spatial and temporal differences in contaminant concentrations detected by the monitoring. For example, would a difference between two values of 10% be considered “real” (i.e. due to some change in environmental factors), or could it be that this is simply due to analytical variability?

Data variability also influences the power of the monitoring programmes to detect trends over time – with more variable data, a greater number of samples (or greater time period) are required to detect trends of a given magnitude.

To address these aspects, the following assessments have been undertaken:

- Examine the accuracy and variability of the sediment contaminant and PSD data to: identify potential data quality issues, assess implications of data variability for trend assessment and provide a basis for defining data acceptance criteria. Assessment included analysis of:
    - Overall variability in metals and PAH data, and comparisons of the SoE, RDP, and UWH programmes;
    - Agreement between blind duplicate results for metals, both within-batch (WB), and between-batch (BB);
    - Accuracy, variability, and trends over time in Certified Reference Material (CRM) results (for metals, and to a lesser degree, PAH);
    - Particle size distribution (PSD) quality assurance data from the SoE programme;
    - Variability in Bulk Reference Sediment (BRS)[1] results, for metals and PSD.
- A comparison of the variability in Auckland Council monitoring programme data with that obtained in similar NZ and overseas monitoring programmes to assess whether the Auckland Council data have comparable data quality and would meet typical international data quality criteria.
- Document the results from reanalysis of archived samples, to check the validity of metals results obtained in the former SoE monitoring programme between 1998 and 2007. This assessment attempts to resolve whether some unusual trend patterns observed in an

---

[1] Bulk Reference Sediments (BRS) are “in house” reference samples used to track analytical performance between batches of monitoring samples. Samples are analysed with each set of monitoring samples to check whether consistent results are obtained over time.

interim assessment of metals monitoring data conducted in 2008 were real, or whether analytical variations were more likely to be responsible. The results from the reanalysis provide valuable information on the potential impact of analytical variation on trend records, and have implications for how the historical monitoring data should be used for trend assessment;

- Compare the results obtained for the determination of sediment mud content using two PSD analysis methods – laser analysis and wet sieving/pipette followed by drying to constant weight and weighing – employed in the sediment contaminant and benthic ecology programmes. Sediment muddiness is an important environmental variable influencing both sediment chemistry and benthic ecology. Because two analysis methods have been used in the past, the comparability of results obtained by these methods needs to be known so that the effects of changes in PSD analysis method on trends can be assessed. The need for a robust single method to assess mud content was identified in 2008, as part of a process of integrating sediment contaminant and benthic ecology monitoring into the unified regional programme (RSCMP);
- Provide commentary on sampling methods, including potential problems and comparisons of sediment contaminant results obtained from the protocols used in the SoE, RDP, and UWH programmes;

Based on these assessments, recommendations for future monitoring and investigation are made. Commentary on sites, sampling approach, analysis methods, quality assurance protocols and contaminants measured have also been made.

1.4 Data reviewed

The focus of this review is on the primary stormwater contaminants monitored in the AC sediment contaminant monitoring programmes, as defined in the monitoring "blueprint" document (ARC 2004). This is because these chemicals are the main indicators used to assess contaminant status (and potential effects on benthic ecological health), and to monitor trends over time.

These primary contaminants are:

- Metals – copper (Cu), lead (Pb) and zinc (Zn), in both extractable (2 M HCl, <63 µm fraction) and total recoverable (HNO_3/HCl digestion, <500 µm fraction) forms; and
- Polycyclic aromatic hydrocarbons (PAH), which are analysed in the <500 µm fraction.

A comprehensive assessment of the status and trends in these primary contaminants between 1998 and 2010 has been reported by Mills et al. (2012). This report provided detail on the monitoring conducted over this period. A summary of the sampling and analyses conducted in the SoE, RDP, UWH, and RSCMP programmes is provided in Table 1-1 (taken from Mills et al. 2012 and updated to include the 2011/12 monitoring undertaken in the RSCMP).

In addition to this core set of monitoring data, the following information has been included in the data quality review:

- Metals data from the RSCMP and UWH programmes collected in 2011. Metals data from the 2012 and 2013 monitoring had not been finalised at the time of this review, and has therefore not been included;
- Metals data from repeat analyses of archived samples from five sites in the former SoE programme, sampled in 1998, 1999, 2001, 2005, and 2007; and
- Particle size distribution (PSD) QA data from the SoE and RDP programmes sampled between 1998 and 2010, and from Bulk Reference Sediment (BRS) analyses conducted in 2011 and 2012, the latter being the QA samples for the 2012 RSCMP monitoring.

A much smaller amount of data has been collected for secondary contaminants – organochlorine pesticides (OCPs) and polychlorinated biphenyls (PCBs) – and hence a detailed analysis of quality assurance data has not been undertaken. Organochlorine monitoring data have recently been reviewed (Mills 2014a), and some recommendations from that review have been included in this report.

Table 1-1 A summary of sampling and analysis providers and methods used in the SoE, RDP, UWH, and RSCMP between 1998 and 2012.

			Metals Analysis				
Programme	Years	Sampling[a]	Sieving[b]	Digestion[b]	Analysis[c]	PAH[b]	PSD[d]
SoE	1998-2001	ARC	NIWA-H	NIWA-H	NIWA: AAS	NIWA-H	NIWA: Laser
	2003–2007	ARC	NIWA-A	NIWA-A	RJ Hill: ICP-MS	NIWA-H	NIWA: Laser
	2009	DSL	RJ Hill	RJ Hill	RJ Hill: ICP-MS	n/a	NIWA: Sieve
RDP	2002	ARC and KML	RJ Hill	RJ Hill	RJ Hill: ICP-MS	NIWA-H	RJ Hill: Sieve
	2004–2008	DSL	RJ Hill	RJ Hill	RJ Hill: ICP-MS	n/a	NIWA: Laser
	2009–2010	DSL	RJ Hill	RJ Hill	RJ Hill: ICP-MS	n/a	NIWA: Sieve
UWH	2005–2007	NIWA	NIWA-H	RJ Hill	RJ Hill: ICP-MS	RJ Hill	NIWA: Sieve
	2008–2011	NIWA	RJ Hill	RJ Hill	RJ Hill: ICP-MS	RJ Hill	NIWA: Sieve
RSCMP	2011–2012	DSL	RJ Hill	RJ Hill	RJ Hill: ICP-MS	n/a	NIWA: Sieve

a. Sampling providers: Auckland Regional Council (ARC), Kingett Mitchell Ltd (KML), Diffuse Sources Ltd (DSL).

b. Analytical providers: NIWA Hamilton (NIWA-H), NIWA Auckland (NIWA-A), RJ Hill Laboratories.

c. Metals analysis methods: Atomic Absorption Spectroscopy (AAS), Inductively Coupled Plasma Mass Spectrometry (ICP-MS)

d. PSD analysis: Particle Size Distribution analysis methods – Laser particle size analyser (Galai instrument at NIWA); Sieve – wet sieving and pipette analysis into 6 size fractions (NIWA), or wet sieving into 3 size fractions (RJ Hill), followed by weighing.

.

1.5 Quality assurance data

The quality assurance (QA) data available for review varied between programmes and contaminants. A summary of the QA data used is given below. While not necessarily exhaustive (there are other smaller pieces of QA information that haven't been included), this set of QA data provides a comprehensive characterisation of data quality. Addition of smaller amounts of additional QA information is unlikely to make significant changes to the conclusions.

A summary of the QA data used in the review is given in Table 1-2 (appended to this section).

1.5.1 State of the environment programme

Quality assurance for metals and PSD was based on analysis of archived samples taken from five monitoring sites. A single sub-sample of the original archived sediment from each site was reanalysed with each subsequent batch of monitoring samples, providing a record of data consistency over time. Details are given in SoE programme monitoring reports for 1998 to 2007 (Williamson et al. 1999, Mills et al. 2000, Mathieson et al. 2002, Reed and Webster 2004, McHugh and Reed 2006, Reed and Gadd 2009).

Insufficient amounts of the original archived samples were available for repeat analysis in all the subsequent monitoring years, requiring a change to samples taken from later years. Exactly which year's samples were used after 2003 was unclear from the monitoring reports. In addition, only a single sample was reanalysed on each monitoring occasion. This lack of replicated analysis prevented testing the significance of differences between batches. These factors made interpreting the results from the SoE programme QA record difficult.

Because of the ambiguity associated with the SoE QA results for metals, they have not been presented or analysed in any quantitative detail in this review. For illustrative purposes, the PSD QA data from the SoE programme have been presented in section 4.4.1. The SoE monitoring report references listed above should be consulted for the original results.

Quality assurance for PAH was based on analysis of replicate samples taken from the same monitoring batch (within batch (WB) replicates) and, for 1999, 2001, and 2005, from samples archived from previous years (between batch (BB) replicates). Data for Certified Reference Material (CRM) were reported for monitoring conducted in 1998 and 2005. These QA data were more straightforward to interpret than those for metals, and therefore most have been included in this review. However, there was some ambiguity in the 2005 QA data reported by McHugh and Reed (2006) – the archived QA sample IDs did not match those given in the earlier monitoring report, and composite samples of archived sample were reanalysed (rather than individual samples), preventing direct comparisons between batches.

The QA data for PAH were therefore used where possible, but ambiguous data were not included in this review. The WB and BB replicate data enabled an assessment of agreement between replicates (a useful measure of variability), as well as accuracy (from CRM analyses). The results from multiple analyses (N=3) of individual samples were also reported, enabling an assessment of WB variability of analysis to be made.

1.5.2 Regional discharges project programme

The QA protocols for metals analysis in the RDP programme were reasonably consistent over time. Generally, QA involved analysis of WB and BB replicates (usually 5–6 of each per batch of monitoring samples). Multiple CRM samples were also analysed (3–5 samples per batch). In addition, the QA/QC reports provided by R.J. Hill laboratories contained data from their own QA samples – these included sample spike recoveries, blanks, additional CRM data, and duplicate analyses. Details have been given in RDP annual monitoring reports (e.g. Diffuse Sources 2005–2012).

No QA was undertaken for PSD in the RDP. However, at sites where both sediment contaminants and benthic ecology were being assessed, two sets of PSD samples were taken – one set for use with sediment contaminants, and one set for benthic ecology. While these PSD samples were not identical[2], they provided the basis for two useful comparisons:

- For samples collected until 2008, these samples enabled comparisons of laser PSD analysis (done on the contaminant sample set) and the wet sieve/pipette/weighing method (done on the ecology samples).
- For sampling conducted in 2009 and 2010, the contaminant and benthic ecology PSD samples were both analysed by the wet sieve/pipette/weighing method. This provided a useful set of duplicate analyses for this PSD analysis method.

These PSD QA results are described in sections 4.3 and 4.4.2.

1.5.3 Upper Waitemata Harbour benthic ecology programme

Sediment contaminants have been measured at UWH programme sites since 2005. The data collected up to 2010 have been analysed and reported by Mills et al. (2012).

Quality assurance data from the UWH programme included in this review were WB duplicates for metals from 2009 sampling (five pairs of duplicates, analysed in 2010), and one pair of WB duplicates in 2011. UWH programme samples from 2011 were analysed with the RSCMP 2011 sample batch, so the same QA results apply (reported in Diffuse Sources 2012).

Additional UWH programme QA data exist in NIWA files for the 2005–2007 period, but to our knowledge they have not been formally reported.

---

[2] Chemistry PSD samples were composites (1 per site) prepared from the chemistry samples collected along the two “sampling lanes”, as detailed in the monitoring blueprint (ARC 2004). The ecology PSD samples were composites (1 per site) prepared from sub-samples taken from each ecology core location (10 per site). The spatial distribution of sub-samples was therefore different for the two types of PSD samples, but both provided a reasonably robust spatial average at each site. The comparability of the results obtained from the 2009/10 samples (both sets analysed by the same sieve/pipette/weigh method) supports this (see section 4.4.2).

1.5.4 Rural estuaries ecology programmes

Benthic ecology monitoring programmes in the Kaipara and Mahurangi Harbours, and Whangateau, Wairoa, Waiwera, Puhoi, Mangemangeroa, Waikopua, Okura, Turanga, and Orewa estuaries (“7 estuaries”) sampled in 2009/10 included analysis of sediment contaminants. These programmes are not currently part of the on-going sediment contaminant monitoring programme, but metals analyses were conducted using the same methods used in the RDP/SoE programmes at that time. The analytical QA data (WB duplicates) from these analyses were therefore included in this review. PAH QA data from these surveys have been used, because they provide a limited, but useful, contribution to the much smaller body of PAH QA information.

Data for these sites can be found in reports appearing on the Auckland Council website – to date these include: Hailes et al. (2010) for the Kaipara Harbour; Townsend et al. (2010) for the Whangateau Harbour; Halliday and Cummings (2012) for the Mahurangi Estuary; Hewitt and Simpson (2012) for Waiwera, Puhoi, Mangemangeroa, Waikopua, Okura, Turanga, and Orewa estuaries and Lohrer et al. (2012) for the Wairoa embayment. The organochlorine data have been reviewed in Mills (2014a).

1.5.5 Regional sediment contaminant monitoring programme 2011 and 2012

Quality assurance protocols for the RSCMP 2011 and 2012 monitoring followed those used in the RDP (2002 to 2010) and SoE (2009). WB duplicates, CRMs, spike recoveries, and lab blanks were assessed.

In addition, two Bulk Reference Sediments (BRS) were included. Initial characterisation of the BRS was undertaken in 2011, and the first set of repeat analyses performed with the 2012 RSCMP sample batch. At the time of preparing this review, the PSD data from both 2011 and 2012 were available (and have been summarised in section 4.4.3). The 2012 metals data were not yet finalised, and have therefore not been included.

Unlike the CRM samples analysed to date, the BRS are subjected to the entire analysis procedure, including sieving (<63 µm and <500 µm) and drying. The data obtained from the BRS therefore provides a more comprehensive measure of analytical performance over time for “real” samples than the CRMs.

Table 1-2 A summary of key quality assurance data for metals, PAH, and PSD.

		Analytes			
Programme	Years	Metals	PAH	PSD	Comments
SoE	1998-2007	[BB]	WB, BB, CRM	BB	BB metals data unclear, therefore not included in review
	2009	WB, BB	n/a	see text	RDP & SoE sites sampled in 2009 analysed together. BB duplicates were RDP site samples archived from 2007 & 2008.
RDP	2002	[WB], [BB], CRM			Duplicate data appendix not available for review.
	2004–2008	WB, BB, CRM	n/a	see text	No PSD QA data, but comparison of PSD methods made.
	2009–2010	WB, BB, CRM	n/a	see text	No PSD QA data, but sampling/analysis comparisons made.
UWH	2005–2007	[WB], [BB]			Metals & PAH QA data exist in NIWA files but unreported.
	2008–2011	WB, BB, CRM			Analyses included in RDP/SoE batch in 2009.
RSCMP	2011–2012	WB, BB, CRM, [BRS]	n/a	BRS	First BRS analysis undertaken in 2011. BRS PSD data from 2011 & 2012 analyses used. Metals QA data for 2011 used in review. 2012 metals data not finalised, not used.
Rural estuaries	2009, 2010	WB, [CRM]	CRM		Metals CRM data not included in review. This aspect well covered by data from RDP/SoE/RSCMP 2002-2011 data series.

Table notes:

[ ] QA data exist but are not used in review – see section 1.5 for further explanation

CRM – Certified Reference Material

WB – Within Batch duplicates

BB – Between Batch duplicates

BRS – Bulk Reference Sediments

n/a – not analysed

Blank cells indicate no known information

2.0 Accuracy and variability in metals and PAH data

2.1 Overview

This section provides a summary of the accuracy and variability in the metals and PAH data. The variability observed in the monitoring data record is the result of a combination of factors including:

- Environmental factors, including changes in contaminant inputs (e.g. storm events), site disturbance due to natural and human activities, sediment erosion/deposition etc.
- Sampling procedures
- Sample analysis, including sample processing, sub-sampling, and instrumental analysis steps; and
- Human error (e.g. mistakes in sample labelling or data analysis).

As summarised in Table 1-1, the sediment contaminant data has been acquired by a number of providers using procedures that have varied over time and between programmes. No programme has had completely consistent sampling/sample processing/analysis over the whole duration of monitoring.

Probably the most consistently acquired data are from the RDP programme, which has used the same sampling and analysis procedures/providers since 2004 (in 2002 the sampling was conducted by a different provider, but sampling procedures and analysis were the same as in later years).

The UWH programme has had consistent sampling method and provider throughout (2005–2011), but sample processing changed between the 2005-7 and 2008-11 periods. Analysis method/provider remained consistent over the whole monitoring period.

There have been several changes in the SoE programme over time (see Table 1-1), partly reflecting its longer duration (1998–2009/11). In 2009 (and in 2011 as part of the RSCMP), the SoE programme sites were sampled and analysed using the same procedures and providers as the RDP programme sites.

The differences between these programmes have therefore reduced over time. The (former) RDP and SoE programme sites are now (since 2009) sampled and analysed in the same way. RDP sites sampled in the odd years (2009, 2011) have also been analysed in the same batch as SoE site samples, and this is likely to continue in future.

The key difference between the programmes is now the sampling methods used in the UWH and RDP/SoE (now RSCMP) programmes. The UWH programme method is based on a combination of cores taken across each site (Lundquist et al. 2010), whereas the (former) RDP/SoE programme sampling uses the sampling lane “blueprint” method (ARC 2004).

Because of the multiple sources of variability, including changes in sampling and analysis procedures/providers in most programmes over time, analysis of data variability is therefore best done in a fairly general way, with the aim of characterising the overall variability in the data rather than trying to definitively investigate particular sources or causative factors. Comparing variability between programmes should also be treated in the same way, because the operational distinctions between them have blurred over time.

With these provisos, the following sections summarise data variability and provide commentary on the implications for trend detection and programme operation (e.g. sampling, processing, and analysis).

The key issue for the RSCMP associated with data variability is its effect on the ability of monitoring to detect trends. We can use the measures of data variability to estimate the monitoring effort (e.g. sample numbers, length of time, sampling frequency) required for detection of statistically significant trends.

These measures are also useful to examine whether there have been any marked differences in variability between the monitoring programmes (SoE, RDP, UWH). Similar variability across the programmes would provide reassurance about the comparability of the monitoring data acquired in these programmes to date. Conversely, if there are major differences evident between the programmes, further investigation to determine contributing factors may be warranted (e.g. sampling or analytical procedures that may produce lower/higher variability). This could assist in programme improvement.

Comparison with other NZ and overseas monitoring programmes is also useful, allowing the assessment of whether the Auckland Council programme data have markedly different performance characteristics (i.e. data quality) to similar programmes elsewhere.

2.2 Accuracy and variability assessments undertaken

Monitoring and QA data from the former RDP and SoE programmes, and from the UWH programme, were analysed to:

- Calculate overall data variability in metals and PAH from SoE, RDP, and UWH monitoring programmes (using the data from all three programmes grouped).
- Compare variability between the SoE, RDP, and UWH programmes.
- Assess whether variability has changed markedly over time.
- Compare data in the UWH programme from 2005–7 with that from 2008–11, to assess whether marked changes in variability occurred as a result of changes in sample preparation methods and providers (see section 2.4).
- Comment on implications of data variability for trend detection in relation to monitoring effort.
- Assess accuracy and variability of laboratory analysis for metals, as indicated by CRM results.
- Examine agreement between blind duplicates (WB and BB). Compare the results with data quality acceptance criteria from overseas (USA) monitoring programmes to assess whether AC monitoring data would be consistent with an international definition of acceptable quality.

- Compare laboratory analysis within-batch (Lab WB) duplicate results for metals with blind WB duplicates to assess the contribution of instrumental analysis to overall variability.

Variability has been expressed as a coefficient of variation (CV, expressed as a %[3]). Agreement between duplicates was assessed by calculating the relative percentage difference (RPD) between two results (e.g. duplicates, A and B):

$$RPD = \left[\frac{|A - B|}{(A + B)/2}\right] x100\ \%$$

2.3 Variability in monitoring results

The within-year variability, expressed as a coefficient of variation (CV), of metals and PAH at each of the 92 monitoring sites was calculated from the analytical results from sample replicates (generally N=3 replicates per site were analysed on each monitoring occasion). These within-year CVs for each site have been used to avoid incorporating any contribution to variability arising from trends over time. The assessments given in the following sections use these within-year CV values at each site as the basis for statistical summaries of data variability and for comparing variability between programmes and over time (i.e. the variability in the variability data, how this differs between programmes, and how it has changed over time).

2.3.1 Overall variability in metals and PAH

All SoE, RDP, and UWH data were grouped, and overall variability assessed. A summary is presented in Table 2-1 and the data plotted in Figure 2-1.

---

[3] The coefficient of variation (CV) is the ratio of the standard deviation to the sample mean. In this review, it has been expressed as a percentage (100 x standard deviation/mean).

Table 2-1 A summary of overall sediment contaminant variability given as within-year coefficients of variation; CV, expressed as a %). Data are from all sites in SoE, RDP, and UWH programmes, covering data collected from 1998–2011 (inclusive). N is the number of within-year CV values used to generate the statistics.

	Extractable metals (<63 um)			Total Recoverable metals (<500 um)			
CV (%)	Cu	Pb	Zn	Cu	Pb	Zn	HWPAH
N	431	428	431	287	287	287	181
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.57
Maximum	111.0	127.6	46.8	59.5	52.1	49.6	102.1
Median	5.1	4.8	4.5	6.2	5.5	4.3	10.7
Mean	6.9	6.9	5.8	9.1	7.5	6.1	17.7
10% ile	1.4	1.7	1.3	1.2	1.5	1.0	2.8
90% ile	13.0	12.9	11.6	21.6	15.5	11.9	46.3

Median within-year variability values (CVs) were 4–6% for metals and about 10% for HWPAH. Slightly higher within year variability was observed for total Cu, and slightly lower for total Zn.

Within year variability was occasionally high, as shown by maximum CVs of 47–128%. This reflects the occasional outlier result in the 3 replicates analysed at each site each year. 90% of the CV values were in the <10–13% range for extractable metals, <12–22% for total metals, and <46% for HWPAH.

Overall, the median within-year variability (as a CV, %) in metals has typically been of the order of 5%, and for PAH around 10%. The sometimes much higher variability observed was often associated with a single outlier result in the 3 replicates analysed per site on each sampling occasion. Identifying these unusual results and conducting further analysis is worthwhile to reduce data variability. This approach, involving analysis of spare replicate samples[4], is a standard procedure for the “blueprint” monitoring protocol.

---

[4] Five replicates are taken from each site in the “blueprint” protocol. Reps 1–3 are analysed, and Reps 4 and 5 are archived in case further analysis is required.

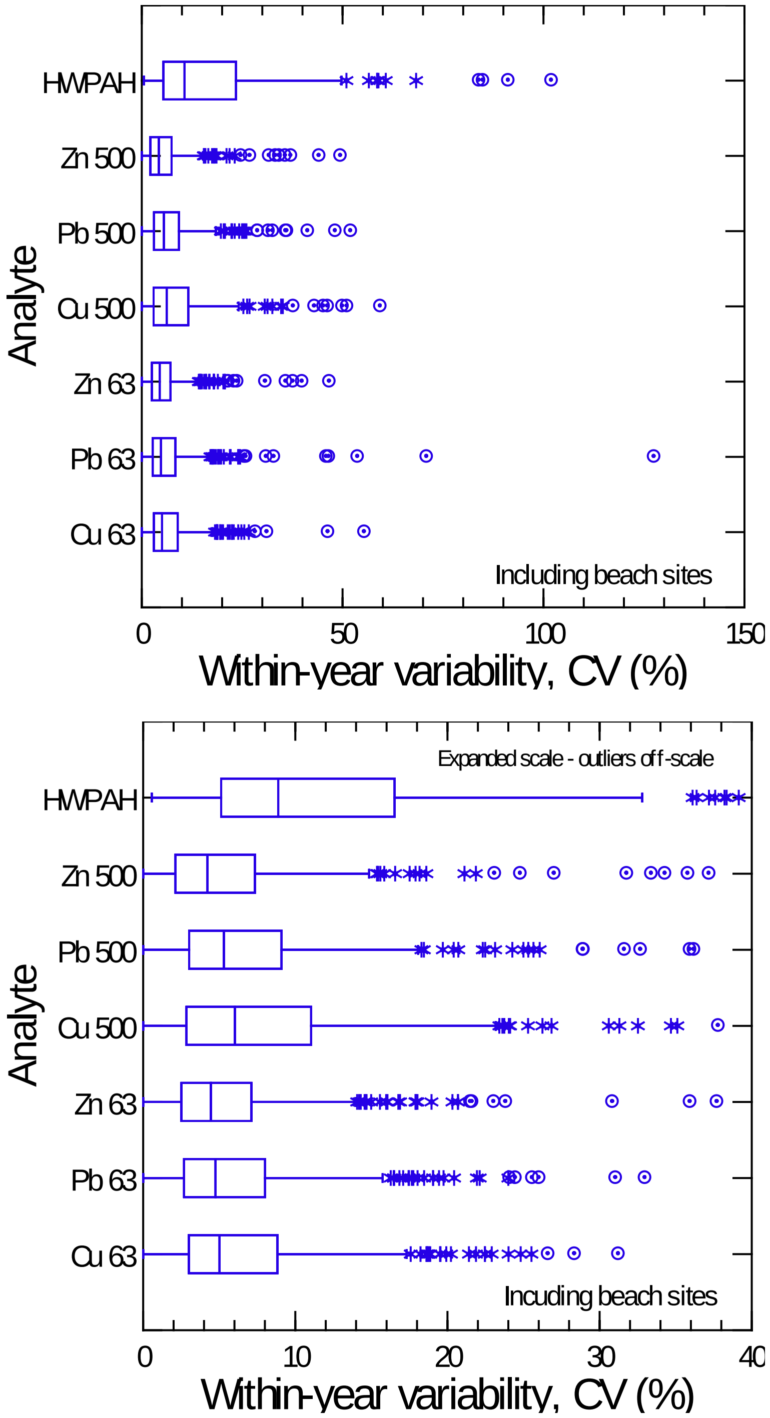


Figure 2-1. Within-year variability for metals and PAH results.

Data are for extractable Cu, Pb, and Zn in the <63 um fraction (Cu63, Pb63, Zn63), total recoverable Cu, Pb, and Zn in the <500 um fraction (Cu500, Pb500, Zn500) and high molecular weight PAH in the <500 um fraction (HWPAH). The lower plot has an expanded scale for better visual comparisons of medians (high outliers are off scale).

2.3.2 Variability relationships between analytes

Within-year variability was correlated between metals in the same forms – i.e. in each of extractable and total recoverable forms – but not between metals in different forms (i.e. the same metal in extractable and total recoverable forms). This is probably not unexpected, as the extractable and total recoverable metals analyses are distinctly different, being undertaken on samples processed by different digestion methods (2 M HCl and HNO_3/HCl) and on different sediment fractions (<63 µm and <500 µm).

Examples of these relationships for Zn and Cu within each of the extractable and total recoverable metal forms are shown in Figure 2-2, and between the extractable and total recoverable forms in Figure 2-3.

Generally, within each metal form, if high variability (i.e. a high within-year CV) was observed for one metal, it was also high for the other two. This suggests that high within-year variability was related to generic "whole sample" factors (possibly including analysis), rather than being specific to each individual metal.

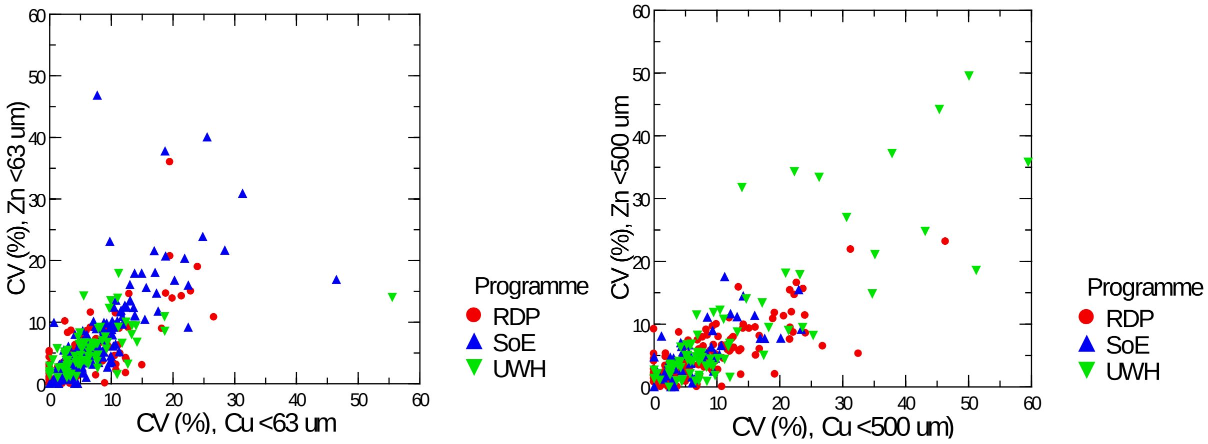


Figure 2-2. Examples of within year variability relationships between metals in the same forms: extractable Cu and Zn, and total recoverable Cu and Zn.

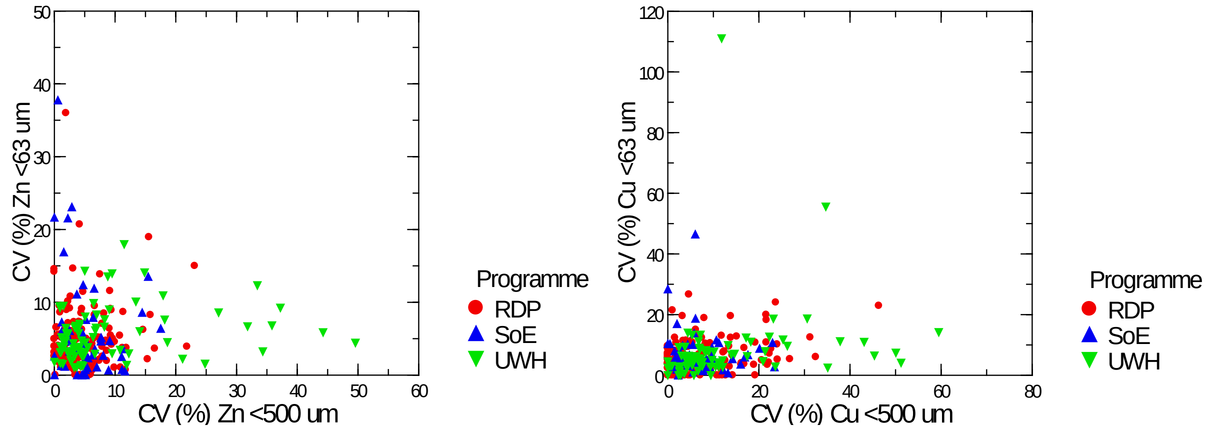


Figure 2-3. Examples of within year variability relationships between the different metals forms: extractable Cu and Zn and total recoverable Cu and Zn.

2.3.3 Relationships between variability and analyte concentrations

Variability declined slightly as metals concentrations increased, but relationships were very weak with lots of scatter (Figure 2-4). Highest variability was often found where metals concentrations were low, but not always. The relationships between the within-year variability and metals concentrations varied between the SoE, RDP, and UWH monitoring programmes. Most notably, total recoverable metals from the UWH programme often showed higher variability at low metals concentrations than in the other programmes, especially for Zn.

For PAH, highest variability was observed at sites with lower concentrations, but, as for metals, the variability versus concentration relationship showed a great deal of scatter (Figure 2-5).

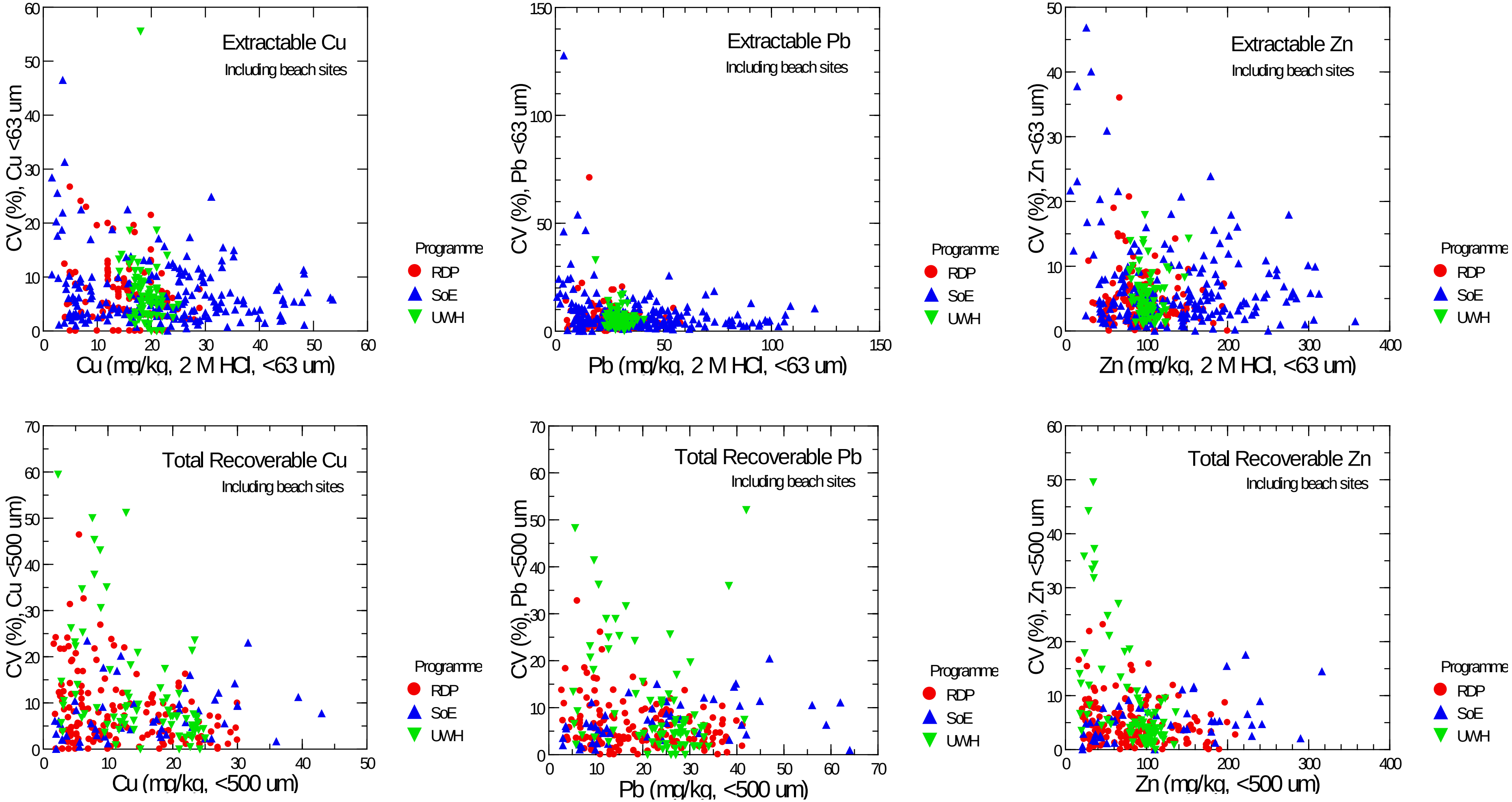


Figure 2-4. Relationships between the within-year variability for extractable and total recoverable metals (as CVs, %) and metals concentrations.

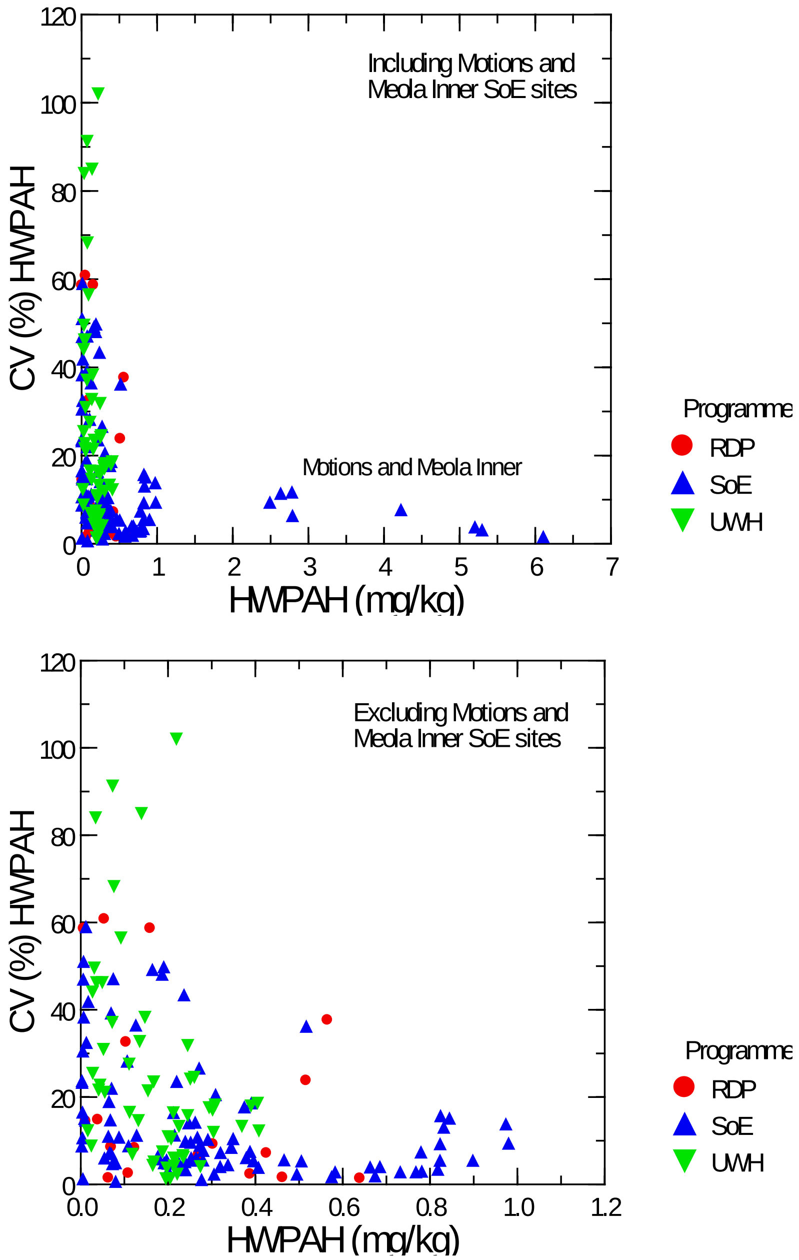


Figure 2-5. Relationships between the HWPAH within-year variability and HWPAH concentrations. Within year variability data are CVs (%). The lower plot has an expanded scale to exclude the high concentrations present at Motions and Meola Inner sites.

2.3.4 Relationships between variability and sediment texture

Sediment texture is a key characteristic that may affect the variability in sediment chemistry results. As described later in this report (in section 4.0), different particle size distribution (PSD) analysis methods have been used at various times in the SoE, RDP, and UWH monitoring programmes – for the 1998–2008 period, samples from the SoE and RDP programmes were analysed by laser analysis, and from 2009 onwards the pipette/sieve/weigh method was used. All samples in UWH programme, from 2005–2011, were analysed by sieve/pipette. The use of the different PSD methods over time and between the programmes complicates detailed interpretation of relationships between sediment texture and analyte variability. The relationships have therefore been shown graphically in two ways:

1. Using all the data from 1998–2011 (Figure 2-6). This provides the most comprehensive picture of texture/variability relationships, but may include added uncertainties associated with the use of different PSD analysis methods, and
2. For 2009–2011 only, when the same PSD analysis method was used for all samples (Figure 2-7). This provides a much smaller data set, but gives better comparability between programmes.

Examination of these plots indicated that there was no strong relationship between variability and mud content for extractable metals, although highest variability tended to occur for sediments with low mud content (e.g. sandy beach samples). This was also observed for total recoverable metals, with samples from the UWH programme showing the greatest variability in low mud content samples (this is particularly evident for the full data set shown in Figure 2-6).

HWPAH data tend to be slightly more variable at lower mud content (sandier) sites (Figure 2-8).

It should be noted that sediment texture and contaminant concentrations may be related, such that coarser textured (i.e. sandier) sediments generally have lower total contaminant concentrations than muddy sediments from a similar location. Therefore the increasing variability in analyte concentrations at lower sediment mud content may also be (at least in part) attributable to lower analyte concentrations in less muddy samples.

Examination of plots of metals concentrations versus sediment mud content revealed that extractable metals concentrations (in the <63 µm fraction) showed no strong correlation with mud content. Total recoverable metals showed a general, but highly variable, increase in concentration with increasing mud content. The relationships showed considerable scatter, indicating mud content was generally responsible for only a small portion of the variation in metals concentrations – this is sensible, since site location in relation to contamination sources is a key influencing factor. Examples of the relationships between metals concentrations and sediment mud content, for Zn, are shown in Figure 2-9).

HWPAH concentrations showed no correlation between concentrations and mud content. Highest concentrations found at the relatively sandy Motions SoE site.

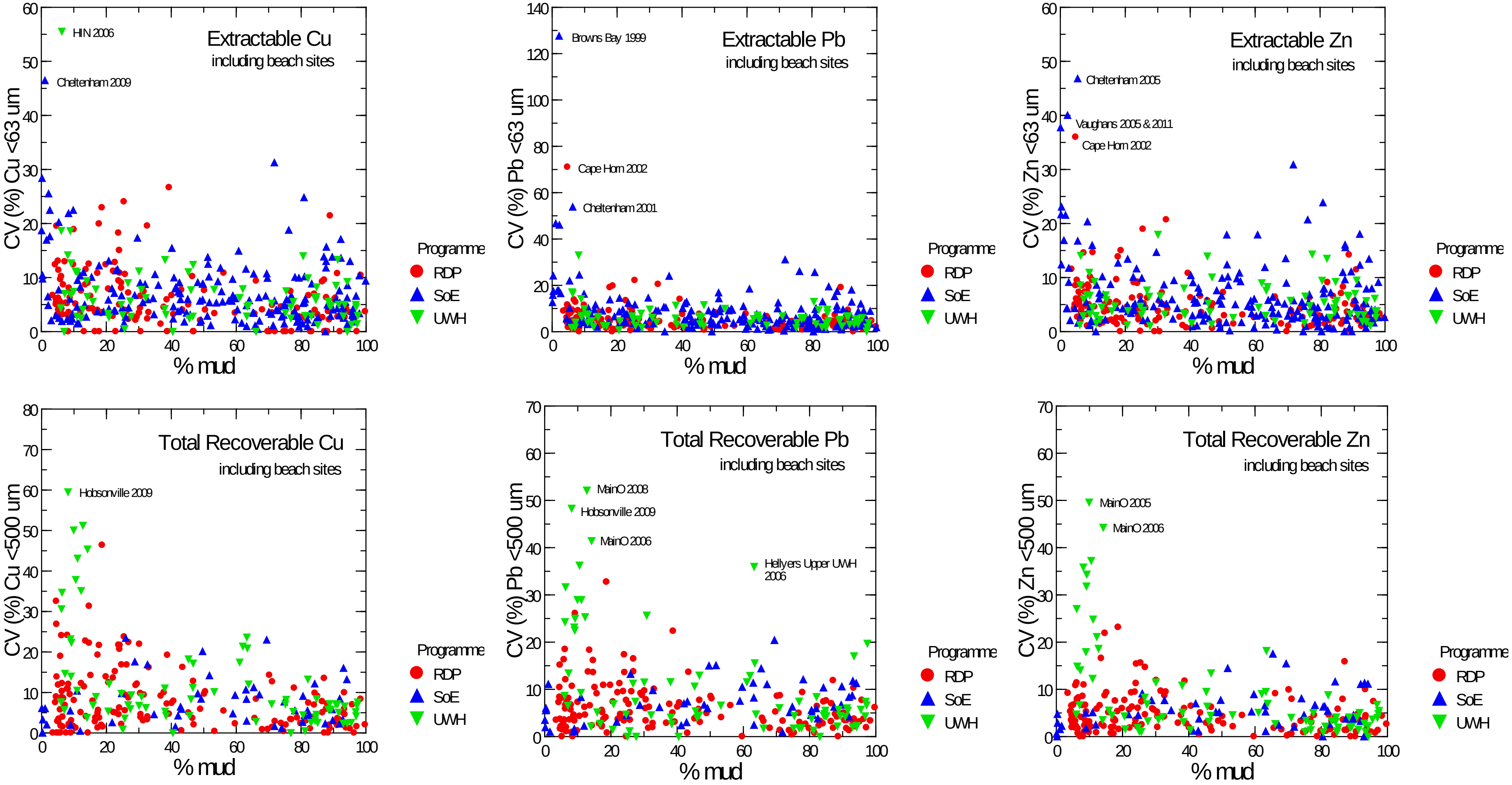


Figure 2-6. Relationships between the within-year variability for extractable and total recoverable metals (as CVs, %) and sediment mud content (% <63 µm) for all data 1998–2011.

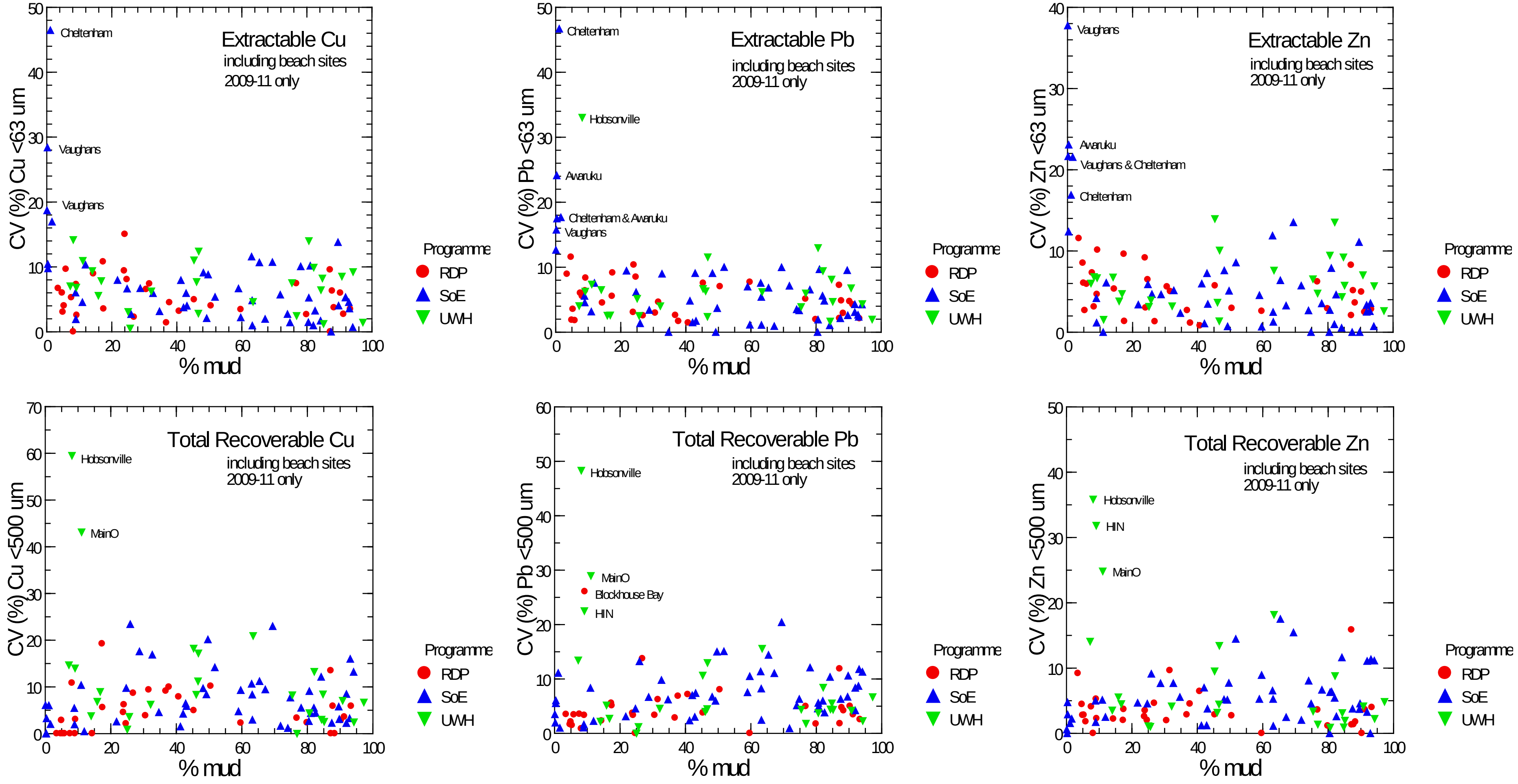


Figure 2-7. Relationships between the within-year variability for extractable and total recoverable metals (as CVs, %) and sediment mud content (% <63 µm) for data acquired from 2009–2011 only.

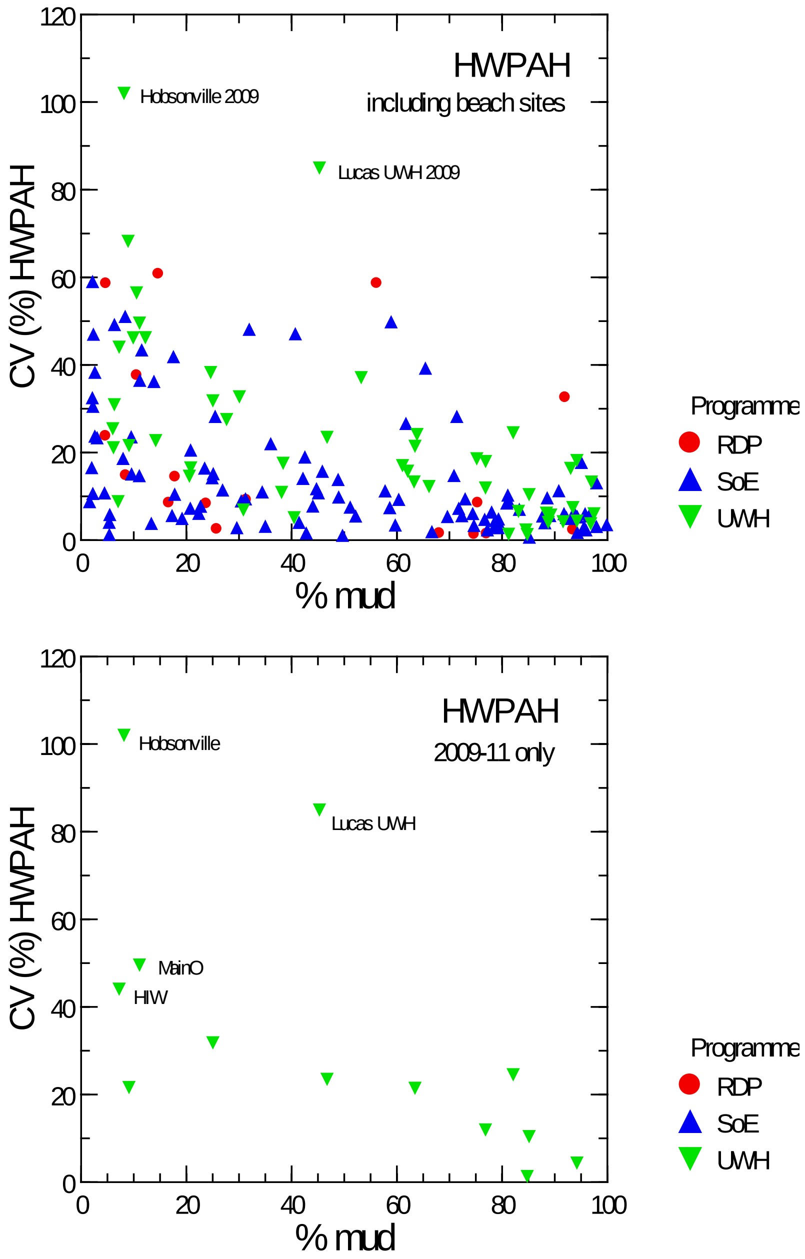


Figure 2-8. Relationships between the within-year variability (as CVs, %) for HWPAH and sediment mud content (% <63 µm). The upper plot includes all the data from 1998–2011. The lower plot is for 2009–2011 only.

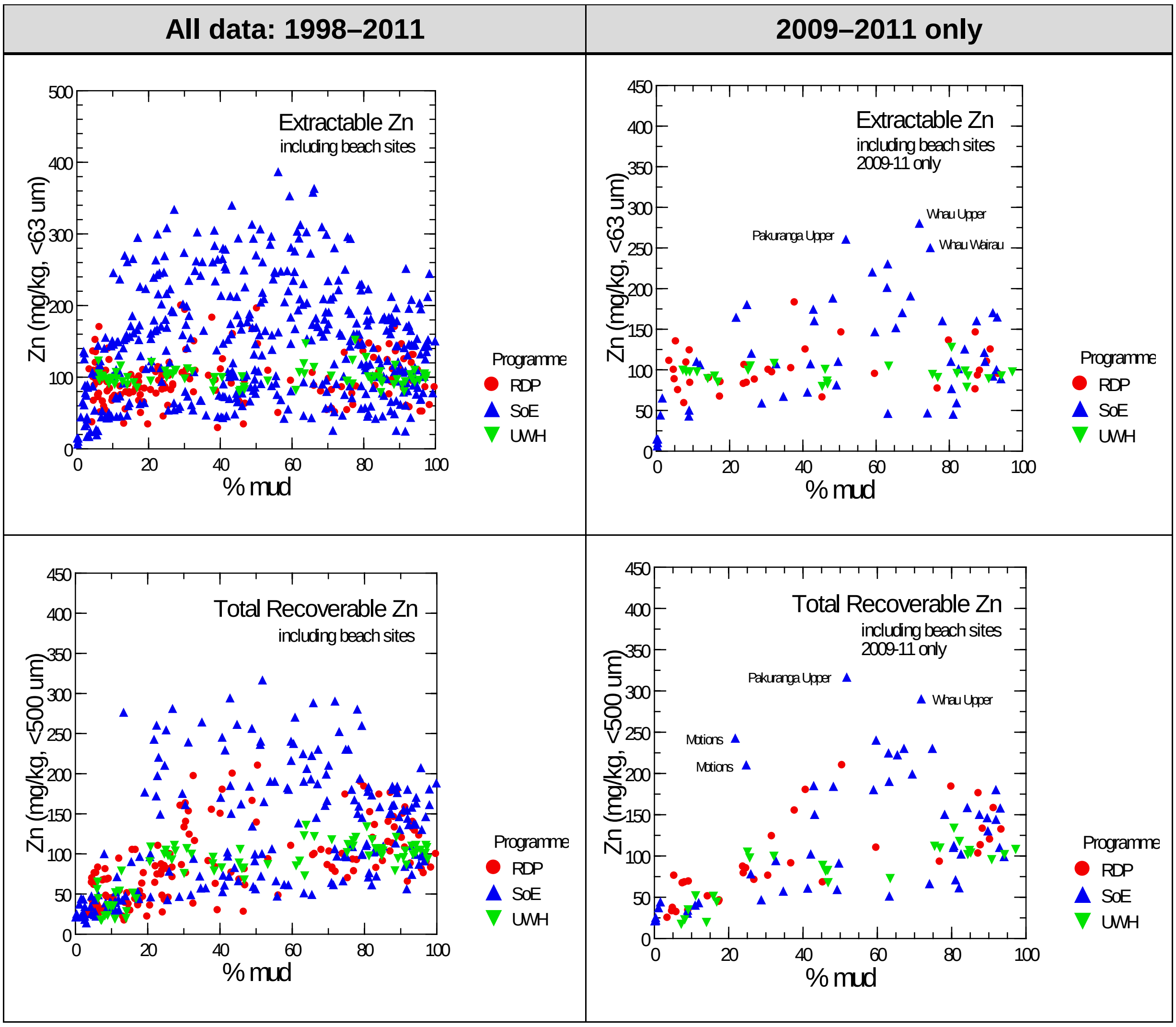


Figure 2-9. Examples of the relationships between metals concentrations and sediment texture (mud content, % <63 µm): Extractable and total recoverable Zn.
The left hand plots include all data from 1998–2011, and the right hand plots are for 2009–2011 only.

2.3.5 Comparison of variability in State of the Environment, Regional Discharges Project, and Upper Waitemata Harbour programmes

Comparisons of variability between the programmes was undertaken to identify any marked differences that might highlight major issues, or lead to identification of key factors that could improve monitoring performance.

Two assessments have been made:

- Firstly, all the data from each programme were grouped (RDP sites covering 2002–2011, SoE sites 1998–2011, and UWH 2005–2011) and variability between the programmes compared. This provided a broad scale comparison, and includes the effects of all the different variables contributing to variability.

- Secondly, a sub-set of the data, with more consistent sample processing and analysis methods, was compared. Since 2008, all programme samples have been processed and analysed in the same way. Before this, there were variations (see Table 1-1). Comparing the three programmes using only the 2008–2011 data provided an assessment of differences arising from site-related factors, and for UWH the effect of the different sampling protocol.

A summary of the within year variability (CV, %) data using all the data from each programme is given in Table 2-2 and shown graphically in Figure 2-10 and Figure 2-11.

Note the smaller number of data (N) for total recoverable metals than extractable metals for the SoE programme. This is because total metals were measured on only a single sample from each site between 1998 and 2007. For 2009 and 2011, 3 replicates per site were analysed (as for extractable metals). The SoE total recoverable metals data therefore really only represent the 2009–2011 period.

Table 2-2 Summary of variability (given as within-year coefficient of variation; CV, expressed as a %) for extractable metals (<63 µm fraction), total recoverable metals (<500 µm), and high molecular weight PAH (<500 µm) at RDP, SoE, and UWH programme sites. N is the number of within-year CV values used to generate the statistics. Data are from 1998 to 2011.).

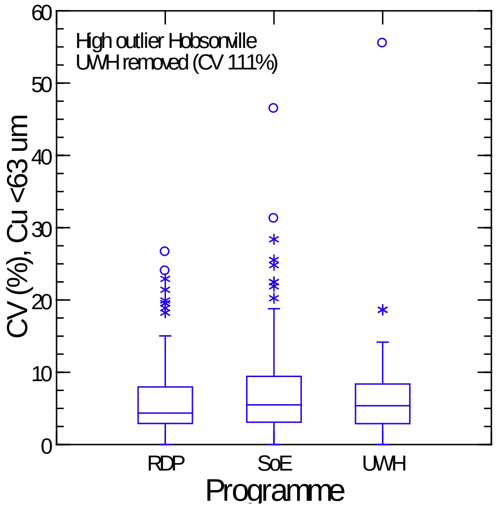
		Extractable metals (<63 um)			Total Recoverable metals (<500 um)			
Programme	CV (%)	Cu	Pb	Zn	Cu	Pb	Zn	HWPAH
RDP	N	155	154	155	155	155	155	19
	Maximum	26.6	71.0	36.0	46.4	32.7	23.1	60.7
	Median	4.3	3.9	3.9	5.6	4.6	3.6	8.5
	90% ile	12.4	9.4	9.2	19.2	12.0	9.6	58.6
SoE	N	195	193	195	51	51	51	107
	Maximum	46.5	127.6	46.8	23.4	20.4	17.5	58.9
	Median	5.5	5.4	4.7	5.8	6.3	5.0	8.7
	90% ile	13.8	16.6	13.5	16.4	12.5	11.2	37.8
UWH	N	81	81	81	81	81	81	55
	Maximum	111.0	33.0	18.0	59.5	52.1	49.6	102.1
	Median	5.5	5.2	4.7	7.2	5.7	4.6	17.7
	90% ile	12.9	10.5	9.9	32.2	27.0	25.7	56.5

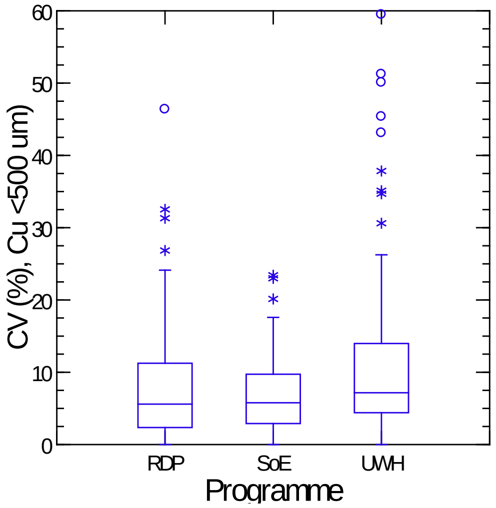
Generally, the median within-year variability in metals was similar in all programmes. 90%-ile variability was higher for extractable Pb and Zn in the SoE programme, in part reflecting higher variability data obtained at the coarse textured open coastal beach sites (Awaruku, Browns Bay, Cheltenham, and Vaughans). Total recoverable metals in the UWH programme had higher 90%-ile variability. UWH sites with highest total metals variability were sandier sites, including Outer Main Channel, Herald Island North, Waiarohia and Hobsonville.

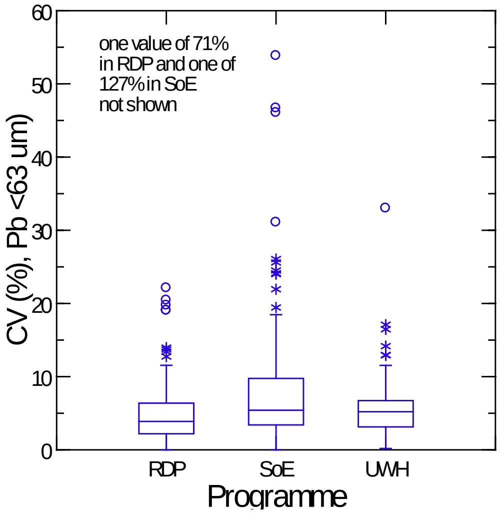
Variability in HWPAH was slightly higher at UWH programme sites (Figure 2-11). This may, in part, reflect differences in analysis procedures and method performance for the different labs used in these programmes (SoE and RDP samples were processed and analysed by NIWA, UWH samples processed by NIWA and analysed by R.J. Hill labs; see Table 1-1). As observed for total metals, sites with higher PAH variability were mostly the coarser textured sites.

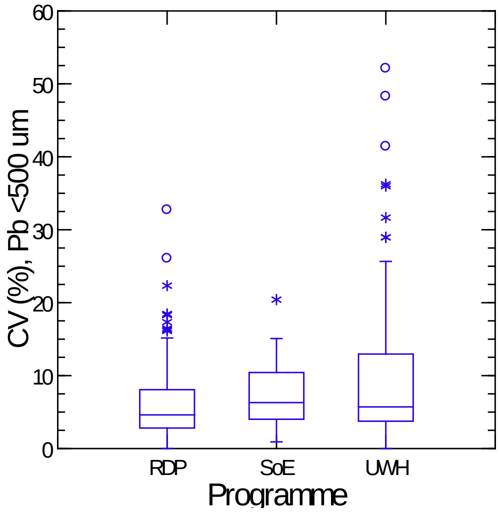
Extractable Metals (<63 µm fraction

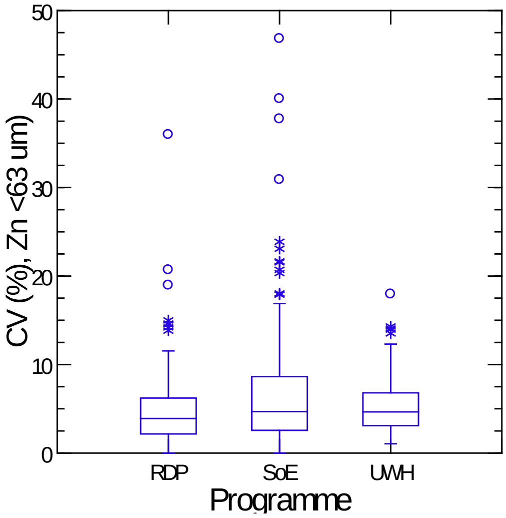
Total Recoverable Metals (<500 µm fraction)











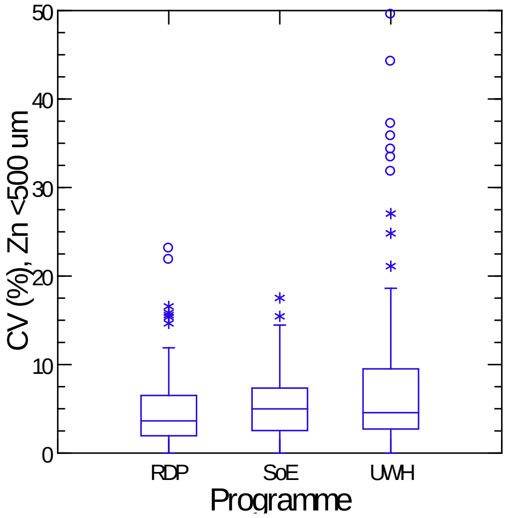


Figure 2-10 Comparison of variability in extractable metals (<63 µm) and total recoverable metals (<500 µm) at RDP, SoE, and UWH programme sites.

Displaying the within-year coefficient of variation; CV, expressed as a %.

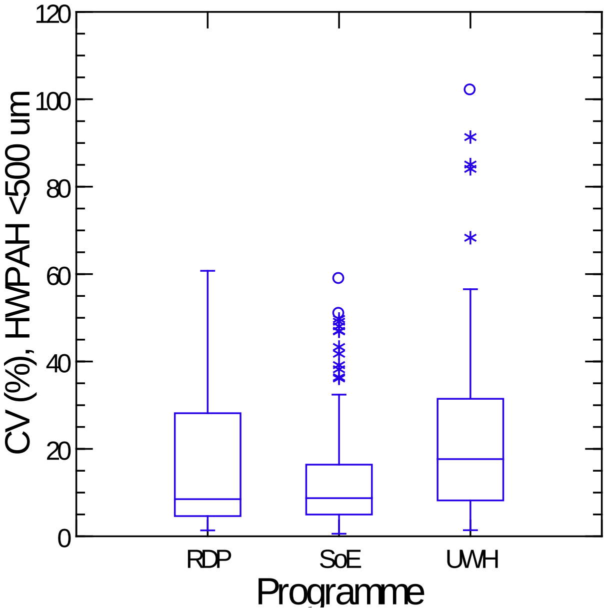


Figure 2-11 Comparison of variability in high molecular weight PAH (<500 µm) at RDP, SoE, and UWH programme sites.

Displaying the within-year coefficient of variation; CV, expressed as a %.

Comparison of variability in the metals data in the three programmes using only the 2008–2011 data again showed that there was no great difference in the median variability between programmes (Table 2-3, Figure 2-12). However, the UWH programme results for total recoverable metals showed markedly higher 90 percentile and maximum CVs than observed for the RDP and SoE programmes. No PAH analyses were carried out in this period at the RDP and SoE sites, so a comparison of PAH variability was not possible.

The similarity of the median variability across the SoE/RDP and UWH programmes in 2008–2011, when sample processing and analysis methods were essentially the same in each programme, suggests that the difference in sampling procedure between the UWH and the RDP/SoE programmes does not markedly affect the variability in metals results.

Table 2-3 Summary of variability in extractable metals (<63 µm fraction), total recoverable metals (<500 µm), and high molecular weight PAH (<500 µm) at RDP, SoE, and UWH programme sites. Displaying the within-year coefficient of variation; CV, expressed as a %. N is the number of CV values used to generate the statistics. Data from 2008–11 only.

		Extractable metals (<63 um)			Total Recoverable metals (<500 um)			
Programme	CV (%)	Cu	Pb	Zn	Cu	Pb	Zn	HWPAH
RDP	N	58	58	58	58	58	58	.
	Maximum	15.0	12.7	14.6	24.1	26.1	15.8	.
	Median	4.7	4.6	4.4	5.3	4.6	3.7	.
	90% ile	9.6	8.8	8.9	15.4	13.1	9.9	.
SoE	N	51	51	51	51	51	51	.
	Maximum	46.5	46.7	37.7	23.4	20.4	17.5	.
	Median	5.3	4.8	4.2	5.8	6.3	5.0	.
	90% ile	12.5	13.9	14.9	16.4	12.5	11.2	.
UWH	N	39	39	39	39	39	39	13
	Maximum	14.2	33.0	14.3	59.5	52.1	35.8	102.1
	Median	6.4	6.0	5.7	7.1	5.5	4.5	23.6
	90% ile	12.6	10.7	11.4	24.7	24.0	29.0	88.5

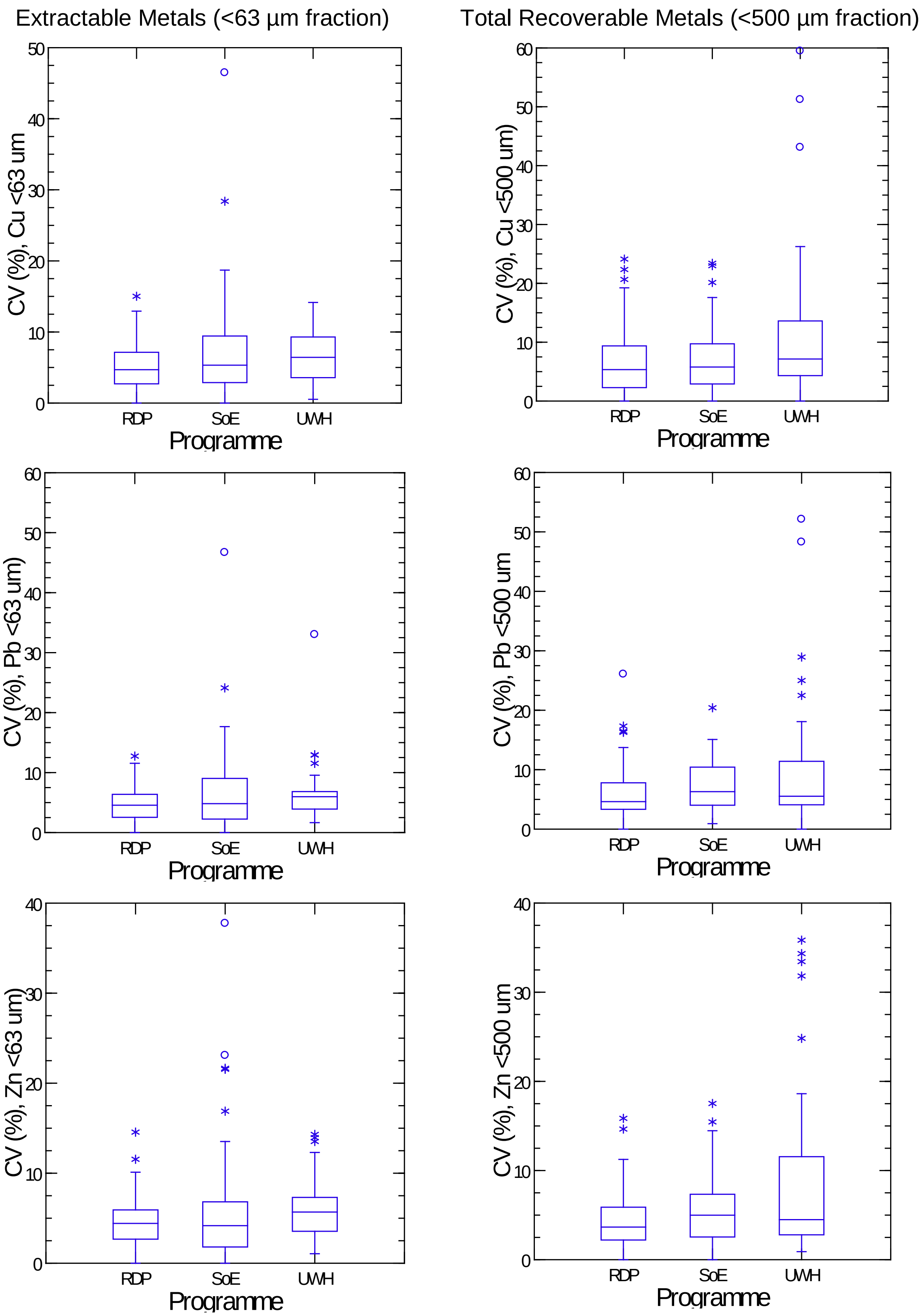


Figure 2-12 Comparison of variability in extractable metals (<63 µm) and total recoverable metals (<500 µm) at RDP, SoE, and UWH programme sites. Data from 2008–2011 only. Displaying the within-year coefficient of variation; CV, expressed as a %.

2.3.6 Changes in variability over time

It is useful to assess whether there have been any marked changes in within year variability over time, given the changes in sampling/processing/analysis that have occurred over time (as discussed previously). Observation of marked changes in variability might signal a need to investigate potential causes, whereas a relatively consistent record would provide some reassurance that changes in programme operation over time have had no major overall effect on data quality (at least in data variability).

The results of this assessment need to be interpreted in only a general sense, because variability can be affected by changes in the numbers and types of sites sampled each year In the RDP programme , the numbers of sites monitored varied from year to year, with the most contaminated sites (in the Environmental Response Criteria (ERC) amber and red ranges) generally sampled at two-yearly intervals, and less contaminated sites (in the ERC green range) sampled approximately every five years. New sites were also added over time to improve spatial coverage. Therefore only a general analysis is valid. For the SoE, the same core sites have been monitored every two years, but beach site monitoring has been less consistent (the data record showed occasional missing data values, and sometimes the sites were not sampled). The same set of sites in UWH programme was monitored on all occasions.

Because of these variations in monitoring in the RDP programme (and to a lesser degree in the SoE), the time series data have not been quantitatively analysed. Instead, changes in variability over time for metals and PAH are shown in Figure 2-13 and Figure 2-14. SoE beach site data have been shown in red.

These plots indicate data variability has changed from year-to-year, but with no major consistent changes over time – e.g. there appeared to be an increasing trend in data variability for extractable Cu and Zn in the SoE programme between 1998 and 2007, but a marked decrease followed in 2009, and then an increase again in 2011. Apart from this, the only large changes evident from the data were in the UWH programme, where markedly lower variability for total recoverable metals was recorded in 2011, and lower variability for extractable metals in 2007. The reasons for these differences are unknown.

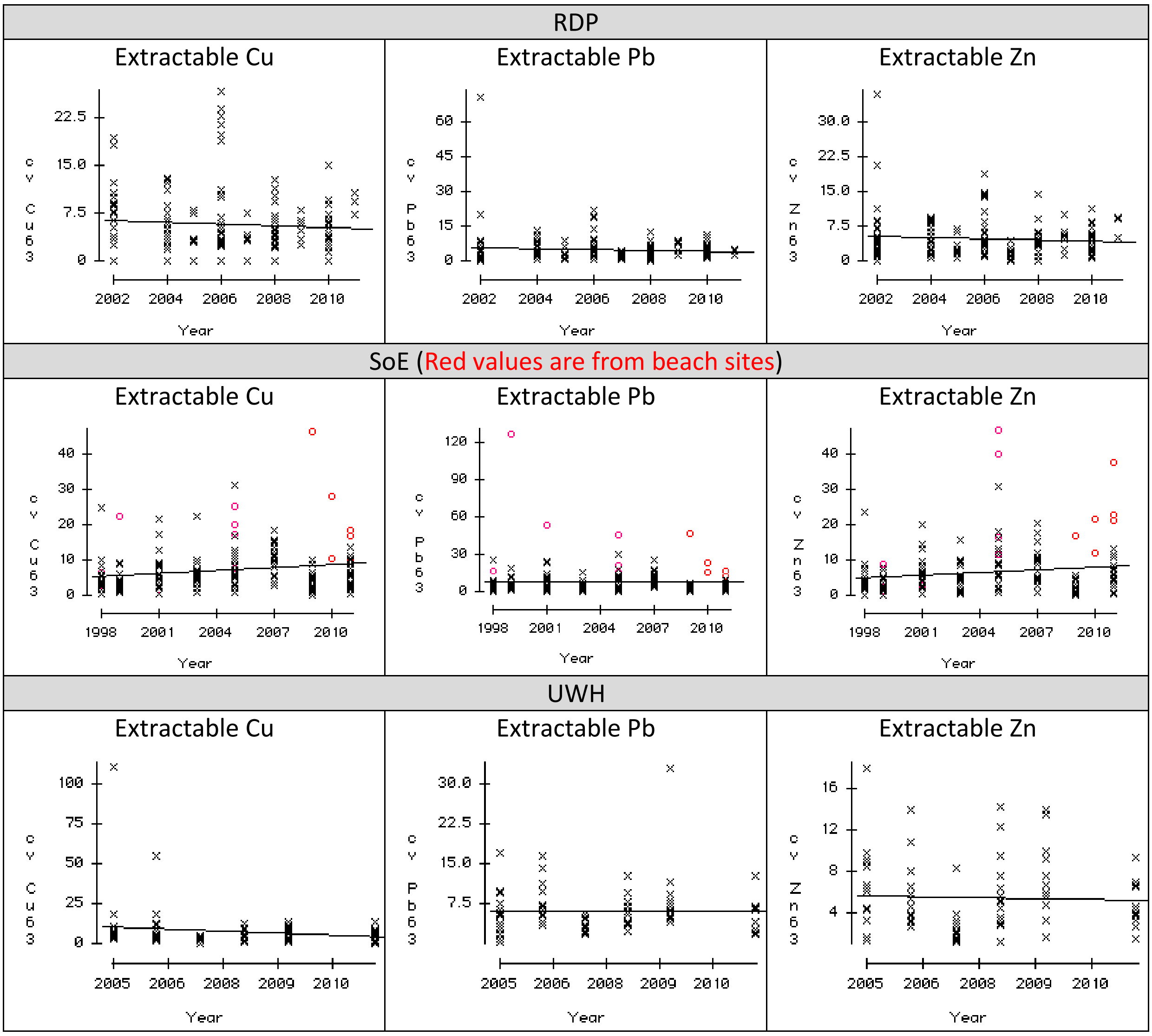


Figure 2-13 Changes in variability (within-year coefficient of variation, CV, expressed as a %) over time for extractable metals (<63 µm fraction) in the RDP, SoE, and UWH programmes.
A linear regression line is shown to provide an indication of trends.

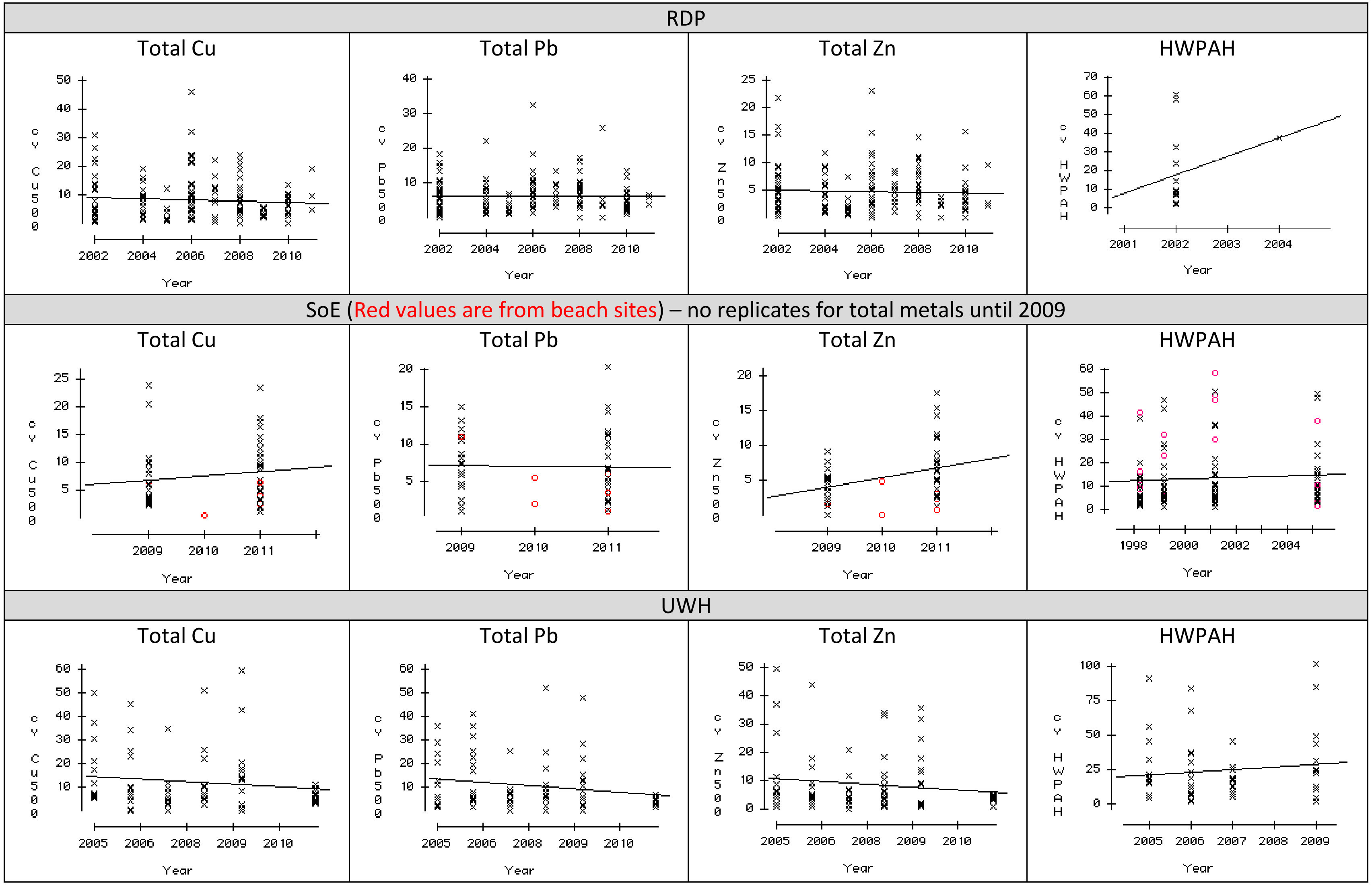


Figure 2-14 Changes in variability over time for total recoverable metals and HWPAH (both in <500 µm fraction) in the RDP, SoE, and UWH programmes. A linear regression line is shown to provide an indication of trends. Displaying the within-year coefficient of variation; CV, expressed as a %.

2.4 Effect of changes in sample preparation in the UWH programme

Between 2005 and 2007 inclusive, sample preparation (homogenisation, sieving, and freeze drying prior to analysis for metals and PAH) was carried out by NIWA. R.J. Hill laboratories were contracted to carry out the analyses of the sieved and dried samples (see Table 1-1).

In 2008–11, samples were provided to Diffuse Sources Ltd (DSL) for processing and analysis. DSL homogenised the samples, transferred them into sample pottles, and submitted them to R.J. Hill laboratories, who carried out the sieving and drying prior to metals analysis.

There was, therefore a difference in sample preparation methods between these periods. Sampling protocols and provider, site locations, and lab analysis of metals remained the same.

For PAH, samples in 2008–11 were homogenised by DSL, then freeze dried at NIWA, before analysis at R.J. Hill labs. This was essentially the same processing and analysis as conducted in 2005–7.

It is useful to compare the data variability (expressed as the variability in within-year CVs from each monitoring site) for metals between the 2005–7 and 2008–11 periods to assess whether marked changes in variability occurred as a result of the changes in sample preparation methods.

Note that while there was a change in only one part of the sampling/processing/analysis protocols, there may also have been changes in the environmental variability at the monitoring sites between the two periods. Aggregating the data from each three year period should reduce the effects of these changes.

The findings are relevant not only to the UWH programme data, but also to the SoE programme. In the SoE, the same change in sample processing protocol occurred from 2009 onwards[5].

Comparisons of variability between the two periods are shown graphically in Figure 2-15 (metals) and Figure 2-16 (PAH). Data are summarised in Table 2-4.

---

[5] Note that prior to 2009, the sieved/dried samples were extracted/digested at NIWA (rather than at R.J. Hill labs, who then undertook the analysis of the NIWA extracts). Differences between the 1998–2007 and 2009–2011 data for the SoE may therefore arise because of the change in extraction/digestion provider.

Table 2-4 Comparison of variability (within-year coefficient of variation; CV, expressed as a %) in extractable metals (<63 µm) and total recoverable metals and HWPAH (both <500 µm) at UWH programme sites for the periods 2005–7 and 2008–11.
N is the number of CV values used to generate the statistics. One high CV value of 111% for extractable Cu (at Hobsonville) in 2005–7 given in brackets was removed for clarity before plotting in Figure 2-15.

		Extractable metals (<63 um)			Total Recoverable metals (<500 um)			
Period	CVs (%)	Cu	Pb	Zn	Cu	Pb	Zn	HWPAH
2005-7	N	41	42	42	42	42	42	42
	Maximum	55.5 (111)	17.1	18.0	50.1	41.4	49.6	91.3
	Median	4.8	4.6	3.8	7.4	6.1	4.6	16.5
	90% ile	12.4	10.4	9.4	34.8	29.8	22.9	49.4
2008-11	N	39	39	39	39	39	39	13
	Maximum	14.2	33.0	14.3	59.5	52.1	35.8	102.1
	Median	6.4	6.0	5.7	7.1	5.5	4.5	23.6
	90% ile	12.6	10.7	11.4	24.7	24.0	29.0	88.5

There was little difference in variability for extractable or total recoverable metals between the two periods. The extractable metals data were slightly more variable in 2008–11 than in 2005–7, but the differences were small – none of the differences in medians (Kruskal Wallis test) or means (t-test) were significant at p<0.05. For PAH, variability was slightly higher (but not significantly so; t-test p=0.245) in the 2008–11 period than in 2005–7. As noted previously, PAH sample preparation and analysis was conducted in essentially the same way in each of these periods.

These results indicate that the change in sample processing for metals analysis made little practical difference to the variability in results.

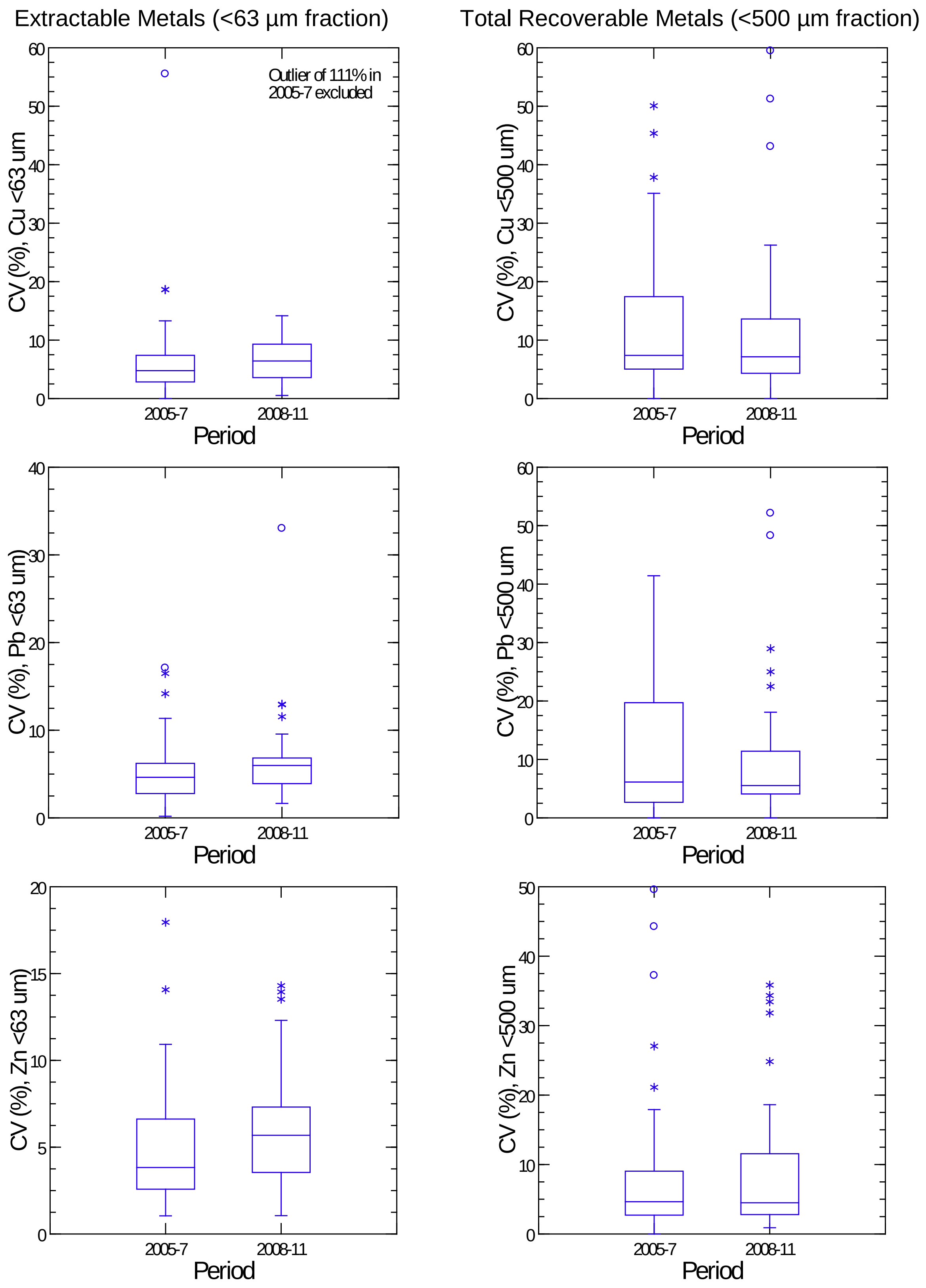


Figure 2-15 Comparison of variability in extractable metals (<63 µm) and total recoverable metals (<500 µm) at UWH programme sites for the periods 2005–7 (sample processing by NIWA) and 2008–11 (sample processing by R.J. Hill labs).

Displaying the within-year coefficient of variation; CV, expressed as a %.

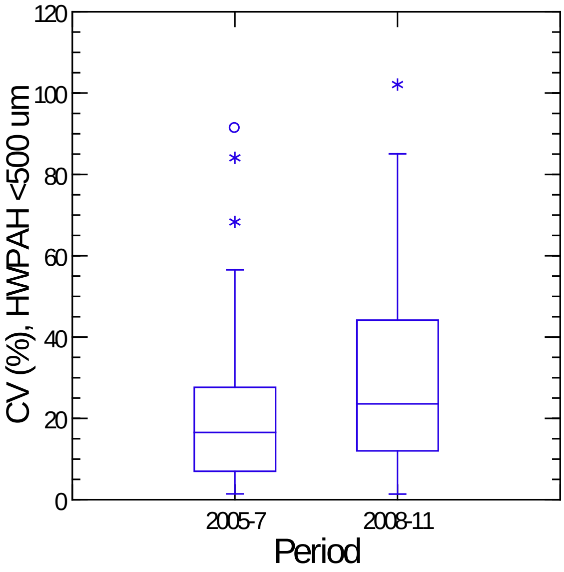


Figure 2-16 Comparison of variability (coefficient of variation; CV, expressed as a %) in HWPAH (<500 µm) at UWH programme sites for the periods 2005–7 and 2008–11.
Samples in both periods were processed and analysed by essentially the same procedures (see text).

2.5 Certified Reference Material analysis

Certified Reference Material (CRM) or Standard Reference Material (SRM) for sediment analysis is homogeneous material that has been characterised within stated uncertainty limits. A certificate provides the list of analytes for which the CRM is certified, and gives a certified concentration ± uncertainty (confidence) bounds for each analyte. The CRM therefore provides a basis for testing the accuracy (and precision, if multiple samples are analysed) of a particular analytical protocol.

CRMs have been used in the sediment contaminant monitoring programme to assess accuracy and variability of laboratory analysis for metals and PAH, as discussed in the following sections.

2.5.1 Metals – accuracy and variability of CRM results

The CRM AGAL 10 has been analysed with each batch of RDP (2002–2010), SoE (2009/11), RSCMP (2011), and UWH (2009–11) monitoring samples. This material is a dried and <212 µm sieved sediment (Hawkesbury River) prepared by the Australian Government Analytical Laboratories. A detailed description of the preparation and characterisation of this CRM is given in AGNMI (2012).

Three to five CRM samples were included in each analytical batch and analysed as “unknowns” by the lab (R.J. Hill). In addition, the lab also included CRM samples in its QA/QC assessment process. These results, which are given in the lab QA/QC reports each year, have been included in the overall assessment of the CRM data. Results are presented in annual monitoring reports (e.g. KML 2003, Diffuse Sources 2005–2011).

The CRM data from the RDP/SoE/RSCMP data series have been analysed and plotted to assess:

- The accuracy of the metals data – how well the lab results compare with the certified concentrations;
- The variability of the metals data –the spread of results obtained each year; and
- Trends in concentrations over time – from examination of the data series acquired from 2002 to 2011.

Important points to note about the CRM analyses are:

- The CRM analyses involved extraction/digestion and ICP-MS analysis only, and did not include the sieving (<63 µm and <500 µm), drying, and sub-sampling steps undertaken for total analysis of the field monitoring samples. The CRM data therefore provide an estimate of the best achievable laboratory analysis performance for extraction/digestion and ICP-MS analysis of homogeneous (<212 µm sieved), dry, sediment.
- Extractable metals have no certified concentrations. The CRM data cannot therefore provide a direct measure of accuracy. However, spike recoveries reported in the R.J. Hill labs QA/QC reports (and documented in RDP annual reports – e.g. Diffuse Sources 2008–2012) provide some assurance of the accuracy of this method. The CRM results do allow the consistency of the results for extractable metals to be tracked over time, which is an important QA requirement for validating trends measured in the monitoring programmes.

The accuracy of the CRM results for total recoverable metals has been assessed by comparison with reference values in each annual monitoring report (KML 2003, Diffuse Sources 2005–2012). In most cases the results were within ±1 standard deviation (s.d.) of the certified value (as defined in the CRM reference certificate). The data and the certified limits are plotted in Figure 2-17.

Average concentrations determined by the lab in the CRM were close to the certified concentrations – average total Cu was 1.6% lower, Pb was within 1% (0.25% higher), and Zn was 6.1% lower. Total Zn concentrations were mostly lower than the certified concentration, and the average (53.6 mg/kg) was close to the lower certified limit (-1 s.d.) of 52.8 mg/kg. Most individual results were within 10% of the certified concentrations, and all were within 20% of the certified concentrations (except for one Pb value in 2008, which was 27% higher than the certified concentration).

Overall, the CRM results indicate good accuracy (on average), although there appears to be a small (a few %) negative bias for total recoverable Zn.

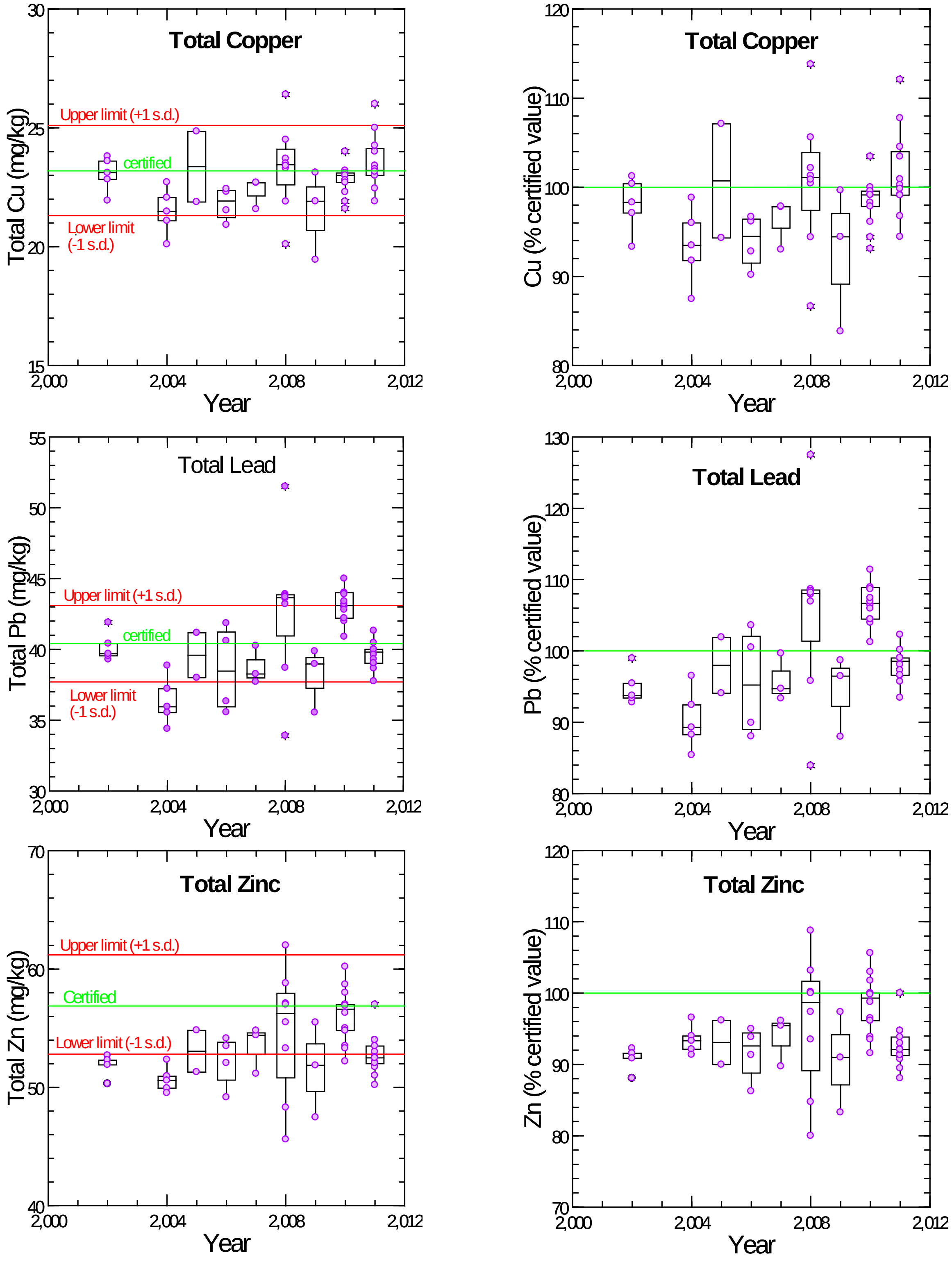


Figure 2-17 Certified Reference Material (CRM) data for total recoverable metals for RDP and RSCMP samples analysed from 2002 to 2011.

Plots show concentrations, with certified values and upper and lower limits (±1 standard deviation; s.d.), and as percentages of the certified values.

The variability in the CRM results have been analysed by calculating the CVs for each monitoring batch (as done for the monitoring data, as described in previous sections), and preparing a statistical summary of these within-year CVs (Table 2-5).

Table 2-5 A summary of within year variability in metals concentrations in CRM AGAL10, from samples used as quality assurance checks in the RDP/RSCMP programmes between 2002 and 2011. N represents the number of CV values used to calculate the statistics.
Displaying the within-year coefficient of variation; CV, expressed as a %.

	Extractable metals (<63 um)			Total Recoverable metals (<500 um)		
CV (%)	Cu	Pb	Zn	Cu	Pb	Zn
N	7	7	7	9	9	9
Minimum	1.92	0.42	0.24	2.88	2.34	1.81
Maximum	8.90	7.11	8.15	8.99	11.70	9.95
Median	5.00	4.01	3.00	4.60	4.70	3.98
Arithmetic Mean	5.02	3.35	3.35	5.23	5.22	4.62
10% ile	2.15	0.43	0.49	2.88	2.38	1.95
90% ile	8.63	6.60	7.48	8.87	10.24	9.08

These results show that the median annual variability in metals concentrations (expressed as a CV in %), ranged from 3–5%. This is approximately 1–2% lower than the median variability found in the monitoring data. The 90%-ile CVs for the CRM analyses ranged from 7–10%, which is somewhat (but not greatly) lower than those observed in the field monitoring data (e.g. 9–19% for the RDP data; Table 2-2). These results suggest that analytical variability (even without sample processing steps such as sieving, drying and homogenising) is a significant contributor to the overall variability in the monitoring data.

While the overall variability in CRM analysis was relatively low, the range between lowest and highest individual values recorded each year was sometimes large – for extractable metals, up to 30%, and for total metals up to 41%. Variability in results obtained in different analytical batches (e.g. between years) is likely to be even higher – maximum differences between individual CRM results for extractable metals were 18–30%, and 30–43% for total metals.

The sometimes large differences between individual analytical results means that comparison of any two individual results could be prone to considerable uncertainty. For this reason, replicated sampling and analysis is required, so that averages (means or medians) can be compared, and statistical tests used to determine the significance of differences between sites and trends over time.

The variability in the metals concentrations observed for the CRM results was similar to that given in the CRM “Report of Analysis” (AGNMI 2012). The relative standard deviations (rsd, %, which is equivalent to a CV) in the reference values are 8.3%, 6.7%, and 7.3% for total Cu, Pb, and Zn respectively (although this variability was measured from analyses conducted by a number of different laboratories, rather than one in-house assessment). For the entire monitoring period (not the annual variations summarised in Table 2-5), the CV values obtained from CRM analyses conducted in the RDP/RSCMP monitoring programmes (2002 to 2011) were 5.6%, 7.7% and 5.9%

for total Cu, Pb, and Zn respectively. For extractable metals, the CVs were 5.7%, 4.5% and 4.4% for Cu, Pb, and Zn respectively. This overall variability includes a contribution from the small trends observed for some metals (see section 2.5.2).

Overall, therefore, it seems likely that the routinely attainable variability for metals analysis is, at best, of the order of 4–8% (as a CV). Because the CRM analyses described above did not include the sample processing steps (sieving, drying, sub-sampling), variability for “real” field monitoring samples may be higher than this. This is currently being investigated through on-going analysis of bulk reference QA samples (BRS). Only one set of BRS results was available at the time of writing this report so no results are presented here. Based on the data analysed in this review, it would seem unrealistic to expect within-batch variability from field monitoring samples (as a CV) for metals of less than 5–10%.

2.5.2 Metals – trends over time in CRM results

The metals CRM data showed no significant trends over time (Mann Kendall test, $p<0.05$), except for total recoverable Cu, which increased at 0.14 mg/kg/year (+0.6% per year). Trend data are summarised in Table 2-6 and plotted in Figure 2-18.

The CRM data indicate that analysis (excluding sample sieving/drying) may have contributed up to ±0.6% per year to temporal trends in the monitoring data, although only the total Cu trend is statistically significant ($p<0.05$). The data also suggest that trends of up to approximately this rate could be associated with analysis alone, and therefore the monitoring programme is unlikely to be able to reliably detect “real” changes of less than this magnitude. Based on these data, it may be prudent to assume that trends of up to approximately ±1% per year over an approximately 10 year period could be associated with sample analysis.

Table 2-6 Trends in extractable (<63 µm) and total recoverable (<500 µm) metals in CRM (AGAL-10) analysed from 2002–2011.
Results from Mann Kendal trend test. Significant trends ($p<0.05$) are shown in red.

Metal	Period	Median (mg/kg)	P	Median annual Sen slope (mg/kg/yr)	5% confidence limit	95% confidence limit	RSSE (% median value per year)
Extractable Cu	2005 - 2011	18.1	0.683	0.03	-0.09	0.17	0.15
Extractable Pb	2005 - 2011	36.5	0.164	-0.22	-0.38	0.05	-0.59
Extractable Zn	2005 - 2011	41.9	0.512	-0.12	-0.53	0.15	-0.29
Total Cu	2002 - 2011	23.0	0.024	0.14	0.03	0.23	0.60
Total Pb	2002 - 2011	40.1	0.130	0.23	-0.03	0.46	0.56
Total Zn	2002 - 2011	53.3	0.139	0.21	-0.01	0.45	0.40

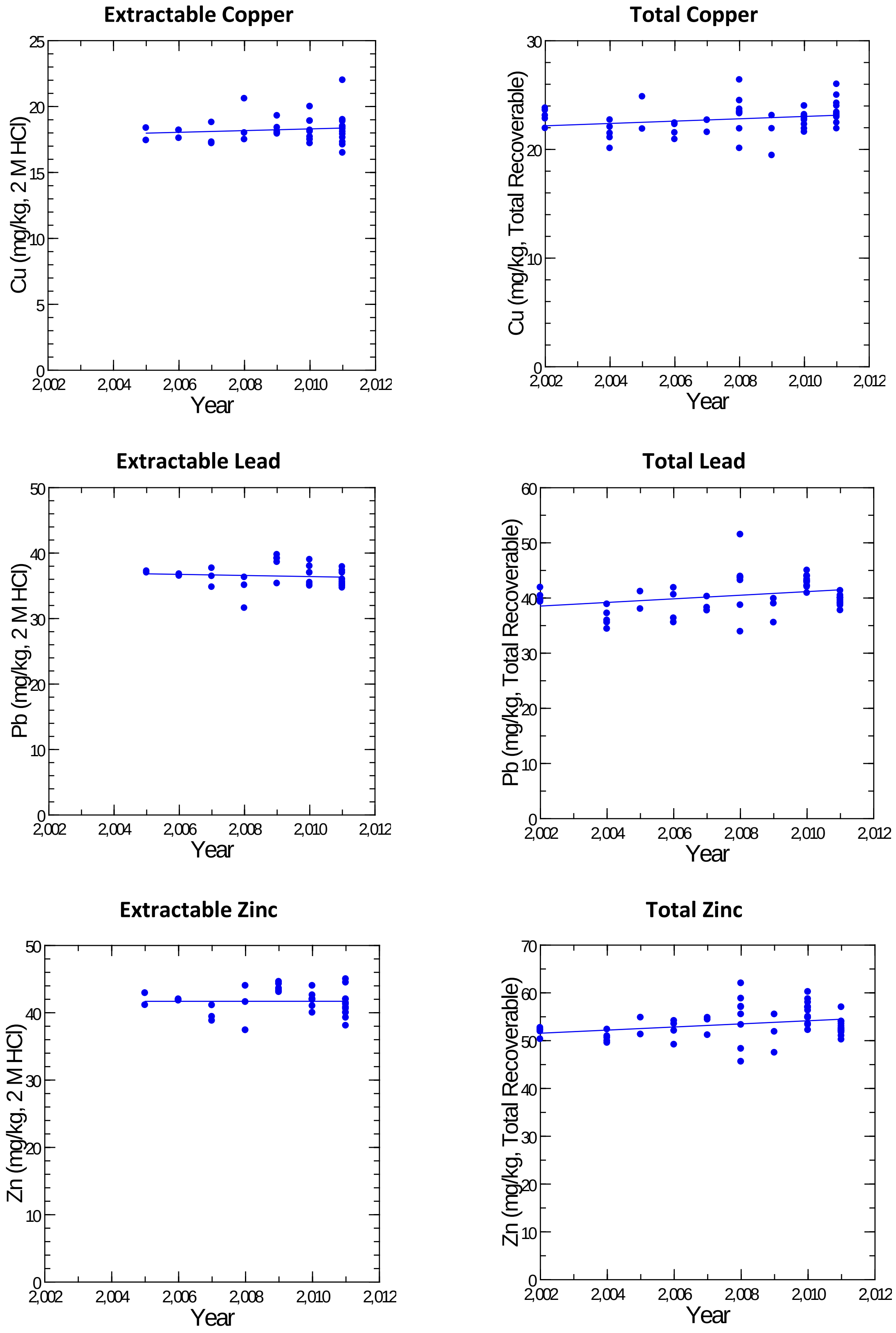


Figure 2-18 Trends in extractable and total recoverable metals in Certified Reference Material (CRM AGAL-10) for samples analysed from 2002–2011.

2.5.3 PAH

PAH analysis has been performed much less frequently than metals. Consequently, there are fewer QA results for PAH, and no regular QA series such as that described above for metals.

In the SoE programme, PAH were analysed on 3 replicates per site at all (or nearly all) sites in 1998, 1999, 2001, and 2005. Analysis was carried out by NIWA. In the RDP, PAH were analysed at 18 sites (3 reps per site) in 2002 and at one site (Chelsea, 3 reps) in 2004. Analysis was by NIWA. In the UWH programme, PAH were analysed at 13 sites (3 reps per site) on four occasions (2005, 2006, 2007, and 2009). Analysis was by carried out by R.J. Hill labs. PAH were also analysed at 22 rural estuary sites (1 replicate per site) in 2009/10 (Hailes et al. 2010, Townsend et al. 2010, Halliday and Cummings 2012, Hewitt and Simpson 2012, Lohrer et al. 2012). Analysis was by R.J. Hill labs. PAH analyses have been conducted on freeze dried, <500 µm sieved samples.

In the SoE programme, the accuracy of PAH results was checked in 1998 and again in 2005 using a standard reference material (SRM) prepared by US National Institute of Standards and Technology (NIST) – SRM 1941a (marine sediment). Results for the SRM analyses (and other QA checks such as duplicate analyses) are given in appendices to the SoE monitoring reports for 1998, 1999, 2001, and 2005 (Williamson et al. 1999, Mills et al. 2000, Mathieson et al. 2002, McHugh and Reed 2006).

Total PAH concentrations reported for SRM NIST 1941a in 1998 were 8.41 mg/kg, and in 2005 the two values were 8.11 and 8.22 mg/kg, compared with the certified concentration of 9.15 mg/kg. The 1998 value was within the certified 95% confidence limits (8.26 to 10.04 mg/kg), but the 2005 values were slightly outside the certified range. However, the concentrations were reasonably close to the certified concentration, being 10.1% and 11.3% below the certified level. Based on these limited data, it appears that the SoE PAH data are likely to be reasonably accurate – possibly a little low, but likely to be within 10–15% of the "real" concentration. As outlined in section 2.8.2, USEPA data quality objectives consider differences between analytical results for organic contaminants and a reference target (e.g. CRM certified value) of up to 35% as being acceptable.

In the RDP programme, no published QA data for PAH were found for this review. For the UWH programme, QA for PAH analysis conducted by R.J. Hill labs included analysis of the standard reference material Wag764, which is a sample used for a quarterly inter-laboratory programme run by Wageningen University (Netherlands). There are no certified values for this SRM, but the concentrations obtained were compared with R.J. Hill laboratory in-house limits, which have been derived from multiple analyses. The SRM was analysed five times in the UWH sample batch. Total PAH concentrations averaged 11.44 mg/kg (CV, 6.7%, N=5) which were within the lab's acceptable range (7.94 to 15.71 mg/kg).

For the rural estuaries surveys in 2009 and 2010, QA for PAH analysis conducted by R.J. Hill labs also included analysis of SRM Wag764. The SRM was analysed twice (once per batch of samples). Total PAH concentrations were 12.09 mg/kg and 11.54 mg/kg, which were within the lab's acceptable range (7.94 to 15.71 mg/kg).

Note that the acceptable ranges for the PAH SRMs are quite different. For the NIST 1941a material used by NIWA, the 95% confidence limits are ±9.7% of the certified concentration, while the in-house limits for Wag764 used by R.J. Hill labs are ±33% of the mid-range level.

2.5.4 Overall conclusions from CRM analyses

CRM results for metals analysed in the RDP (2002–2010), SoE (2009), UWH (2008–2011), and RSCMP (2011) programmes showed generally good agreement with certified concentrations, fairly low variability, and small (mostly not significant) trends over time. Based on these results, the metals data for these programmes appear reliable; within the limitations discussed in the previous subsections. Again, it must be noted that the CRM analysis did not include sample sieving/drying/sub-sampling prior to extraction/digestion, and therefore the results may not fully reflect the performance of the total field sample analysis procedure.

While the average results were generally close to certified concentrations, individual results were found to vary considerably from each other (up to approximately 40%) and, to a lesser degree, from the certified concentrations (generally within 20% of the certified values). Single analytical results may therefore be subject to a considerable degree of uncertainty, and replicated analysis is recommended to obtain robust estimates of metals concentrations.

The median overall within-year variability (CV, %) for metals was about 3–5% for the CRM analysis, which is approximately 1–2% lower than the median variability found in the field monitoring data. These results suggest that analytical variability is a major contributor to the overall variability in the metals monitoring data.

While mostly not statistically significant, trends in metals CRM concentrations of up to approximately ±0.6% per annum were measured over the 2002–2011 period. This indicates that analysis (excluding sample sieving/drying) may have contributed up to ±0.6% per year to temporal trends in the monitoring data and it may be prudent to assume that trends of up to approximately ±1% per year could be associated with sample analysis.

Very limited PAH reference material data acquired in the SoE programme indicated that PAH results were reasonably accurate, probably within 10–15% of “real” concentrations. For the UWH and rural estuaries programmes, SRM data agreed with R.J. Hill laboratory in-house confidence limits, and therefore met the lab’s quality acceptance criteria. The data would therefore be consistent with other PAH data produce by R.J. Hill lab. However, because no certified concentration for the SRM used has been given, the accuracy of the PAH results for these programmes is unknown.

2.6 Agreement between blind duplicate analyses

Blind duplicate samples from field monitoring are analysed to assess how well two results obtained from the same sample agree with each other. This provides a measure of the variation obtained by the analytical methods for individual measurements (including the effects of sub-sampling and sample processing), which is useful for assessing the likely real world significance of differences observed between two analytical results. The duplicate data can also be used to compare data quality with that obtained in other programmes.

Duplicates from the same year's monitoring batch – within batch (WB) duplicates – give information on agreement between analyses conducted for that year. Samples archived (frozen) from previous years' monitoring and reanalysed in later years – between batch (BB) duplicates – provide a check on variations between years.

In addition to these duplicates, instrumental analysis duplicates were included in analytical batches reported by RJ Hill Labs. These Lab WB duplicates were for repeat analyses of metals digests by ICP-MS – i.e. the instrumental part of the analysis method[6]. Lab WB data for samples analysed in the RDP (2002–2009), UWH (2008–2009), SoE (2009 only), rural estuaries (2010) and RSCMP (RDP, SoE, UWH in 2011) have been included in the following assessment. The Lab WB data are useful because they provide a measure of the contribution to overall data variability attributable to instrumental analysis alone.

Agreement between duplicates was assessed by calculating the relative percent difference (RPD) between two results (e.g. duplicates, A and B):

$$RPD = \left[\frac{|A - B|}{(A + B)/2}\right] x100\ \%$$

Note that the RSCMP monitoring does not rely on single analytical results. Replicated sampling and analysis is now used for all sediment contaminant status and trend assessments, with N usually equal to three samples per site from each monitoring round. Single samples (sometimes composites made up from a combination of individual replicates) are generally only used for expensive analyses such as organochlorines, or for characterisation of physical properties (e.g. particle size distribution).

---

[6] The R.J. Hill lab QA/QC reports report these data as "duplicates", and it was not previously clear that they were not full duplicates, but were only reanalyses of sample digests (rerun through the ICP-MS instrument). These data have been separated from the data obtained for "real" blind duplicates submitted with each batch of samples (and put through the complete analytical protocol), and the former have been labelled as "Lab WB" replicates in this review. The Lab WB data are useful because they provide a measure of the contribution to overall data variability attributable to instrumental analysis alone.

The following discussion focuses mainly on data acquired in the former RDP programme (2002–2010), the SoE (2009), and RSCMP (2011), because these form the bulk of the duplicates' database. Duplicate analyses were also conducted in the SoE programme between 1998 and 2007 but, as discussed previously (and further in section 4.0), analysis of these QA results was confounded by uncertainty about which year's archived samples had been reanalysed each year. The SoE programme did, however, include some WB and BB duplicate data for PAHs. These have been included in the following sections.

One set of WB duplicate data (five samples) for metals and PAH was also available from the UWH programme for samples taken in 2009, and six sets of metals duplicates from the rural estuaries sampled in 2010. The results for these were included in the analysis.

A summary of the agreement between duplicates in each of the WB, BB, and Lab WB categories is given in Table 2-7, and shown graphically in Figure 2-19 (metals) and Figure 2-20 (PAH).

Table 2-7 Summary of relative percentage differences (RPD, %) for within batch (WB), between batch (BB), and ICP-MS laboratory within batch (Lab WB) duplicates.

		Extractable metals (<63 um)			Total Recoverable metals (<500 um)			
Type		Cu	Pb	Zn	Cu	Pb	Zn	Total PAH
WB	N	54	54	54	54	54	54	31
	Maximum	50.0	51.3	46.7	65.0	29.2	29.0	85.3
	Median	5.4	3.6	5.1	5.6	4.4	4.1	9.2
	90%-ile	13.5	12.8	11.6	25.0	20.3	10.6	30.4
	95%-ile	31	14	17	37	22	18	70
BB	N	49	49	49	49	49	49	20
	Maximum	47.8	40.0	40.0	39.1	29.8	25.7	33.6
	Median	12.8	10.1	10.5	12.3	9.5	7.2	8.8
	90%-ile	26	29	24	29	22	20	19
	95%-ile	38.6	35.9	37.4	34.1	26.8	21.5	28.6
Lab WB	N	25	25	25	30	30	30	.
	Maximum	15.8	13.1	12.5	20.1	19.3	16.7	.
	Median	3.8	3.3	2.3	3.8	2.9	2.9	.
	90%-ile	10.9	9.9	9.6	12.1	9.9	8.6	.
	95%-ile	14.5	12.9	11.4	17.4	13.6	9.9	.

Agreement between blind duplicates was generally quite good. Median WB RPDs were 4–6% for metals and 9% for PAH. Median BB RPDs were 7–13% for metals and 9% for PAH. Lab WB metals duplicates (reanalysis of digests only) were lower, median RPDs ranging from 2–4%. Between batch RPDs were higher than WB, except for PAH, which were similar.

However, while the agreement was usually good, RPDs were sometimes very high – for metals, maximum RPDs were 48% for BB duplicates, and 65% for WB duplicates. The greatest RPD for PAH was 85% (a WB duplicate for a low level PAH sample, Big Muddy in the SoE programme in 1999). The maximum RPD for repeat ICP-MS analysis (Lab WB duplicates) was 20%.

USEPA Measurement Quality Objectives (MQO) for agreement between duplicates is 95% of the RPDs should be <30% (USEPA 2010). The WB duplicate data met these objectives for total and extractable Pb and Zn, but not for Cu (RPDs of 31% and 37% for extractable and total Cu respectively) or PAH (95%-ile RPD of 70%).

The BB duplicate data met the USEPA MQO for total Pb and Zn and for PAH, but not for total Cu or any extractable metals.

There were no significant relationships between metals concentrations and RPDs apparent from the data – i.e. lower concentration samples did not seem to have poorer agreement between duplicates than higher concentration samples (a weak relationship, with lots of scatter, was observed for total Cu, but this was not significant for either WB or BB duplicates).

Highest RPDs for PAH were generally found with low concentration samples, but there were relatively few duplicate data for analysis (N=51; 31 WB duplicates and 20 BB duplicates). For WB duplicates the correlation was nearly significant (p=0.056) but not for BB duplicates (p=0.309).

Agreement between blind duplicates and sediment texture (measured as mud content, the % <63 µm) showed variable results (Figure 2-21). For within-batch blind duplicates (WB), there were no significant correlations between RPDs and mud content for extractable or total recoverable metals, nor for PAH. For between-batch duplicates (BB), significant correlations were observed between RPDs and extractable metals in the <63 µm fraction (linear regressions, p = 0.0004–0.0022), with higher RPDs (i.e. poorer agreement between duplicates) being recorded for samples with lower mud content. For total recoverable metals, Cu and Pb showed no significant correlation with mud content, but Zn showed increasing RPDs between duplicates with increasing mud content (p≤0.0001) – the opposite effect of that observed for extractable metals. PAH showed a nearly significant relationship between duplicate agreement and mud content, with higher BB RPDs at lower mud content (p=0.062).

These results suggest that within an analytical batch, the agreement between blind duplicates was largely independent of sediment texture. However, for analysis in different analytical batches, there was significant textural effect, with extractable metals results generally being more variable for samples with lower mud content (i.e. sandier sediments with lower <63 µm fraction content).

There were highly significant correlations (p≤0.0001) for the RPDs between metals in each of the <63 µm and <500 µm fractions, but the relationships showed considerable scatter (R^2 = 0.30 to 0.50). Generally, this indicates that, within the same fraction, metals RPDs are correlated with each other, so that if high RPD was observed for one metal (say Zn) it would also be found for the others (Cu and Pb).

There was no consistent trend in RPDs over time, indicating no major change in agreement between blind duplicates over the 2002–11 period. No consistent differences in RPDs for samples analysed in the RDP, SoE, UWH, or 7 estuaries monitoring programmes were observed.

Overall, the duplicate data indicate that agreement was usually good (median RPDs <13%), but occasionally poor agreement (RPDs of up to 85%) was evident from the data. Instrumental analysis of metals was not a major contributor to the higher variability results. Between batch variability was associated with sediment texture for extractable metals (in the <63 mm fraction) with sandier sediments having poorer agreement between duplicates. For total recoverable metals, there was no significant effect of sediment texture, except for total Zn, which showed poorer agreement between duplicates for muddier sediments. Reducing the frequency of the high RPD duplicate results is required to consistently meet USEPA quality objectives for data precision.

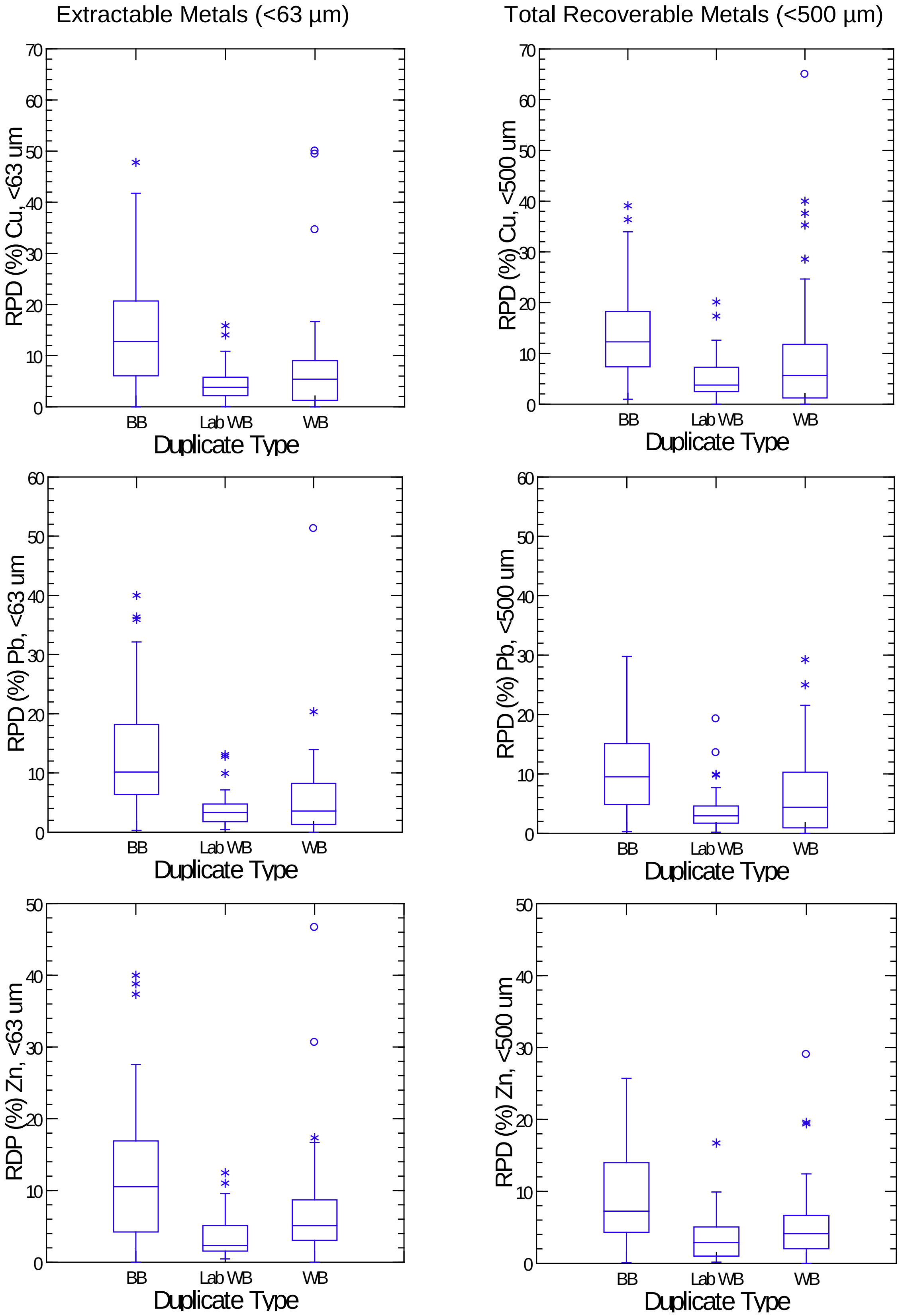


Figure 2-19 Relative percentage differences (RPD, %) for within batch (WB), between batch (BB), and ICP-MS laboratory within batch (Lab WB) duplicate analyses of extractable (<63 µm) and total recoverable (<500 µm) metals.

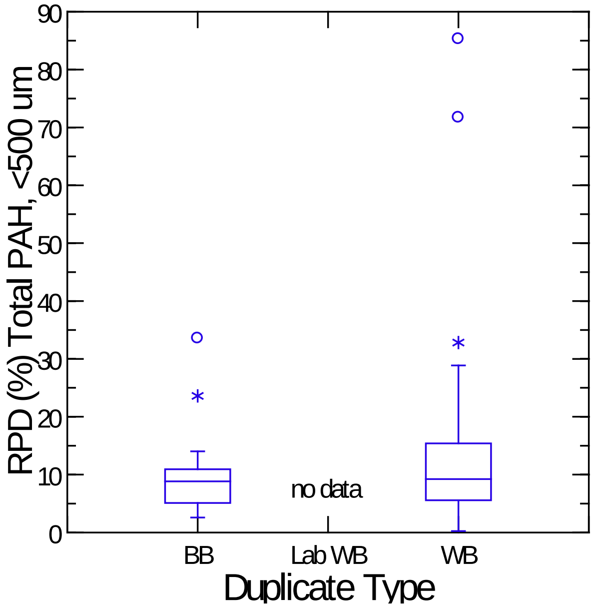
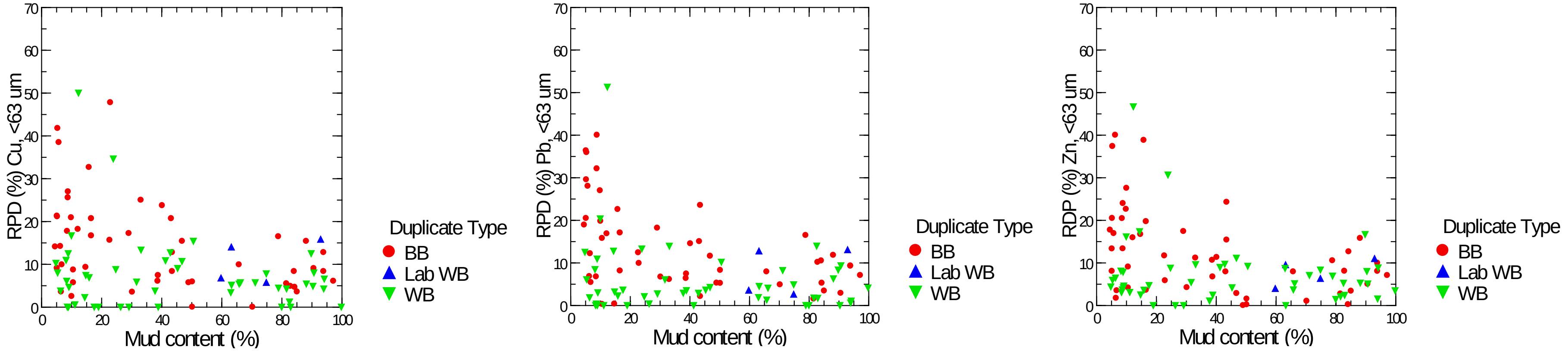


Figure 2-20 Relative percentage differences (RPD, %) for within batch (WB) and between batch (BB) duplicate analyses of PAH.

Extractable metals (<63 um)



Total recoverable metals (<500 um)

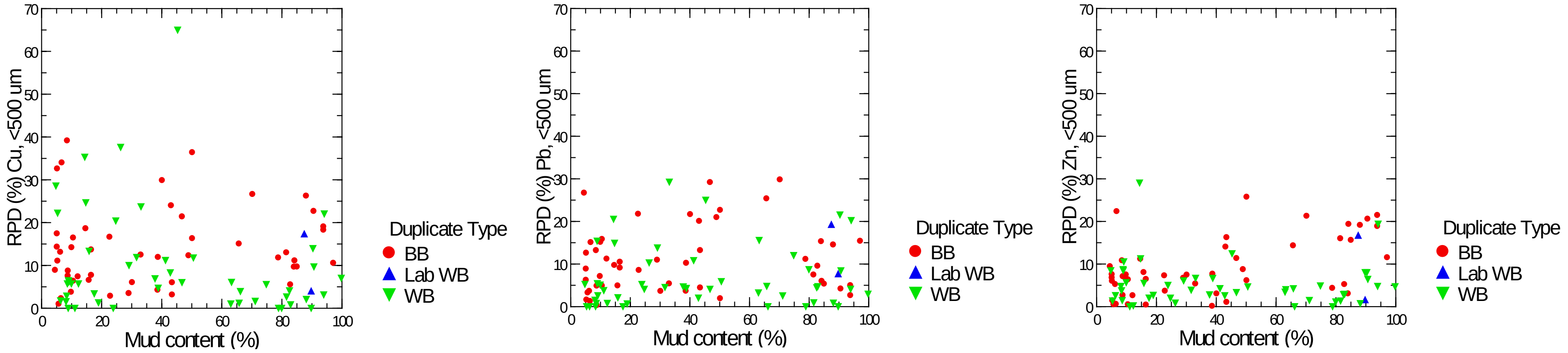


Figure 2-21. Variation in relative percentage differences (RPDs) between blind duplicates for extractable and total recoverable metals with sediment texture (mud content; % <63 um).

Data are shown for between-batch (BB), within-batch (WB), and laboratory within-batch (Lab WB – reanalysis of extracts by ICPMS only) duplicates for all results from 2004–2011.

2.7 Implications of data variability for trend detection

Reliable trend analysis requires accurate monitoring data, with minimal shifts (bias) in results over time due to analytical variation. The CRM results suggest that analytical results (in the RDP and RSCMP programmes) have been reasonably consistent over time, although trends due to analytical drift of up to ±0.6% per annum between 2002 and 2011 were recorded (section 2.5.2).

For sensitive trend detection, low data variability is required. As the variability increases, the numbers of samples required to detect a statistically significant trend increases. Given a constant sampling frequency (e.g. biannually, as at most of the Auckland Council sites), the time required to detect trends also therefore increases.

A useful planning tool for estimating the effect of data variability on sample numbers, or time, required to detect a trend of a given magnitude, with a given probability is the sample size formula given in Ward et al. (1990):

$$n = \frac{12s^2\left[t_{\alpha/2,(n-2)} + t_{\beta,(n-2)}\right]^2}{\Delta^2}$$

where:

n – number of samples required to detect a linear trend

s – standard deviation of data <u>in absence of a trend</u>

Δ – trend over whole period of record

α – Type I error risk - inferring a trend equal to Δ when none exists

β – Type II error risk - failing to detect a trend when it does exist, equal to Δ

t – values are one-tailed.

Once the user has defined acceptable error risks (α and β), and the variability (standard deviation) in the data is known, the numbers of samples required to detect trends of a given magnitude can be calculated. Alternatively, if the trend detection target is defined (e.g. a 20% change over 10 years), then the data variability required to meet this target can be estimated.

An example is shown in Figure 2-22, where the effect of increasing data variability on the numbers of samples required to detect trends of 10% and 20% of the analyte concentration are shown.

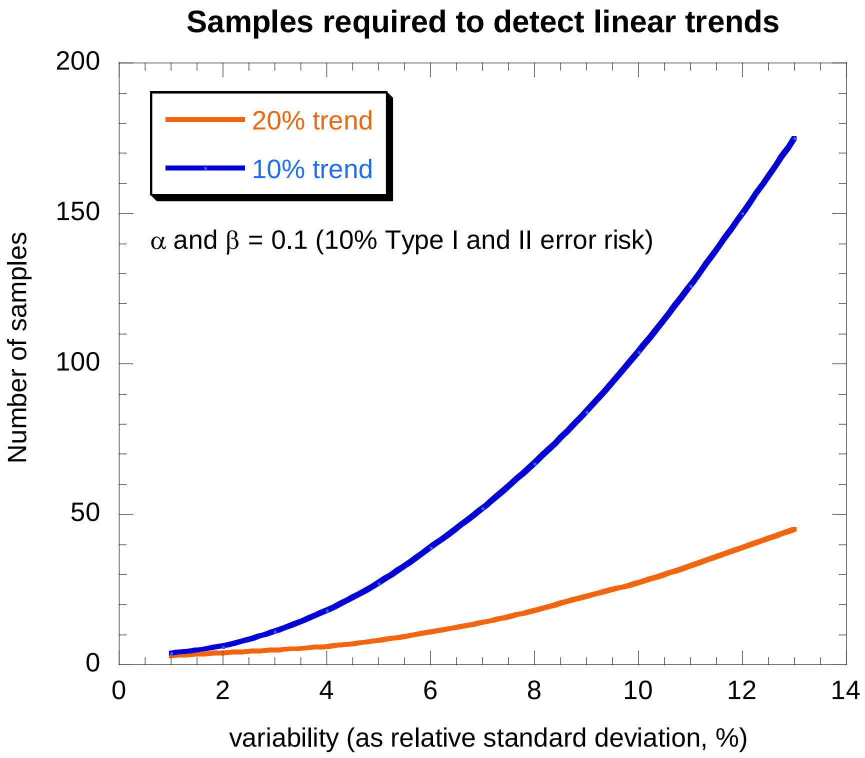


Figure 2-22 Effect of data variability on numbers of samples required to detect linear trends of magnitudes of 10% and 20% of the analyte concentration.
Data expressed as relative standard deviation, equivalent to the coefficient of variation (CV) expressed as a %). Data plotted using the formula of Ward et al. (1990).

Assuming α and β = 0.1 (10% Types I and II error risk), these data could be used to explore trend detection and sampling scenarios. The following examples are used to illustrate this approach, assuming the data variability referred to is the variability in the results when only one data value is recorded from each sampling occasion (i.e. from single samples if only one sample is taken and analysed each time, or from the mean (or median) result from replicate samples and analyses from each sampling occasion):

- If the data variability (as a relative standard deviation, or CV expressed as a %) is 5%, it would require 27 samples to detect a linear trend of 10% of the analyte concentration. At a sampling rate of once every two years (as at most AC monitoring sites), this would take 54 years. The trend slope would therefore be a 10% change over 54 years = 0.19% per year. This is less than the slope (trend) observed in the CRM data between 2002 and 2011 (section 2.5.2), suggesting that analytical performance would need to be improved to be confident that monitoring could detect a trend of this magnitude.
- If the data variability (as a CV, %) is 10%, it would require 104 samples to detect a linear trend of 10% of the analyte concentration. At one sampling every two years (as at SoE sites), this would take 208 years. The trend slope would therefore be 10% over 208 years = 0.3% per year. Again, this slope may be too small for current analytical methods to reliably detect. In addition, the time frame for the trend to become significant is very long – possibly beyond a useful period for making resource management decisions.

- If trend results were required for reporting in 20 years, and sampling occurred every 2 years (i.e. 10 samplings), predict the data standard deviation to achieve detection of a:
  - 10% change (equivalent to a 10 mg/kg increase in the average Zn concentration of about 100 mg/kg in Auckland sediments) – this would require the data variability (CV, %) to be approximately 2–3%, and
  - for a 20% trend, variability (CV, %) would need to be about 5–6%.

Clearly there is a need to minimise data variability in order to reduce the period, and sample numbers, required to detect trends.

It is important to note that the AC monitoring programme approach involves taking and analysing multiple replicate samples (usually N=3) from each site on each sampling occasion in order to provide a more robust estimate of the mean (or median) result from each sampling event and to reduce the temporal variability in the trend series. This should reduce the time required to detect meaningful trends (or reduce the numbers of samples required to do so), but at this stage it is unknown how this replicated sampling and analysis affects trend detection capabilities.

The average annual variability (as a CV, %) in the monitoring data acquired to date is typically approximately 5–10%[7]. The CRM analysis results provide an indication of the lowest possible variability, resulting from analysis only (not including sample preparation) of a uniform, sieved (<212 µm) sediment. The overall variability (as a CV, %) in CRM metals concentrations, for all analyses conducted between 2002 and 2011, ranged from 4.4% (for extractable Zn, N=36) to 7.7% (for total Pb, N=56).

From this initial examination of sample numbers required for trend estimation, it would seem that being able to detect a 10% trend with a 90% probable degree of certainty, may take about 50–200 years, if only one data point was produced every two years. For a 20% trend, the monitoring period would be about 20–50 years. For sites with greater variability, detecting trends of this order of magnitude within a time frame that is useful for resource management may be unrealistic. However, again it must be stressed that the AC monitoring programme involves analysis of replicate samples taken on each occasion, and this is likely to reduce these time frames.

This approach to predicting the likely outcomes of, and/or requirements for, monitoring should be further explored so that end users of the monitoring understand the limitations of the programmes, and have realistic expectations of the resources and time frames required to obtain meaningful results. Assessing the level of Type I and II error risk that users would be prepared to accept in order to report on trends within reasonable time frames, or to reduce sampling effort and cost to acceptable levels should be undertaken. The effect of replicated sampling on each monitoring occasion (as done in the AC programmes) on overall data variability and hence sample number predictions also needs to be determined.

The sample number estimation tool highlights the importance of reducing data variability for effective detection of trends. Since a significant proportion of the variability in the monitoring data

---

[7] The standard deviation used in the sample number equation is the standard deviation in the absence of a trend. These values would need to be calculated for each site to get accurate sample number predictions.

collected to date is associated with sample analysis (including the effects of within-sample heterogeneity, sample processing, and chemical analysis), exploring ways to reduce analytical variability would seem worthwhile. There may be little point continuing to monitor and analyse for trends if the analytical methods for measuring target analytes are too variable. This may be particularly relevant for more complex analyses of low concentration chemicals (e.g. organic contaminants such as PAH and organochlorines), for which higher analytical variability is often normal. It may be unrealistic to expect to be able to measure anything other than major changes in these contaminants over time.

2.8 Comparison with other programmes

To put the Auckland Council monitoring data quality into context, this section compares the Auckland Council data with that from similar programmes elsewhere. Unfortunately, there are few estuarine monitoring programmes of a similar nature for comparison. In New Zealand, the only programme we know of is that run by Greater Wellington Regional Council (GWRC) in the Porirua and Pauatahanui estuaries (Milne et al. 2009).

Overseas, the US National Coastal Condition Assessment program run by the USEPA, as part of the nationwide assessment of sediment contamination (USEPA 2004), provides information on very broad scale monitoring and analysis of levels and trends in sediment contamination. The assessment quality assurance plan (USEPA 2010) provides measurement quality objectives (MQO) for sediment indicators that can be compared with Auckland Council data quality.

The US National Oceanic and Atmospheric Administration’s (NOAA) mussel watch programme includes collection of sediments from over 250 sites across the USA. The National Benthic Surveillance Project (NBSP) ran from 1984 to 1993, and also sampled sediments and bottom dwelling fish from over 100 sites. The quality assurance protocols are documented in the analytical methods’ manuals (e.g. Lauenstein and Cantillo, 1993).

A status and trends sediment monitoring has been conducted in San Francisco Bay since 1993 (SFEI 2011). Details of the monitoring projects, including revisions to sampling designs can be found in Appendix 9 of SFEI (2011), and at http://www.sfei.org/projects/3564.

2.8.1 Greater Wellington Regional Council programme

Sediment contaminant and particle size distribution (PSD) data from the GWRC monitoring programme in Porirua and Pauatahanui estuaries was made available by GWRC for comparison with Auckland Council data.

The variability (within-year CVs) in the GWRC data is compared with that in the Auckland Council programmes (SoE – excluding beach sites, RDP, and UWH) in Table 2-8 and Figure 2-23. The data cover the period 2004 to 2010 (inclusive) – a period for which data was available for all programmes. Comparisons were only possible for extractable metals, because total recoverable metals were analysed only on a single composite sample in the GWRC programme and also for the SoE years 2005 and 2007.

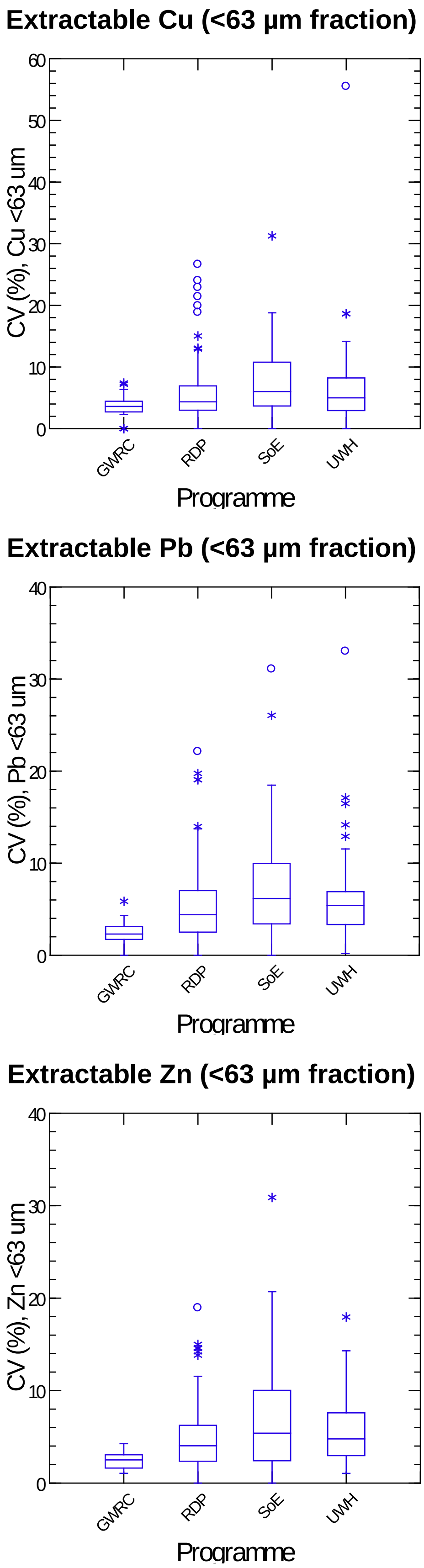


Figure 2-23 Comparison of variability in extractable metals (<63 µm) in Greater Wellington Regional Council (GWRC), and Auckland Council – RDP, SoE, and UWH – monitoring programmes (2004–2010). Variability data are coefficients of variation (CV, expressed as %) from within-year replicates (n=5 reps per site per year for GWRC and n=3 per site per year for Auckland Council data).

Table 2-8 Variability in extractable metals results from Greater Wellington Regional Council (GWRC), and Auckland Council – Regional Discharges Project (RDP), State of the Environment (SoE), and Upper Waitemata Harbour (UWH) – monitoring programmes (2004–2010).
Variability data are coefficients of variation (CV, expressed as %) from within-year replicates (n=5 per site per year for GWRC and n=3 per site per year for Auckland Council data).

		Cu (<63 um)			Pb (<63 um)			Zn (<63 um)		
Programme	N	min	max	median	min	max	median	min	max	median
GWRC	20	0.0	7.4	3.6	0.0	5.9	2.3	1.1	4.3	2.5
RDP	113	0.0	26.7	4.4	0.0	22.1	4.4	0.0	19.0	4.0
SoE	63	0.0	31.2	6.0	0.0	31.1	6.2	0.0	30.9	5.4
UWH	69	0.0	55.5	5.0	0.2	33.0	5.4	1.0	18.0	4.8

These data indicate that the GWRC programme extractable metals data was less variable than those obtained in the Auckland Council programmes (and also lower than CRM variability obtained in AC monitoring). Possible reasons for this lower variability could include:

- The smaller number and range of site types in the GWRC programme. The Auckland Council programme sites cover a much broader range of sediment textures, and levels of contamination. Note that SoE programme open coastal beach sites were excluded from the comparison, because of their very different physical characteristics from all other sites;
- The GWRC sites are sub-tidal and are therefore not prone to the same potential level of physical disturbance (e.g. from channel erosion, mangrove encroachment, human access and interference) as the Auckland Council's intertidal sites, which can be visibly heterogeneous (e.g. humps and hollows, variable sediment textures, litter, debris on surface and beneath the sediment). However, both GWRC and AC monitoring protocols aim to reduce the effects of within-site heterogeneity by compositing multiple replicates from across each site. The GWRC sampling protocol takes 25 core samples from within a 20 m diameter circle, and combines 5 randomly selected samples to each of 5 composited replicates (which are the samples analysed). The AC approach takes 50 individual samples from along two (mostly) 50 m transects at each site, and combines 10 samples into each of 5 composited replicates (3 of which are analysed, leaving 2 for follow up work if required);
- Depth sampled – the GWRC samples were taken from the top 3 cm of the core, while AC samples are surface samples of approximately 2 cm depth. The GWRC method, using small cores from which the samples are extruded to measured depths, provides a more accurate and consistent depth of sampling than the AC method.
- Sample processing prior to metals analysis. The GWRC samples were all processed by NIWA (<63 µm sieving and freeze drying) before analysis by R.J. Hill labs. The RDP samples were sieved, dried and analysed at Hills, while the SoE sediments were sieved and extracted at NIWA, then analysed at Hills (2005 and 2007), or sieved, dried ,and analysed at Hills (as per RDP samples) in 2009. Since analysis of the digests from all programmes was undertaken at Hill labs, the difference in variability may reflect lower variability with the sieving/freeze drying procedures used at NIWA. However, as discussed

in section 2.4 for the UWH programme, little difference in data variability was observed when sample processing changed in this way; and

- A greater number of replicates per site in the GWRC programme (n=5) compared with the Auckland Council programmes (n=3). Increasing the number of replicates analysed would reduce the within-year sample standard deviation and reduce the uncertainty in the annual mean (or median), which would be beneficial for trend monitoring.

Given that reducing variability is a key factor for improving trend detection power, establishing the reason(s) for the lower variability obtained in the GWRC monitoring would be worthwhile.

2.8.2 US National Coastal Condition Assessment data quality

The US Coastal Condition programme reports indices for sediments, based on toxicity, contaminant concentrations and TOC content, rather than contaminant concentrations directly (e.g. USEPA 2008). There are, therefore, no useful status or trend data that could be compared directly with the Auckland Council monitoring results.

However, the assessment quality assurance plan (USEPA 2010) provides measurement quality objectives for sediment indicators that can be compared with the Auckland Council data quality.

The measurement quality objectives (MQO) for sediment indicators (listed in Table 5.5-9, USEPA 2010) give:

- A maximum allowable accuracy (bias) goal of 20% for inorganics and 35% for organics. The bias is defined as the mean of the test results (e.g. CRM results) minus the target concentration (e.g. CRM reference value) divided by the target value (x 100, to express as a percentage).
- The maximum allowable precision goal for inorganics and organics of 30%. The precision is defined as the relative standard deviation (standard deviation divided by the mean, x100 to express as a %) for more than two results, or as the relative percent difference (RPD) between two results (e.g. duplicates, A and B):

$$RPD = \left[\frac{|A - B|}{(A + B)/2}\right] x100\ \%$$

- A 95% completeness goal – the percentage of expected results that are obtained successfully (i.e. 95% of the results should meet the above QA specifications).

These results are similar to those obtained in the Auckland Council programme, with the possible exception of agreement between replicates, which (as discussed in section 2.6) did not meet the USEPA MQOs for all analytes.

CRM results in the Auckland Council programme were also mostly within 20% of the certified concentration, only a single Cu result lying outside this range (27%). This degree of accuracy would meet the US accuracy goals for inorganics (20% for 95% of results).

Overall, therefore, it is probable that the Auckland Council results would have met the accuracy (bias) MQOs of the US National Coastal Condition monitoring programme, but not the variability (precision) goal – the latter because of occasionally high RPDs between duplicates.

2.8.3 San Francisco Bay sediment metals data quality

Trends in sediment contaminants have been reported in San Francisco Bay for 1993 to 2005 (SFEI 2006). Nearly all the trends were either "no trend" or decreases over time. Out of 10 elements (excluding methylmercury) monitored at seven sites, only three increasing trends (at two sites for Cd, and at one site for Cu) were recorded. Only one increase (PAH at one site) and one decrease (PCBs at one site) were recorded for organic contaminants.

SFEI (2011) reported on sediment contaminant data quality. Precision for replicates was shown to be good; RPDs between replicates, or relative standard deviations (RSDs) of replicated analyses, were <25% for all target analytes. Reference material sample data were good for the target analytes, with only aluminium being outside the average error target of 25% (due to the digestion method used). Average concentrations of elements were 80–120% of previous years' averages. Based on these measures, the quality of the San Francisco Bay monitoring results is similar to those obtained in the Auckland Council programme.

The use of a ±20% variation between the latest monitoring results and the average of previous years' data is a simple QA approach for defining data acceptance criteria and identifying suspect data that require further checking. This approach could be explored as a possibility for Auckland Council.

2.8.4 US National Status and Trends (NOAA) programme data quality

The NOAA reporting has focused on contaminant concentrations in mussels for trend assessment (Kimbrough et al. 2008). No trend data for sediment contaminants were found in this review. However, the programme has QA protocols, the most relevant of which are summarised below.

The analytical methods for biota and sediments have been well documented (Lauenstein and Cantillo 1993; Lauenstein and Cantillo 1998), including quality assurance targets for organic contaminants, but not for inorganics.

Acceptable limits of precision for organic control materials are ±30% on average for all analytes, and ±35% for individual analytes. These limits apply only where the concentrations of the compounds of interest are at least 10 times greater than the method detection limits (MDLs). Professional judgement on the part of the analyst is used in application of the QA guidelines, especially in cases where the analyte levels are near the limit of detection.

Acceptable limits of accuracy are ±30% of known certified concentrations that are at least 10 times above the limit of detection of an analyte. The certified values and uncertainties found in the NIST Certificate of Analysis for SRMs describe, statistically, the range in which there is a 95% probability the true value is found. The ±30% range should therefore be calculated as 30% above and below the uncertainty bounds listed in the NIST Certificate of Analysis.

A minimum of 8% of an analytical sample string should consist of blanks, reference or control materials, duplicates, and spike matrix samples. The use of control materials does not entirely replace the use of duplicates and spiked matrix samples. A minimum of 2% of the standard inorganic sample string should consist of calibration materials and reference or control materials.

Because there are no quality assurance targets specified for inorganic elements (e.g. metals) in sediment in the NOAA NS&T programme, no direct comparisons between the NOAA and Auckland Council sediment data quality can be made. However, if the NOAA inorganic element QA targets are the same as the organic compound targets, it would appear that the Auckland Council data would probably comply with NOAA QA guidelines. The PAH SRM data acquired in the SoE programme, while limited, met the NOAA accuracy target.

2.9 Data accuracy and variability summary

The analysis of data variability has shown that:

- Data variability was similar across the RDP, SoE, and UWH programmes, suggesting that differences in sampling, sample processing, and analysis procedures has had no major effect on variability. No major changes in variability have occurred over time, indicating that changes in monitoring and analytical practices have also not greatly affected variability.
- Agreement between blind duplicates has usually been good, but occasionally large differences between individual results have been recorded. Analysis of replicates is therefore required to ensure robust data are obtained.
- CRM results for total recoverable metals and PAH have been good, with nearly all individual results being within 20% of the certified reference values. Concentrations of these primary contaminants are therefore likely to be reasonably accurate.
- Trends in CRM results have been up to ±0.6% per annum. Trends of less than ±1% per year could therefore be significantly affected by analytical variation over time.
- Data quality appears to be similar to that required in three US monitoring programmes. The Auckland Council monitoring data would probably meet the quality objectives set for these programmes, except for agreement between duplicates (precision). Reducing the frequency of high variability results would be required to meet these US measurement quality objectives.
- Data variability in the Auckland Council programmes appears to be higher than that obtained in the GWRC programme. Further investigations are recommended to establish why this may be.
- A first look at estimating sample numbers required for trend assessment indicates a need to reduce data variability in order to have a reasonable chance of detecting even moderate sized trends, with a reasonable level of statistical certainty, within useful time frames. Data users need to be involved in defining their needs in terms of key variables: size of trends that are important for meeting resource management objectives, time frames required for trend detection/reporting, and budgets available for monitoring.

3.0 Archived sample reanalysis

3.1 Background

In 2008, an assessment of temporal trends in metals (Cu, Pb, and Zn) and PAH was undertaken, focusing on the former SoE programme monitoring between 1998 and 2007. This analysis, which was not published, revealed unexpected trend profiles for extractable metals (2 M HCl extraction, <63 µm fraction).

Data for Cu and Zn commonly showed a plateau-shaped profile, in which concentrations increased over time (generally between 1998 and 2003), then levelled off or, in some cases, decreased from 2003 to 2007 (e.g. Hellyers Creek, Figure 3-1).

Around 13 of the 23 SoE sites included in the analysis (beach sites were excluded because of incomplete data sets) showed this pattern, and hence the overall regional trend pattern also followed this general form. Sites with this profile included Ann's Creek, Lucas Upper, Meola Inner, Motions, and Pakuranga Upper.

In addition, some sites, including Big Muddy Creek, and possibly Weiti River, had unusually elevated concentrations in 2007 compared with previous years (see Big Muddy plot in Figure 3-1). The result for Big Muddy was considered unusual because Big Muddy is a rural, reference catchment site, where such increases would not be expected.

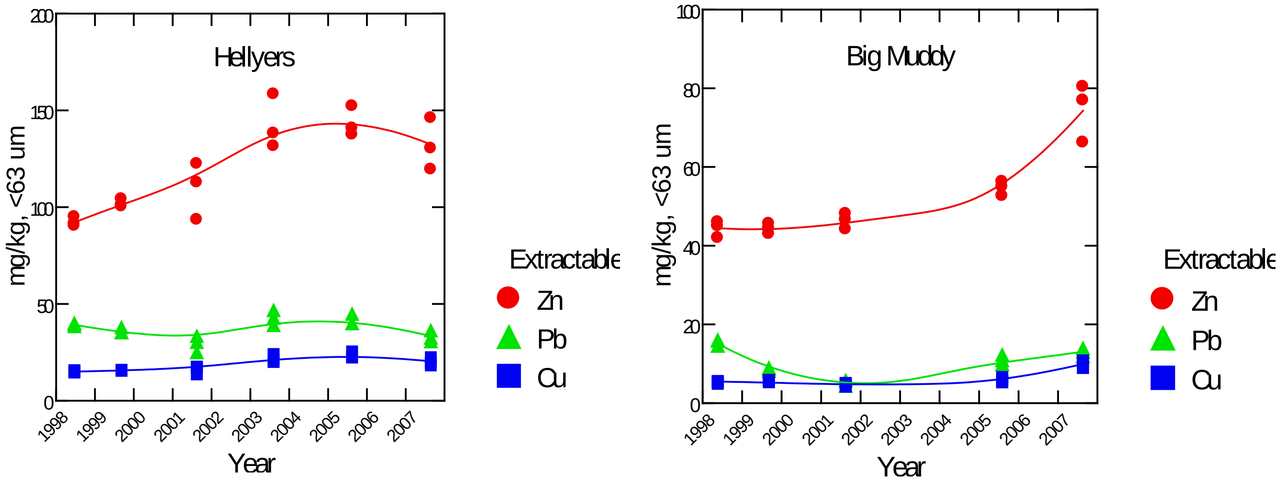


Figure 3-1 Examples of trend profiles from trend assessments at Hellyers and Big Muddy SoE monitoring sites for the 1998–2007 period.
Note the plateau profile for Hellyers, and the high 2007 concentrations at Big Muddy.

In addition to these unexpected trend profiles, results for extractable Pb concentrations, showed decreasing concentrations from 1998 to 2001, a significant increase between 2001 and 2003, then a decrease again to 2007. Cu and Zn also followed the same pattern from 2003 onwards, which raised questions about the reasons for the trend patterns – were these "real", or were they the result of artefacts associated with variability in sampling and analysis?

The plateau trend profile (levelling off and/or decreasing concentrations after 2003) was considered somewhat at odds with the prevailing view of on-going accumulation of contaminants in estuarine sediments resulting from urban runoff, and with the earlier trends observed between 1998 and 2003. Exploration of possible reasons to account for this unusual trend profile was considered warranted before publishing the results of the trend assessments.

Possible explanations for the plateau trend profiles were proposed, including:

- declining loads of Cu and Zn from many urban catchments, which after mixing with existing estuary sediments would result in a gradual slowing of metal accumulation, and eventually a decrease in sediment concentrations;
- mixing of site sediments with less contaminated estuary sediment. Possible sources of less contaminated sediment included:
    - clean catchment soil (e.g. sub-soils) – it would be unusual for significant quantities to be exposed and lost in runoff from established urban catchments, so this should be a relatively minor influence (especially in the older, fully developed, catchments);
    - deeper sediments from within the estuary, which are often less contaminated than the surface sediments being sampled in the SoE and RDP programmes. Deeper sediments could be mixed into the surface layer by physical disturbance of the monitoring sites (a concern when sites are intensively sampled) or possibly by large erosional events (e.g. storms and floods) followed by redeposition of sediments;
    - cleaner sediments brought into the estuary from outer harbour zones. Our understanding of estuarine sediment transport and deposition processes would suggest that this is unlikely to be significant.
- Analytical variation, which may contribute to differences observed between, and within, years. This source of variability is not necessarily consistent between sites or between years, so may be a significant contributor to the variation in trend profiles observed.

Declining metal loads from at least some urban catchments, in particular those where high proportions of exposed metal roofing are being replaced with newer coated versions, is possible. Modelling of loads from Auckland City to the Middle Waitemata Harbour (Timperley & Reed 2008) predicted marked decreases in Zn (but not necessarily Cu) loads from a number of catchments between around 2000 and 2020. This reflects the balance between major sources, mainly roofs and vehicles, which are predicted to change in relative importance over time as old metallic roofs are replaced with newer coated versions and vehicular use increases.

This would not, however, explain the following additional feature shown by the monitoring data. Between 1998 and 2001, the concentrations of Cu and Zn at urban sites generally increased, and Pb levels declined – i.e. Cu and Zn behaved somewhat independently of Pb, reflecting the different sources of these metals. This pattern was consistent with expectations given our understanding of key urban sources of these metals. However, from 2003 onwards, the concentrations of Zn, Cu, and Pb appeared to follow very similar patterns at many sites. This was somewhat surprising, and contrary to the 1998–2001 period. Possible explanations could include:

- Cu, Zn and Pb are now being generated within, and exported from, catchments in very similar proportions, so that concentrations in the receiving environment sediments covary.

This seems unlikely to occur at so many locations, and would be inconsistent with model predictions; or

- internal site (or sediment) factors, rather than external influences (e.g. catchment inputs) are having the major influence on the extractable metals concentrations in the estuary sediments. For example, sediment mixing at the sampling sites during sampling could cause this kind of variation. If deep mixing upon site disturbance was a significant factor, sandy sites would not be expected to show this phenomenon, as these firmer sites are disturbed very little during sampling. Unfortunately, only one sandy site had a reliable trend data record, and this site (Hobson Bay, Newmarket) showed variable metals concentrations over time. However, a limited comparison of sampling methods, which took samples from the sampling lanes (the normal “blueprint” method) and from locations adjacent to ecological sample cores (taken at 10 random locations across each site), showed similar metals concentrations were obtained by the two approaches (see section 5.2.1). This suggests that site disturbance from repeated samplings may not be a major contributor to the variations in the trend monitoring results, however these sites had only previously been sampled once or twice.
- Analytical variation or issues with sample processing and extraction

On balance, the most plausible and readily assessable factors contributing to the variation in trend profiles observed for extractable metals were considered to be analytical variation or issues with sample processing and extraction, especially for the unusually high 2007 results observed at Big Muddy. Therefore reanalysis of archived samples was undertaken at a selection of sites.

3.2 Reassessment approach

Archived (freeze-dried, <500 µm sieved) samples from five sites in the SoE sediment contaminant programme were analysed for Cu, Pb, and Zn to assess whether laboratory analysis was, at least partially, responsible for the plateau trend profiles and the unusually high results in 2007 at rural reference/low urban development sites.

The five sites were chosen based on the nature of the trends observed in the original data:

- Big Muddy showed a large increase in concentrations in 2007, which was unexpected given it is a rural/bush catchment reference site;
- Weiti because the increase observed in 2007 was unexpectedly large given its low degree of urbanisation at the time; and
- Hellyers, Pakuranga Upper, and Whau Upper because they exhibited the plateau profile.

The availability of sufficient archived sample for reanalysis was also a factor in site selection.

Three replicates of extractable metals (<63 um) and total recoverable metals (<500 µm) from each year were analysed. Samples were analysed at R.J. Hill Laboratories.

Note that only one replicate for total recoverable metals was originally analysed each year in the SoE programme (before 2009). Total metals were included in the programme for status, rather than trends, assessment and therefore at that time, no measure of precision was considered necessary. Three replicates for total recoverable metals were analysed in the reanalysis of archived material to assess variability and to permit trend comparisons with the <63 µm extractable metals.

Samples from 1998, 1999, 2001, 2005, and 2007 were analysed. Samples taken in 2003 had not been archived, and were therefore not available for reanalysis – this was unfortunate because the peaking of extractable metals concentrations occurred at many sites around 2003.

Five blind duplicates (one from each site) were included in the reanalysis to check analytical variability.

3.3 Results

3.3.1 Blind duplicate agreement

The results from the analysis of five blind duplicates included with the samples are summarised in Table 3-1. These results showed that agreement between duplicates was generally good, with most RPDs <10%. Exceptions were observed for total Cu at Weiti, and extractable Cu, Pb, and Zn at Whau Upper - these showed 15–20% differences between duplicate results.

Table 3-1 Results for within-batch blind duplicates included with the sediment reanalysis batch. Data are concentrations of total recoverable metals (<500 µm fraction) and extractable metals (<63 µm fraction) in mg/kg for each replicate (Rep) and blind duplicate (Dup). Colour coding reflects relative percentage differences (RPD, %) between duplicate results: Green <10%, Amber 10–20%

		Total Recoverable			Extractable		
Site (Year)	Rep	Cu	Pb	Zn	Cu	Pb	Zn
Big Muddy (1998)	3	8.3	10.5	50.2	7.3	11.2	48.6
	Dup	8.0	10.2	49.0	7.6	10.6	48.4
RPD (%)		3.7	2.9	2.4	3.1	5.5	0.4
Hellyers (1999)	1	15.5	28.3	88.9	19.1	37.0	102
	Dup	15.3	28.9	87.3	19.6	38.4	107
RPD (%)		1.3	2.1	1.8	2.6	3.7	4.8
Pakuranga Upper (2001)	3	27.6	32.5	164.0	34.0	44.5	205
	Dup	26.0	31.5	155.0	34.1	47.4	210
RPD (%)		6.0	3.1	5.6	0.3	6.3	2.4
Weiti (2007)	2	14.7	10.9	57.3	16.1	12.6	64.4
	Dup	17.0	11.9	60.9	16.4	12.5	61.9
RPD (%)		14.5	8.8	6.1	1.8	0.8	4.0
Whau Upper (2007)	2	39.7	90.9	280	37.4	86.1	289
	Dup	38.4	88.0	261	45.7	105.0	344
RPD (%)		3.3	3.2	7.0	20.0	19.8	17.4

The more variable Whau Upper data were consistent with generally high variability found for monitoring results from this site. This suggests a variable sediment matrix is present at this site, which affects the weak acid extractable metals fraction, but not the total recoverable (stronger acid digestion) fraction.

Overall, the duplicate results indicate that the analytical results had good precision, with the probable exception of extractable metals in the Whau Upper samples.

3.3.2 Comparison of trend profiles between original and repeat analyses

Comparison of the original data series with that obtained from the repeated analyses is shown in Figure 3-2 to Figure 3-4 (extractable metals) and Figure 3-5 to Figure 3-7 (total recoverable metals). Line plots (rather than scatter plots with lines of best fit) have been used to more clearly compare the two data sets, with standard error in means used to show variability. Note that the 2003 data from the original data series have not been included, to provide a direct comparison between the two data sets (the 2003 samples were not archived, and therefore could not be included in the reanalyses).

The original 1998–2001 extractable metals data and the repeat analyses from those years' samples generally showed reasonably good agreement. This was the first time archived material had been reanalysed, so it was reassuring that the archived material could be analysed after 10 years storage and give results that were generally comparable with those originally obtained. This result was also satisfactory because two different labs analysed the original (NIWA) and repeat (Hills) samples. Exceptions were observed for extractable Pb at Weiti and Big Muddy (where 2001 repeat results were lower than the original), and extractable Zn at Whau Upper (2001 repeat results were lower than the original).

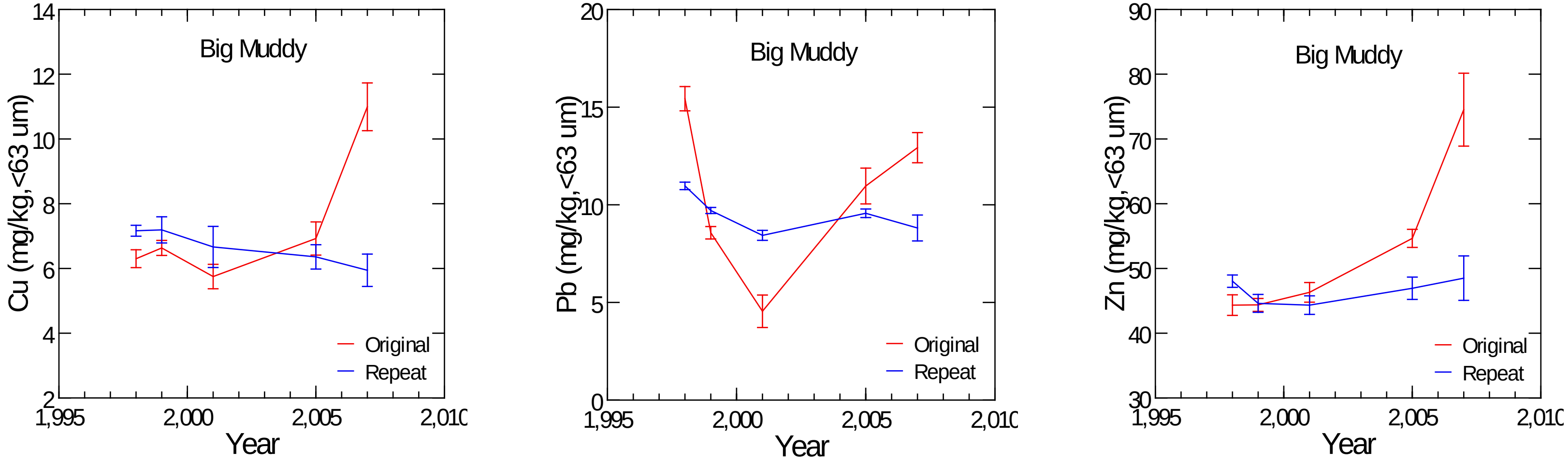
Most of the unusual and/or high results obtained in the original data after 2001 were not confirmed by the repeat analyses – e.g. elevated Cu and Zn at Big Muddy and Weiti in 2007. Extractable Pb results in the reanalysed sample set were generally less variable than in the original data.

Total Cu and Zn results were generally comparable between the two data sets at all sites, given that 3 replicates were analysed in the repeats, compared with only a single sample in the original data set. Total Pb seemed more consistent in the repeat analyses (e.g. see Big Muddy, Hellyers and Weiti results).

Some results for Whau Upper showed high variability in both the original & repeat data (and in the blind duplicate results). This suggests a variable sample matrix-related effect at this site.

Overall, the repeat analyses provided a generally more consistent data set. This was to be expected because all samples were analysed in the same batch, rather than being spread out over five batches spanning nine years. The repeat data showed smaller changes over time at urban sites, less change at the non-urban reference site (Big Muddy), and more consistent decreasing Pb levels. The markedly high results obtained in the original data set were not reproduced in the repeat analyses, supporting a hypothesis that analytical variation between batches was a key factor contributing to the plateau trend profiles. Greater confidence in these findings would have resulted if reanalyses included the 2003 samples.

Big Muddy



Hellyers

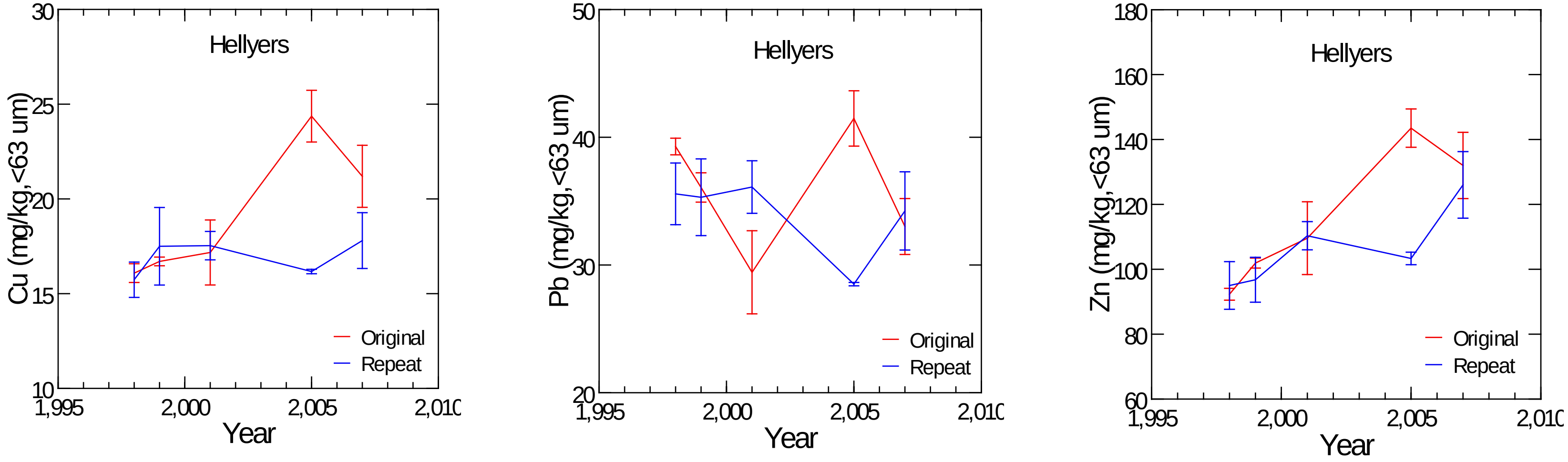


Figure 3-2 **Original** (——) and **repeat** (——) analysis data from Big Muddy and Hellyers Creeks. Extractable Cu, Pb, and Zn (2 M HCl, <63 μm fraction). Values plotted are means (n=3 replicates) ± S.E. in means.

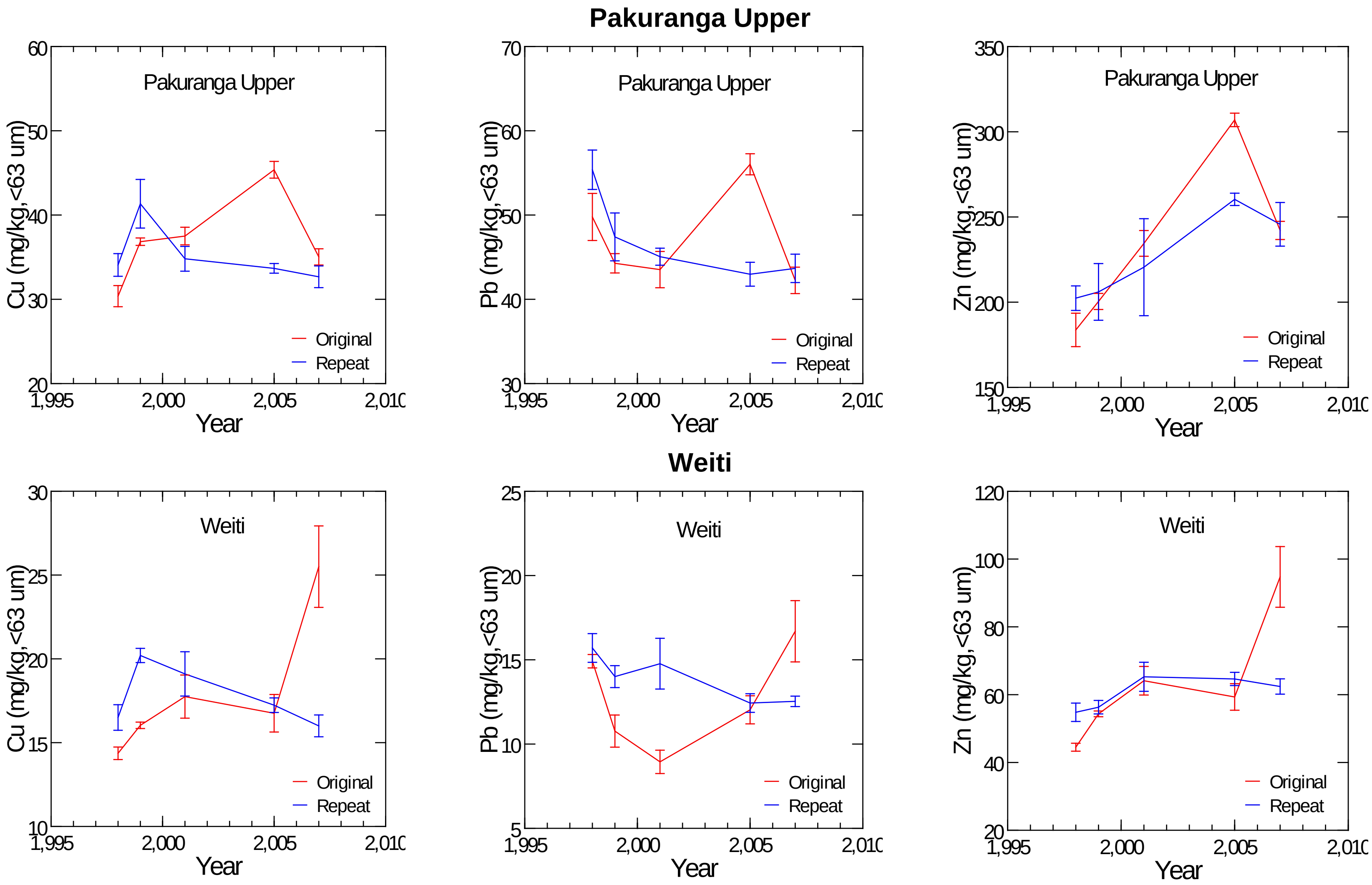


Figure 3-3. **Original** (——) and **repeat** (——) analysis data from Pakuranga Upper and Weiti.
Extractable Cu, Pb, and Zn (2 M HCl, <63 µm fraction). Values plotted are means (n=3 replicates) ± S.E. in means.

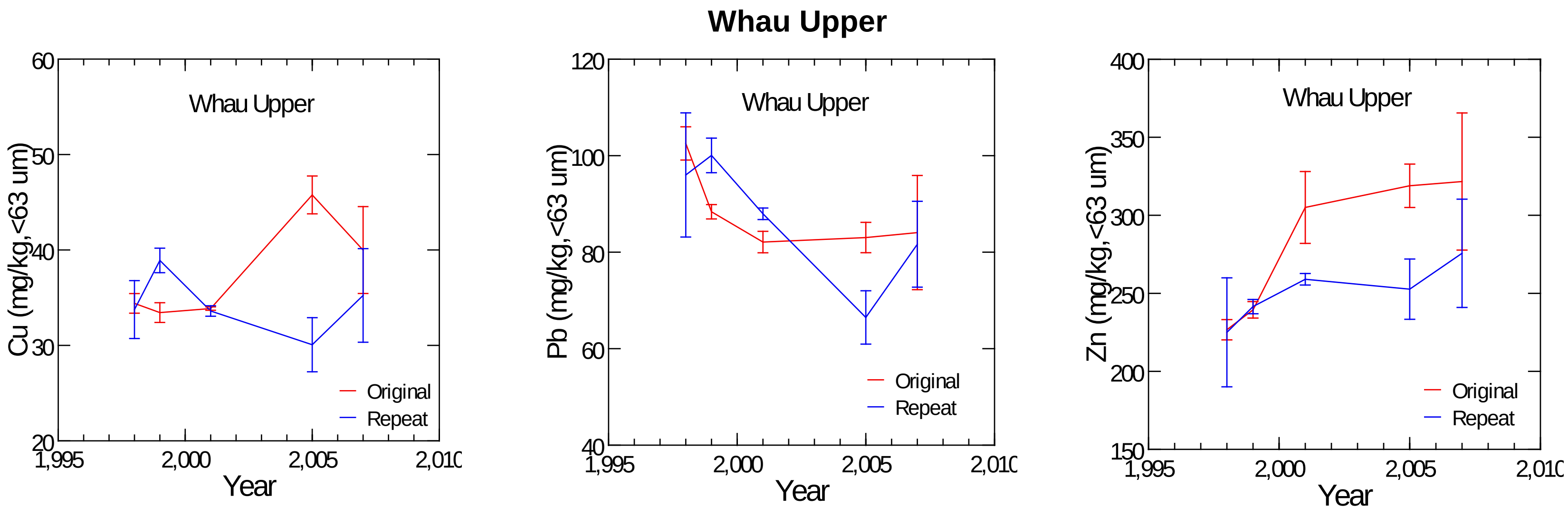


Figure 3-4 **Original** (——) and **repeat** (——) analysis data from Whau Upper.
Extractable Cu, Pb, and Zn (2 M HCl, <63 µm fraction). Values plotted are means (n=3 replicates) ± S.E. in means.

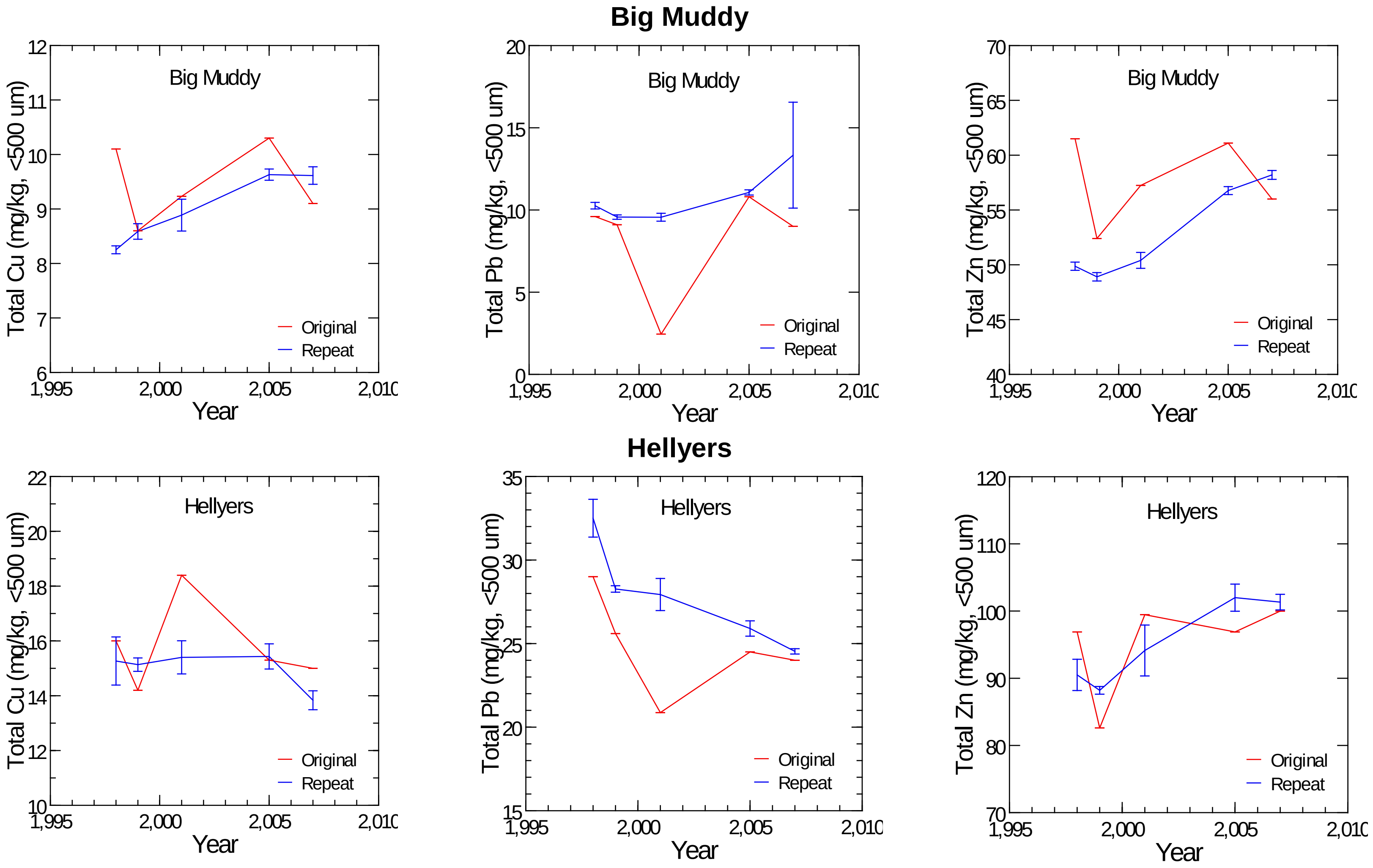
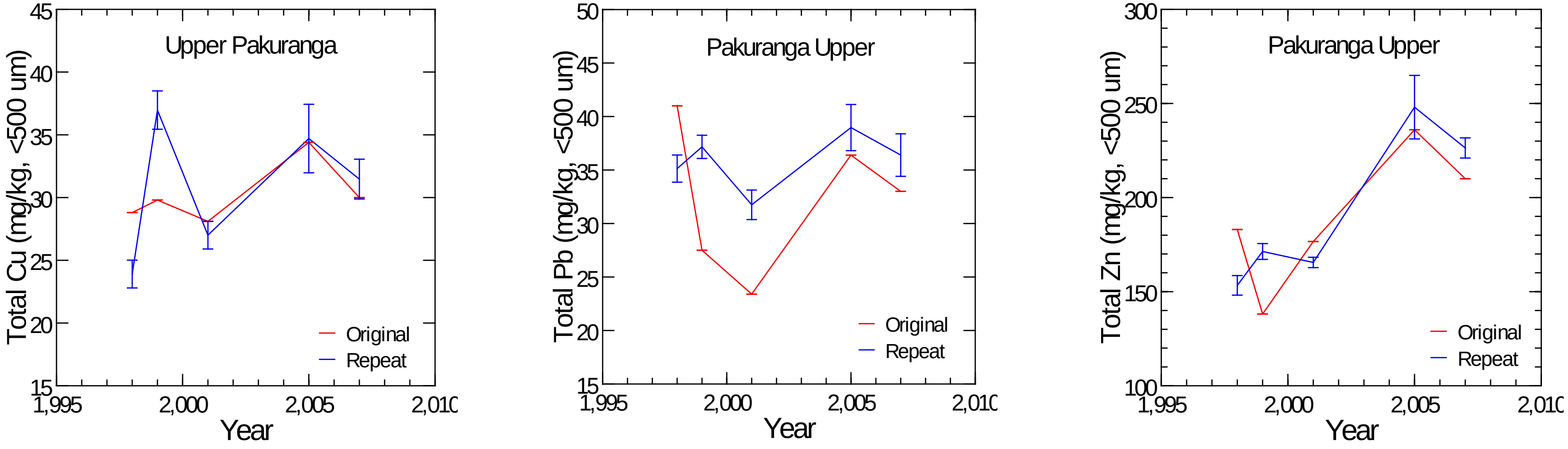
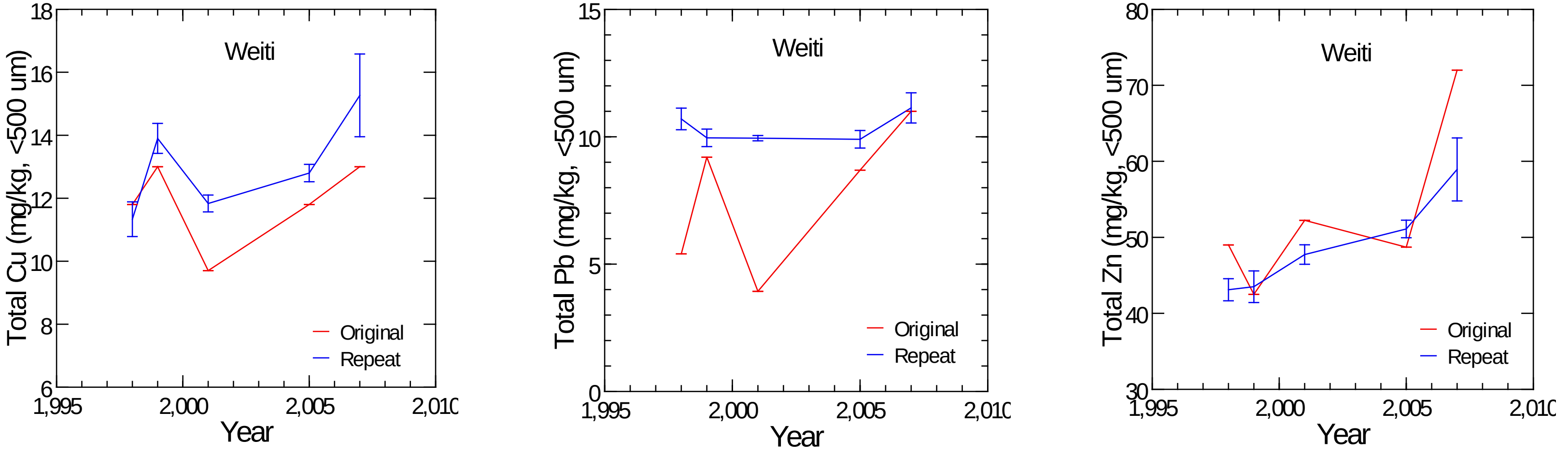


Figure 3-5 **Original** (——) and **repeat** (——) analysis data from Big Muddy and Hellyers Creeks.

Total Recoverable Cu, Pb, and Zn (<500 µm fraction). Values plotted are means (n=3 replicates) ± S.E. in means for the repeats, and a single result for the original data.



Pakuranga Upper



Weiti

Figure 3-6 **Original** (——) and **repeat** (——) analysis data from Pakuranga Upper and Weiti.

Total Recoverable Cu, Pb, and Zn (<500 μm fraction). Values plotted are means (n=3 replicates) ± S.E. in means for the repeats, and a single result for the original data.

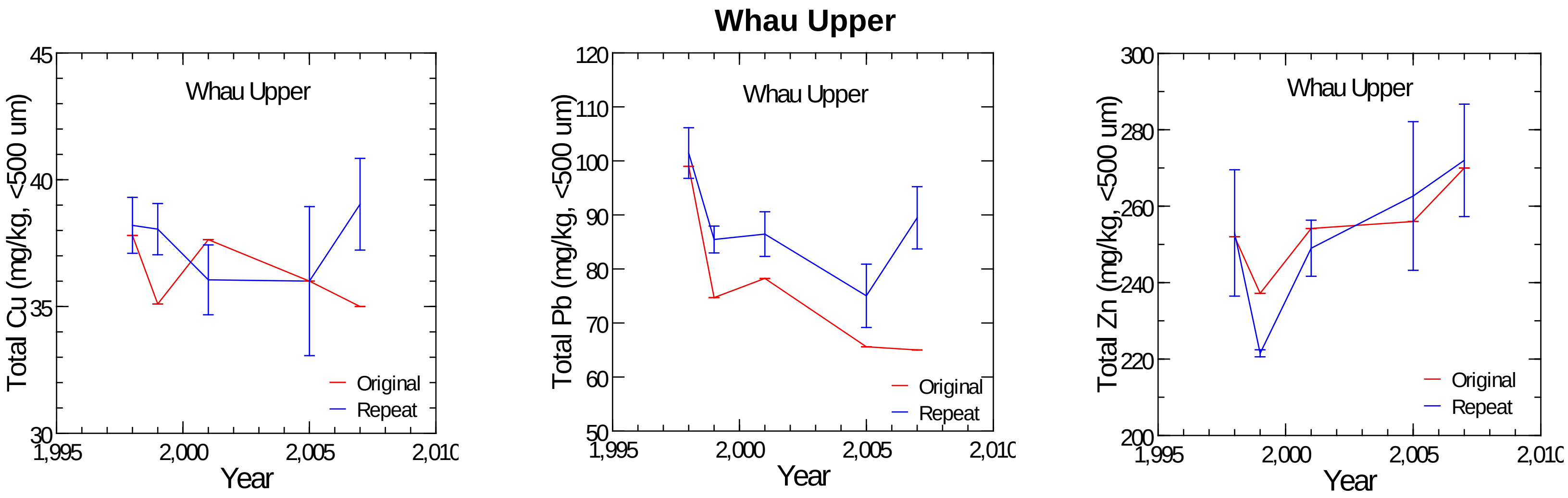


Figure 3-7 **Original** (——) and **repeat** (——) analysis data from Whau Upper.

Total Recoverable Cu, Pb, and Zn (<500 µm fraction). Values plotted are means (n=3 replicates) ± S.E. in means for the repeats, and a single result for the original data.

3.3.3 Comparison of trends obtained by original and repeat analyses

The impact of sample reanalysis on the magnitude and significance of temporal trends obtained from the 1998–2007 period was assessed by conducting Mann Kendal trend tests (undertaken with Time Trends software, version 3.31, 2012; Jowett Consulting and NIWA) on the original and repeat analysis results. A comparison of the trend results is shown in Table 3-2 and the results shown graphically in Figure 3-8 and Figure 3-9.

Table 3-2 Trends (mg/kg/year) between 1998 and 2007 in A) extractable metals and B) total recoverable metals from original and repeat analyses of SoE sediments from five sites.
Red values are statistically significant (Mann Kendall test, p<0.05).

A. Extractable metals (<63 um)

	Cu		Pb		Zn	
Site	Original	Repeat	Original	Repeat	Original	Repeat
Big Muddy	0.26	-0.14	-0.11	-0.19	2.04	0.02
Hellyers	0.87	0.10	-0.21	-0.40	6.13	2.58
Pakuranga Upper	1.45	-0.34	0.05	-1.10	15.26	6.25
Weiti	1.01	-0.21	0.40	-0.33	4.81	0.98
Whau Upper	0.73	-0.28	-1.44	-3.17	11.89	5.12

B. Total Recoverable metals (<500 um)

	Cu		Pb		Zn	
Site	Original	Repeat	Original	Repeat	Original	Repeat
Big Muddy	0.00	0.16	-0.04	0.14	-0.13	0.99
Hellyers	-0.11	-0.12	-0.25	-0.68	0.81	1.42
Pakuranga Upper	0.32	0.72	-0.47	0.18	6.91	8.84
Weiti	0.04	0.28	0.46	0.01	1.05	1.44
Whau Upper	-0.25	-0.05	-3.20	-0.86	2.65	3.57

Trends from the extractable metals results differed between the two data sets:

- Cu trends were small increases in the original data, but reduced to very small decreases at four of the five sites and a very small increase at one site (although most trends were not statistically significant);
- Pb showed mostly significant decreases over time at four of the five sites in the repeated data set, whereas in the original data significant trends were found at only one site (Whau Upper); and
- Zn showed significant increases over time at all sites (including at the Big Muddy reference site) in the original data set. The trends in the repeat data were all substantially smaller than those from the original data, and Big Muddy (rural reference site), Weiti, and Whau Upper showed no significant trends.

Trends from the total recoverable metals were generally more comparable than those for the extractable metals. The original data trends were all not statistically significant, probably because only a single sample was analysed each year. This limits detailed comparisons between the two data sets for total recoverable metals.

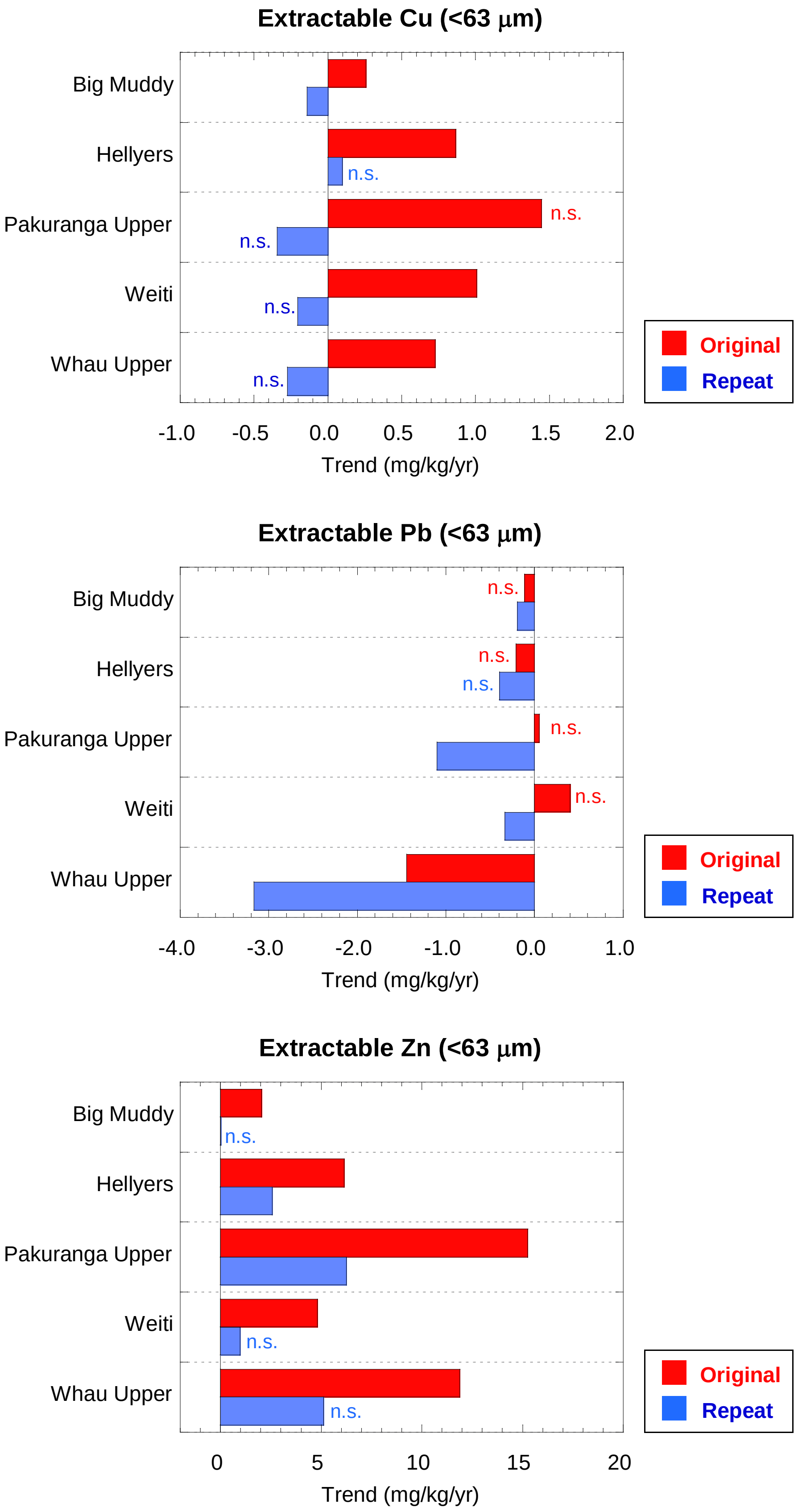


Figure 3-8 Trends (mg/kg/year) in extractable metals between 1998 and 2007 from original and repeat analyses of sediments from five SoE programme monitoring sites.
n.s. indicates not significant (Mann Kendall test, $p<0.05$).

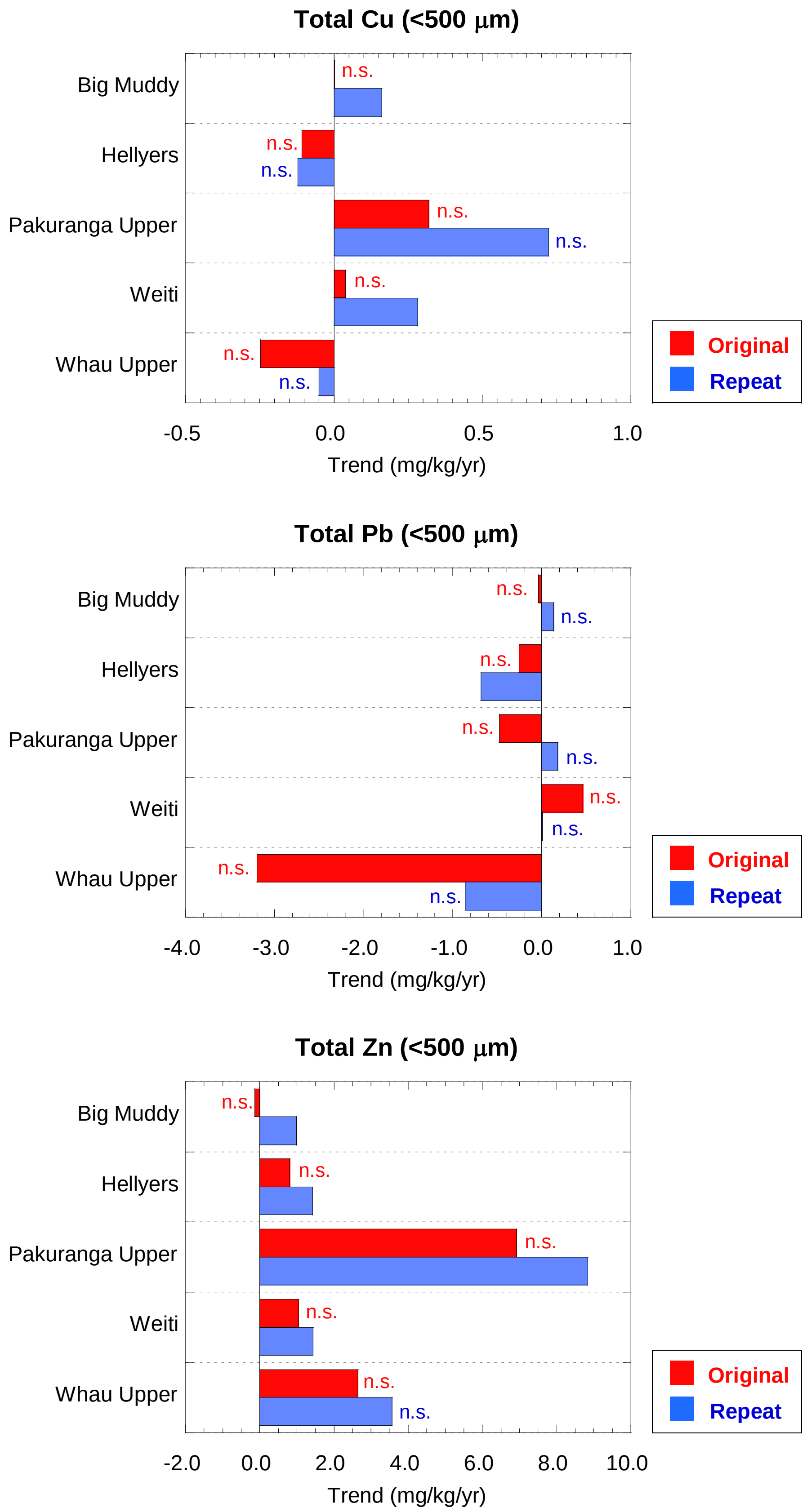


Figure 3-9 Trends (mg/kg/year) in total recoverable metals between 1998 and 2007 from original and repeat analyses of sediments from five SoE programme monitoring sites.
n.s. indicates not significant (Mann Kendall test, p<0.05).

3.3.4 Conclusions from archived sample reanalysis

Reanalysis of archived samples showed that analytical variation, particularly for extractable metals data from 2005 and 2007, had a marked effect on year-to-year changes in metals concentrations, and on trends measured from the data.

The high concentrations responsible for the apparent outliers originally observed at Big Muddy and Weiti in 2007 were not reproduced in the reanalysis, strongly supporting the hypothesis that these were analytical artefacts in the original data.

The plateau trend profiles evident in the original data series for extractable metals (and tested by reanalysis of samples from Hellyers, Pakuranga Upper and Whau Upper) were generally not reproduced in the reanalysis. However, the lack of samples from 2003 somewhat limited the conclusiveness of the comparisons.

Trends obtained from the original and repeat analyses were different for extractable metals. Trends in Cu and Zn were smaller than measured from the original data, and stronger (and more consistent) for Pb.

Overall, the reanalysis highlighted the impact that year-to-year between batch analytical variation could make on assessment of temporal trends.

The data also supported, but did not conclusively prove, the hypothesis that the plateau trend profiles for extractable metals analysis observed as a result of elevated concentrations between 2003 and 2007 were likely to be an analytical artefact rather than a real phenomenon. Because of this, and some indication of possible issues with the weak acid extraction process between 2003 and 2007, assessment of temporal trends may require excluding these results from trend analysis. This was the approach adopted in the status and trends assessment undertaken for monitoring conducted between 1998 and 2010 (Mills et al. 2012).

The effect of occasional unusual data or outliers will reduce over time as more monitoring results are accumulated. However, minimising the number of such data (by quality assurance checks) is still required to maximise the reliability of the trend monitoring.

4.0 Particle size distribution

4.1 Background

Sediment particle size distribution (PSD) is an important characteristic influencing benthic ecology and sediment chemistry. Many contaminants tend to accumulate to a greater extent in the fine particle size fractions. Comparisons of contaminant concentrations between sites, or assessment of changes in concentrations over time, need to take these differences in PSD into account. Benthic ecology is also influenced by sediment PSD, with benthic communities varying in response to changes in sediment texture (Hewitt et al. 2012).

Monitoring PSD is therefore an important component of the RSCMP. The methods used for PSD need to be reliable, and the data produced in both the sediment contaminant and ecological monitoring programmes needs to be consistent, so that contaminant–ecology effects can be investigated.

The particle size indicator of most interest for routine sediment contaminant monitoring is sediment muddiness – the proportion (%) of the sediment in the <63 µm range (silt plus clay fractions). An increase in sediment muddiness is a potential impact of the early stages of catchment urbanisation or land use intensification and therefore change in sediment mud fraction content is a key environmental indicator of land use change.

4.2 Particle size distribution analysis methods

Particle size analysis has been conducted by a variety of methods in the (former) SoE and RDP programmes, in ecological monitoring programmes (e.g. the Upper Waitemata Harbour – UWH – programme), and more recently in the RSCMP. These include:

- Wet sieving (3 size ranges; >2000 µm, 63–2000 µm, <63 µm) and weighing the dried fractions – this method was used in the RDP in 2002 (KML 2003). Analyses were conducted by R.J. Hill Laboratories;
- Laser particle size analyser (Malvern Mastersizer; Earth Sciences Department, Waikato University) – used by URS in the RDP in 2002 (URS 2002). Data were reported as % volumes in each of eight particle size classes ranging between 0.05 and 3500 µm;
- Laser particle size analyser (Galai instrument) – used in the SoE sediment contaminant programme from 1998 to 2007, and in the RDP for sediment contaminant PSD analysis from 2004 to 2008 (Diffuse Sources 2005, 2006, 2007, 2008, and 2009). Analyses were conducted by NIWA (Hamilton). The standard protocol was to record particle areas, volumes, and numbers (expressed as % of the total sediment) in eight size classes. Data were acquired in two particle size ranges: 0–300 µm and (for samples with significant amounts of coarse particles of >300 µm) 2–600 µm. Samples were pre-sieved at 500 µm to remove large debris such as shell hash, so effectively the data reported were for the <500 µm range (rather than for the whole sediment).
- For the 2008 RDP samples, NIWA used a replacement laser instrument for the Galai analyser (Ankersmid EyeTech). This provided the same basic information as the Galai

analyser, but the 2–600 µm range was analysed as 10–2000 µm (really 10–500 µm, since samples were 500 µm sieved before analysis).

- Wet sieving/pipette separation into 6 particle size ranges (gravel (>2000 µm), coarse sand (>500–2000 µm), medium sand (>250–500 µm), fine sand (>62.5–250 µm), silt (>3.9–62.5 µm), clay (0–3.9 µm), followed by drying to constant weight at 60°C and weighing (Hailes et al. 2010). This has been the standard protocol used by NIWA for benthic ecology monitoring, and was used in the benthic ecology component of the RDP between 2004 and 2008. It was also adopted for sediment contaminant and ecology monitoring at SoE sites in 2009 (Diffuse Sources 2010) and in the RSCMP in 2010 and 2011 (Diffuse Sources 2011, 2012).

Of these methods, the laser and sieve/pipette methods have been used for the vast majority of the monitoring conducted to date.

4.3 Comparison of the laser and wet sieve/pipette methods

The wet sieve/pipette method is now used across all the Auckland Council RSCMP and benthic ecology programmes. It is therefore useful to assess the relationship between the data generated by this method and the laser method used prior to 2009 in the sediment chemistry monitoring programmes. If there is a consistent (preferably 1:1) relationship between the results generated by the two methods, then we can be more confident about the validity of trends measured over time periods where a combination of both methods were used. Conversely, if the results generated by the two methods are inconsistent (i.e. not well correlated), reliable trends will not be obtained.

4.3.1 Approach used to compare the PSD methods

The particle size indicator of most interest for the routine sediment contaminant monitoring is sediment muddiness – the proportion (%) of the sediment in the <63 µm range (sum of the silt and clay fractions).

The proportion of the sediment in the <63 µm fraction was calculated from each PSD analysis method as the sum of the:

- 0–3.9, 3-9–7.8, 7.8–15.6, 15.6–31.3, and 31.3–62.5 µm data from the laser analyser; and
- silt (3.9-62.5 µm) and clay (<3.9 µm) fractions from the sieve/pipette method.

The proportion <63 µm from each method was expressed as a % of the total sediment (with a minor modification, as outlined below).

The Galai laser analyser recorded particle areas, volumes, and numbers in 8 size classes. Data were acquired in two particle size ranges: 0–300 µm for muddy samples, and also at 2–600 µm for coarser textured samples with larger sized particles, >300 µm. Because samples were pre-sieved at 500 µm to remove large debris such as shell hash, effectively the data were assessed for the <500 µm range (rather than for the whole sediment).

For the 2008 RDP samples, NIWA used an Ankersmid EyeTech laser instrument. This provided the same basic information as the Galai analyser, but the range reported for coarser textured samples was 10–2000 µm (really 10–500 µm – since samples were sieved before analysis – rather than the 2–600 µm range formerly reported).

Because the sediments analysed by laser were pre-sieved at 500 µm, the proportions of the sediment in the <63 µm fraction determined from the sieve/pipette method were also calculated as a proportion of the <500 µm fraction (rather than the total sediment). This was usually very similar to the whole sediment, but it may differ for coarser (shelly or gravelly) sediments.

The data used for the comparisons of the laser and sieving methods were taken from RDP sites that had both sediment contaminant and benthic ecology samples taken on the same sampling occasions. These sites therefore had PSD measured by both laser (for the sediment contaminant component) and sieve/pipette (for the benthic ecology component). Samples used in the comparison were from RDP programme monitoring conducted in 2004 (8 sites), 2005 (4 sites), 2006 (18 sites), 2007 (5 sites), and 2008 (18 sites).

Note that while the samples used in the PSD analysis method comparison were taken from the same sites on the same dates, the analyses were not conducted on exactly the same samples – the benthic ecology PSD samples were composites made up from 10 sub-samples taken from adjacent to each ecology core (where cores are spread across the monitoring site), while the contaminant PSD samples were composites made up from 10 sub-samples taken from along the two contaminant sampling lanes (see ARC 2004 for contaminant sampling protocols). Unless the sampling sites have significant spatial gradients in sediment texture (such that the two sampling lanes do not adequately represent the texture across the whole site), the two types of PSD sample should be very similar. This was verified in a separate correlative investigation (see section 4.4.2).

4.3.2 Relationships between laser and sieve/pipette method PSD results

The Galai analyser reported particle areas, volumes, and numbers. The area data correlated best with the sieve/pipette method (% by weight) data, giving a reasonable linear fit across the full range of sample mud content encountered in the RDP sediments. The relationship between the Galai particle area data and the sieving method data is shown in Figure 4-1.

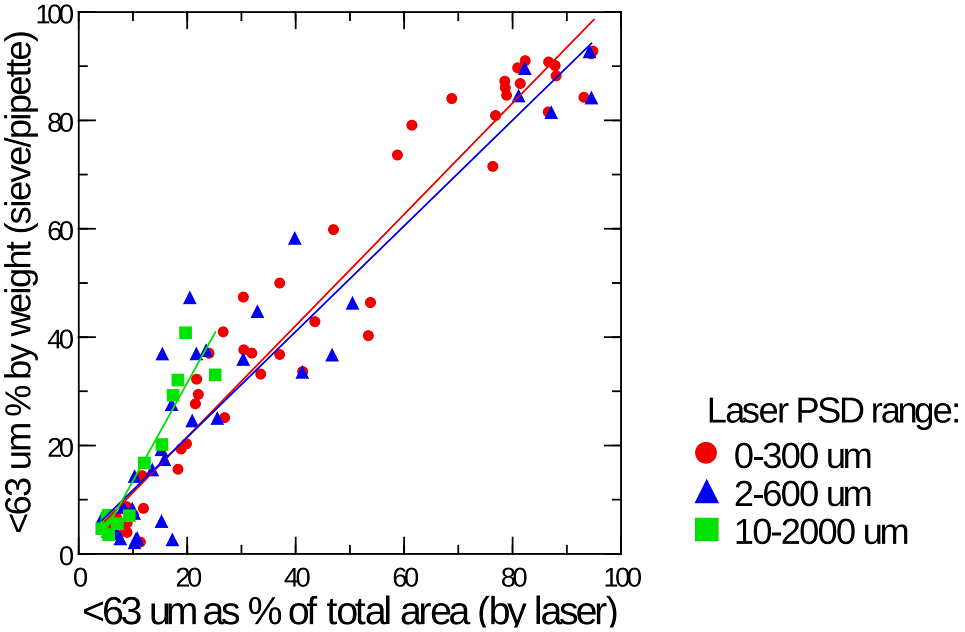


Figure 4-1 Relationship between sediment mud fraction content (% <63 µm) obtained by laser analysis (particle area data) and by weight (using the sieve/pipette method).

The slopes of the linear regressions of laser areas and % by weight for the 0–300 µm and 2–600 µm particle size ranges were similar (slopes of 1.03 and 0.98 respectively) and indicated approximately 1:1 relationships between the laser areas and the % by weight data, while that for the 10–2000 µm range was markedly different (slope = 1.81). However, this difference in slope may simply reflect the more limited sample numbers and relatively lower range of muddiness in the samples analysed in the 10–2000 µm range in 2008. All the linear regressions between % area and % by weight were significant ($p \leq 0.0001$, n=53, 36, and 12 for the 0–300 µm, 2–600 µm, and 10–2000 µm ranges respectively) and correlations coefficients were reasonably close to 1 (R^2 values of 0.950, 0.883, and 0.870 for the 0–300 µm, 2–600 µm, and 10–2000 µm ranges respectively). However, while these relationships were approximately linear, the scatter was too great for reliable prediction of % by weight from particle area data (or vice versa).

The relationship between the sieve/pipette method results and the Galai volume data were more complex, and were non-linear, at least above approximately 20% of the total volume (Figure 4-2). The relationships between the two laser methods (area and volume) and the % by weight (by the sieve/pipette method) were unexpected. Assuming constant particle density, weight would be expected to be more comparable with volume, rather than particle surface area.

There was no useful relationship between the sieve/pipette % weight data and the laser number data, because the particle numbers in the finest fractions represented over 90% (and for most samples >98%) of the total particle numbers in the sediments.

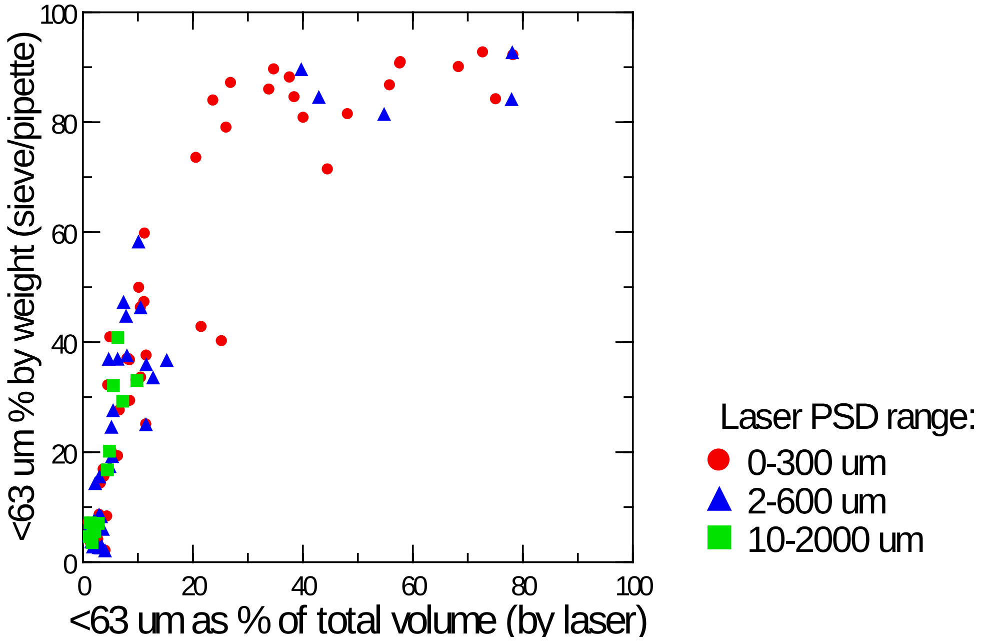


Figure 4-2 Relationship between sediment mud fraction content (% <63 µm) obtained by laser analysis (particle volume data) and by weight (using the sieve/pipette method).

Overall, therefore, there was a reasonably robust linear relationship between laser particle area results and % by weight from sieve/pipette analysis. However, there was too much variation (scatter) in the area versus weight relationship to allow reliable predictions of mud content (as a % by weight) from laser area data on an individual sample basis. The volume versus weight relationship was (somewhat unexpectedly) non-linear over the full range of sediment mud contents found in the samples, and therefore the volume data could not easily be converted into % by weights. This unexpected relationship warrants further investigation. The area results were better correlated with the % weight results, and therefore area data from the laser analysis provide a better option than volume for tracking trends where no % weight data are available. However, the potential differences in results generated by the laser and sieve/pipette methods means that trend data generated from a mixture of these methods are unlikely to be very reliable. It is therefore not recommended that trends be measured from results generated by a mixture of these PSD analysis methods.

4.4 Variability in particle size distribution analysis

The variability in PSD analysis is important to know in order to assess the significance of differences observed between sites and over time (temporal trends) – are the observed differences real, or does analytical variability contribute significantly?

Unfortunately, there has been no systematic assessment of PSD analytical variability in the RDP or SoE programmes that can definitively answer this question. However, information on PSD variability can potentially be obtained from three sources:

- Quality assurance data collected in the SoE programme between 1999 and 2007, where samples were analysed using laser PSD analysis;
- Correlative analysis of data from the RDP and SoE programmes, in which sites sampled for both sediment contaminant and benthic ecology on the same occasion had separate sets of PSD samples analysed by the sieve/pipette method; and
- From the results obtained from sieve/pipette method analysis of Bulk Reference Sediments (BRS), which were introduced as quality assurance samples in the 2011 RSCMP and were also analysed in 2012 (these BRS are to be reanalysed annually).

As outlined below, none of these assessments is definitive, but they do provide useful information on the precision of the two PSD analysis methods.

4.4.1 Quality assurance data from the SoE programme

Quality assurance (QA) assessments were undertaken in the SoE programme using samples from five sites, which were sampled in 1999 and archived for reanalysis with every subsequent SoE monitoring batch (i.e. every two years – in 2001, 2003, 2005, and 2007). The data were updated each sampling occasion, and changes between years assessed. This allowed the changes observed over time in the monitoring programme to be put into context relative to analytical-sourced variations.

Unfortunately, the amount of 1999 sample was insufficient to allow reanalysis after 2001, and therefore later monitoring used samples collected in 2001 for on-going QA assessment.

However, exactly which QA samples (1999 or 2001) were analysed in 2003 is unclear. The 2003 SoE data report (Reed & Webster 2004; Table 9) states *"Distribution of surface area (%) for five samples collected in 1999 and analysed in 1999, 2001 and 2003"*. This indicates the 2003 QA checks were made using the sample taken in 1999.

The 2005 monitoring report (McHugh and Reed 2006; Table 11) gives *"Distribution of surface area (%) for five samples collected in 1999 and 2001 and analysed in 1999, 2001, 2003 and 2005."* The table footnote states *"Note: 2003 and 2005 data is determined on 2001 historical sample due 1999 sample exhausted from previous studies."* This indicates the 2003 (and 2005) QA checks were made using sample collected in 2001.

Further investigation was unable to shed any light on exactly which QA samples were reanalysed in which years, and therefore detailed analysis of the QA data has not been reported here. However, the QA data showed reasonably consistent batch-to-batch (i.e. year-to-year) results, with one major exception – the % area results from 2001 were markedly different (lower) than in other years. An example of the lower % area results for data collected in 2001 is shown in Figure 4-3, which plots the QA data collected from the Middlemore QA sample(s) analysed in 1999, 2001, 2003, 2005, and 2007. The QA data shown are as reported in Reed & Webster (2004) and McHugh and Reed (2006), with 2007 data supplied by NIWA (J. Reed, pers. comm.).

The lower % area results were also observed in the field monitoring results from 2001 at most sites – an example, again for the Middlemore SoE site, is shown in Figure 4-4. They were also evident when % area versus % volume data for samples analysed between 1999 and 2008 were plotted (Figure 4-5).

These strands of evidence indicate that the 2001 laser PSD % area data were somewhat lower than expected based on QA and field monitoring results. The low 2001 results have therefore been flagged in the Single Site Reports (SSRs) produced as part of the Status and Trends monitoring reporting (e.g. Mills et al. 2012) as probably being associated with analysis (rather than being due to “real” environmental changes).

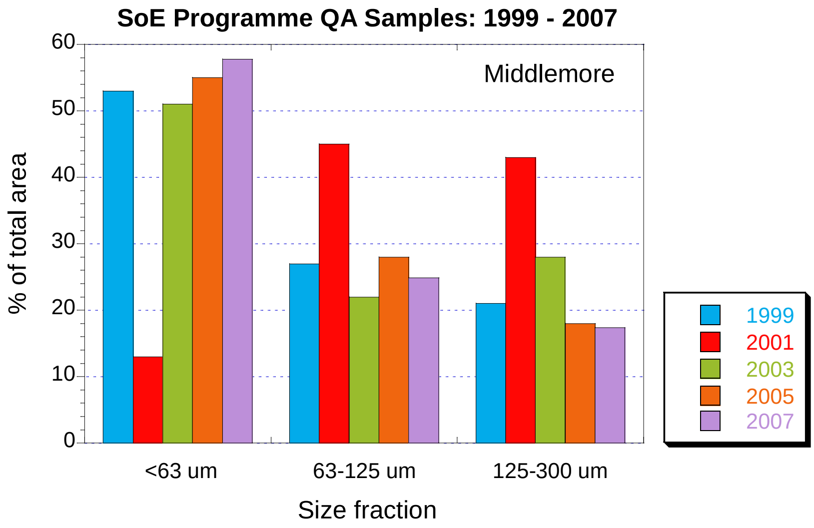


Figure 4-3. Quality assurance results for particle size distribution (PSD) from the former SoE programme. Data are for % area results (grouped into three size fractions) from the Middlemore QA samples. Note the lower results for the <63 µm fraction (and higher proportions of 63–125 µm and 125–300 µm fractions) obtained in 2001.

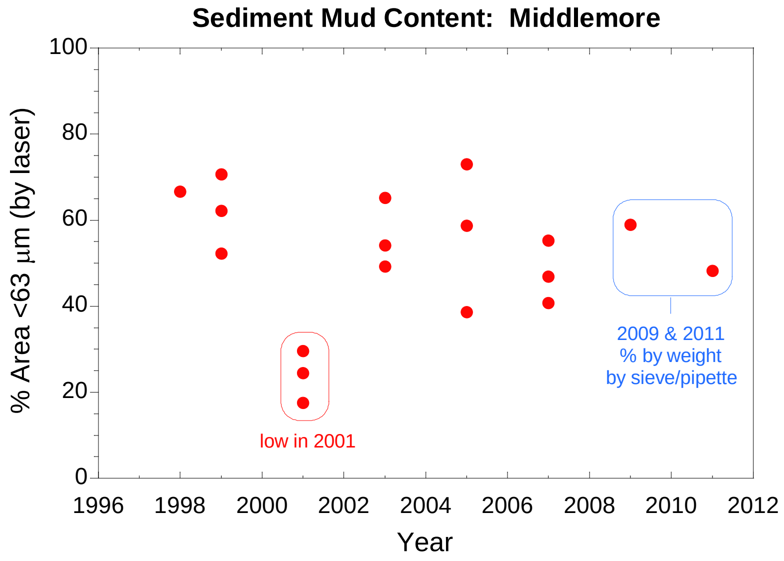


Figure 4-4. Trends in sediment mud content (% <63 µm) at Middlemore SoE monitoring site 1998–2011. Data are % total area (in 0–300 µm range) by laser PSD analyser, except 2009 and 2011 (% by weight from sieve/pipette method). Note low values obtained in 2001. N = 3 replicates per year except 1998, 2009 and 2011 (N = 1).

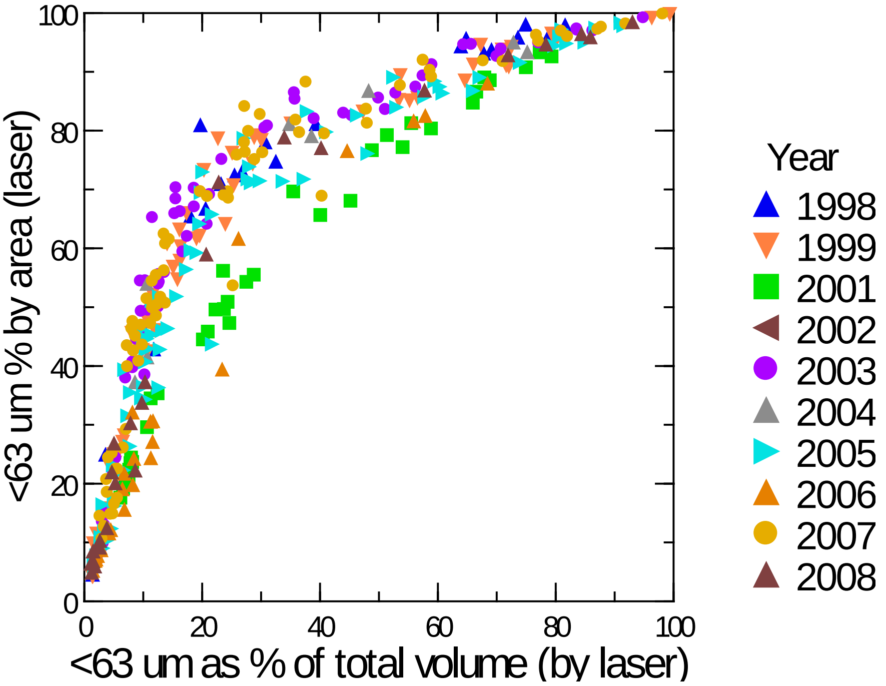


Figure 4-5. % <63 µm, laser area versus volume (0–300 µm range) for samples analysed between 1998 and 2008. Note relatively lower areas for samples analysed in 2001.

4.4.2 Correlation between ecology plot and contaminant lane PSD samples

RDP and SoE programme sites sampled in 2009 and 2010 for both sediment contaminant and benthic ecology on the same occasion had separate PSD samples analysed by NIWA using the sieve/pipette method. This provided two sets of PSD data from each site – one set obtained from a composite sample taken from the 10 ecology core plot locations distributed across the site, and one set obtained from a composite sample prepared from sampling along the two contaminant sampling lanes (as per the sediment contaminant monitoring “blueprint” protocol; ARC 2004).

While these two sets of data were not from exactly the same samples, they did represent very similar “site integrated” samples, which would be expected to have very similar sediment characteristics.

Examination of the relationship between the contaminant lane and ecology core samples showed good correlations, with relatively little scatter, between all fractions except the very fine (clay, <3.9 µm) and coarser (>500 µm, coarse sand and gravel) fractions (Figure 4-6).

The average ecology plot to contaminant lane ratio for the <63 µm fraction was 0.99 (CV = 7.3%, minimum 0.79, maximum 1.15, N = 33).

These data indicated that essentially equivalent sediment mud content (% <63 µm fraction) values could be obtained from composite samples taken from either the contaminant sampling lanes or ecology core locations. They also suggested that the sieve/pipette PSD method provided consistent data for mud content (silt + clay; <63 µm) and for silt (3.9–63 µm), fine sand (63–250 μm), and medium sand (250–500 μm). Greater variability was observed for the clay fraction (<3.9 μm), coarse sand (500–2000 μm), and gravel (>2000 μm) fractions. Data for these fractions are therefore likely to be less quantitatively reliable than the others.

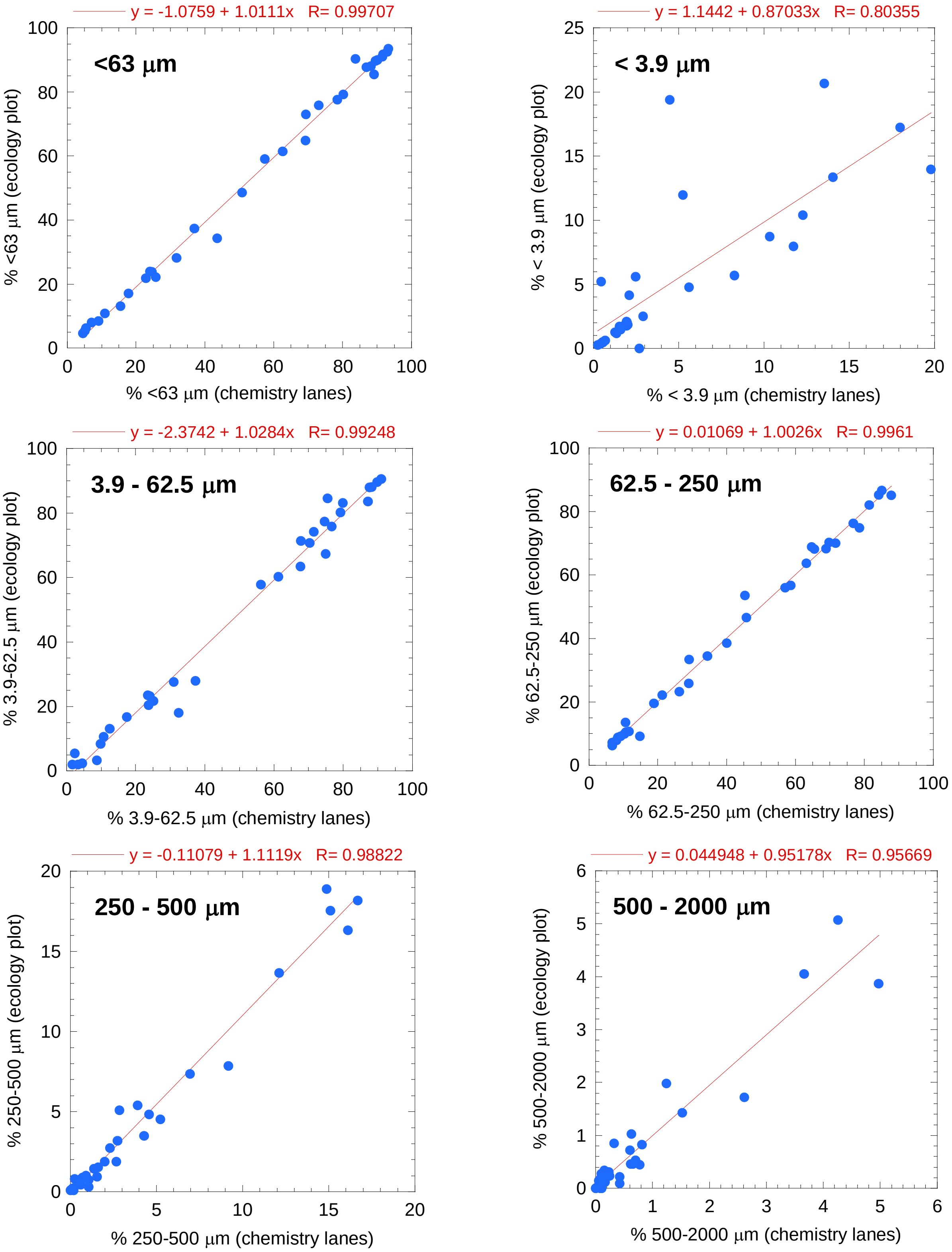


Figure 4-6 Correlations between PSD data (% by weight) obtained from composite samples taken from the benthic ecology core locations and from sediment contaminant sampling lanes at SoE and RDP sites sampled on the same occasions in 2009 and 2010 (N=33).

PSD analysis was by the sieve/pipette method, undertaken by NIWA.

4.4.3 Variability from Bulk Reference Sediment (BRS) analysis

Bulk Reference Sediments (BRS) are large batches of sediments from two Auckland Council estuarine sediment monitoring sites, prepared in 2011 and archived for on-going use in quality assurance. Samples of BRS are to be submitted with each set of monitoring samples, and the contaminant concentrations and PSD results tracked over time to provide benchmarking for temporal trend assessment. They can also be used to assess laboratory performance, for example if changes in laboratories or analytical methods occur.

The BRS samples were prepared from multiple sub-samples taken over areas similar to normal sediment monitoring site plots, and homogenised. The BRS for PSD analysis are stored in frozen form. The results obtained from BRS analyses reflect lab analysis performance for real world samples typical of those collected in the AC monitoring programme – i.e. the variability in the BRS results reflects analytical variation, and the within-sample variability inherent in the estuarine sediments sampled in the Auckland Council programmes.

The two BRS are:

- a muddy sediment taken from adjacent to the SoE monitoring site at Middlemore (upper Tamaki Estuary), and
- a sandy sediment from adjacent to the RDP Meola Outer Zone site (Central Waitemata Harbour).

Particle size analysis of the BRS was first undertaken with the 2011 RSCMP monitoring samples. A second analysis was undertaken with samples submitted with the 2012 RSCMP monitoring round.

Three samples of each of the Middlemore and Meola Outer Zone BRS were submitted for PSD analysis with the 2011 and 2012 RSCMP monitoring samples, and analysed by the sieve/pipette method by NIWA Hamilton. These results, which are the first two sets of BRS PSD analyses, provide a measure of within-batch (WB) variability for the sieve/pipette method, and mark the beginning of a between batch (i.e. year-to-year) variability and temporal trend data series (to be used for trend benchmarking).

A comparison of the 2011 and 2012 data are shown in Figure 4-7.

The PSD in the Middlemore BRS was dominated by fine sand and silt (together making up about 80–90% of the sediment by weight), with smaller amounts of clay (approximately 10–20%). Particles greater than 250 µm made up <1% by weight. The mud (<63 µm, silt + clay) content was about 68%. The Meola Outer Zone BRS was dominated by fine sand (95%), with mud content of about 3%.

Variability in the PSD was relatively low for the mud fraction; CVs were <0.4% for Middlemore, and <6% for Meola OZ. Variability was higher for individual grain size fractions of low abundance (e.g. 35% for the clay fraction in the Middlemore BRS in 2011).

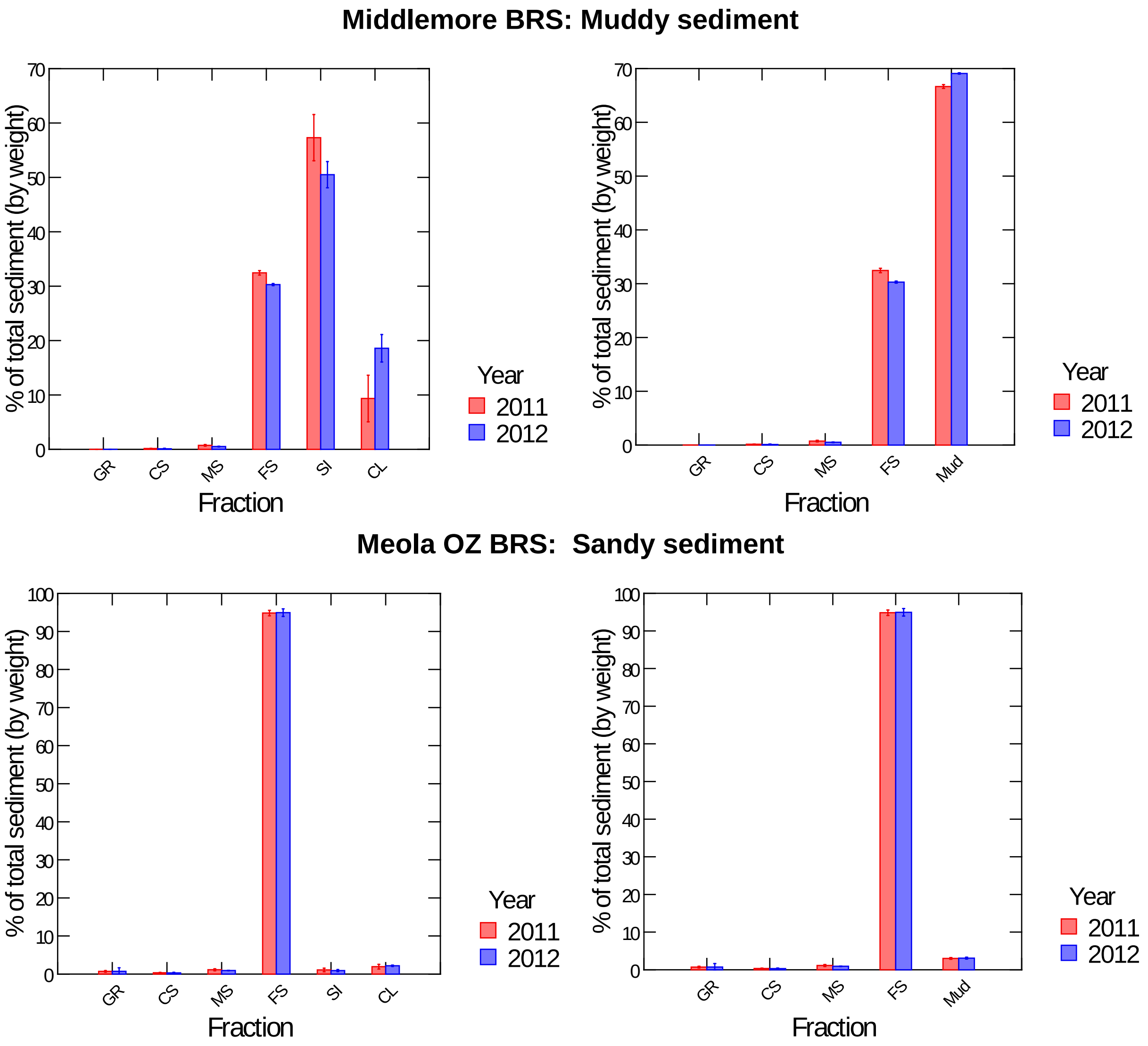
As concluded from the correlations between the ecology plot and contaminant lane samples summarised previously in section 4.4.2, the BRS results indicate that the sieve/pipette method has provided consistent data for mud content assessment, but produced more variable results for individual size class fractions of low relative abundance (proportion by weight).

Comparison of the 2011 and 2012 results showed that:

- for the muddy Middlemore BRS, there was a substantial difference in the results for the individual silt and clay fractions, but the % mud (silt + clay) results were relatively consistent – the mean mud content of the 3 replicates was 66.7% in 2011 and 69.1% in 2012. Although this difference was relatively small (absolute difference between means = 2.4%, RPD = 3.6%), the low variability in each year's results meant the difference in these means was statistically significant (t-test, $p<0.001$).
- for the sandy Meola Outer Zone BRS, the 2011 and 2012 results were consistent for all fractions, even for the minor constituents with <2% abundance. No significant difference in mud content between 2011 and 2012 was recorded (means of 3.01% in 2011 and 3.09% in 2012; t-test, $p=0.592$).

The BRS results obtained in 2011 and 2012 indicated that the sieve/pipette method was giving consistent PSD results across all fractions for the sandy sediment (Meola Outer Zone), but more variable data for individual size-class fractions in the muddy sediment (Middlemore). The mud content (% <63 µm) results were reproducible (low variability within each year), with CVs of <1% for the muddy sediment and <6% for the sandy sediment.

These results, while limited to only two sets of data, look promising for providing the consistent PSD data required for on-going trend monitoring.


Middlemore BRS: Muddy sediment

Meola OZ BRS: Sandy sediment

Figure 4-7 Comparison of Bulk Reference Sediments (BRS) particle size distribution results obtained in 2011 and 2012. Analysis was by the sieve/pipette method (NIWA, Hamilton).

Bars are means (n=3) ± 1 standard deviation. The left hand plots show data for all particle size ranges. The plots on the right combine the silt and clay fraction results into a single mud result. Fraction abbreviations: GR – gravel (>2000 µm), CS – coarse sand (500–2000 µm), MS – medium sand (250–500 µm), FS – fine sand (63–250 µm), SI – silt (3.9–63 µm), CL – clay (<3.9 µm). Mud = silt + clay (% <63 µm).

4.5 Summary of PSD quality assurance

Sediment particle size distribution (PSD) was measured by a range of methods in the former RDP and SoE programmes. However, nearly all the data was obtained by either Galai laser particle size analyser or sieve/pipette analysis.

Examination of correlations between the laser data and the sieve/pipette data has found that a significant linear, approximately 1:1, relationship existed between the laser particle area data (for particles <63 µm, measured over the 0–300 µm or 2–600 µm ranges) and the silt + clay weight fraction data (from sieve/pipette analysis). However, there was too much variation (scatter) in the relationship to allow reliable predictions of mud content (as a % by weight) from laser data on an individual sample basis. The potential differences in results generated by the two methods also means trend data generated from a mixture of these methods may not be very reliable, and hence quantitative trend assessment from data generated by more than one method is not recommended.

The accuracy of the PSD methods, as reported in the monitoring reports reviewed here, is unknown, as no data for standard (or certified) reference sample data were given. Future PSD monitoring should provide the results from analysis of standard samples.

The between-year variability of the laser method was assessed by repeated analysis of archived samples in the SoE programme. However, interpreting these QA data in any detail was not possible due to uncertainties in the identity of the samples used. Therefore, the year-to-year variability in the laser PSD data obtained in the SoE programme is unknown. The QA data, in combination with results from field monitoring, suggested that laser area results generated in 2001 may have been low.

Variability in PSD results generated by the sieve/pipette method has been assessed by correlations between data obtained from ecology plot and contaminant lane sampling at SoE and RDP sites sampled in 2009 and 2010 (n=33), and from analysis of BRS samples analysed in 2011 and 2012. These data indicate that the sieve/pipette method has produced low variability results for mud content (silt + clay fractions), and for size classes with moderate-to-high abundance in the sediments. Measurement of clay content (<3.9 µm) and coarser fractions (>500 µm; coarse sand and gravel) of relatively low abundance (by weight) has been more variable, and should probably be regarded as semi-quantitative. A more robust picture of the performance of the sieve/pipette method will emerge with subsequent sets of BRS data obtained with each annual monitoring round.

5.0 Sampling

This section provides commentary on sampling methods. The potential problems associated with the current blueprint method (ARC 2004) are briefly discussed, and comparisons with alternative approaches made.

5.1 Potential problems with the “blueprint” method

The method used to sample sediment contaminants at (former) SoE and RDP monitoring sites is detailed in the “Blueprint for monitoring urban receiving environments” (ARC 2004). Briefly, this involves sampling the surficial sediment (top 2 cm) along two designated sampling lanes at each (usually) 50 x 20 m site. Samples are taken every 2 m along each lane and collected into five pottles (replicates) – after traversing the approximately 100 m of sampling lanes (50 m x 2 lanes) the pottles each contain 10 sub-samples. Each of the five replicates is homogenised before submitting for contaminant analysis and archiving left over material. Generally, three replicates are analysed, and replicates 4 and 5 retained in case further checks are required.

In practice, this method has worked satisfactorily since its introduction in 1998. The data produced have generally indicated that the replicates are very similar chemically, with analytical variation probably representing the greatest single source of variability between replicates. However, there are some potential issues associated with the method, as outlined below.

5.1.1 Site disturbance caused by on-going resampling

Contaminant sampling only occurs approximately an arm’s length (~ 0.5–1 m) outside the sampling lanes. Samples are taken by reaching out from the sampling lane, taking a small scoop of sediment from alternate sides of the lane every 2 m. This means that there is minimal disturbance of the site, except along the lanes. The lanes run the length of the site, at two locations approximately 10 m apart. Disturbance during contaminant sampling is restricted to within the lane, a width of approximately 0.5 m.

These lanes have been traversed at most sites every two years since 1998 for SoE programme sites, and since 2004 for most RDP sites. This means that the sampling lanes have been walked several times at each site. At muddy sites, this walking is likely to have mixed surface and deeper sediments within the lane, in addition to the normal mixing by bioturbation and physical processes. Since no contaminant sampling is done from within the lane zones, this disturbance is of little importance; assuming that mixed sediment from the lanes is not spread into the adjacent sampling zones. At this stage, there is no information to test whether this has occurred and, if it has, what difference it has made to surface sediments in the sampling zone.

In the original method for sediment contaminant sampling carried out between 1998 and 2004, there was no other use of the site. However, benthic ecology monitoring at some of these sites has been undertaken since 2004 (RDP sites) and 2005 (SoE sites) since the development of the monitoring ‘blueprint’ method (ARC 2004). Ecology sampling involves taking 10 cores from across the site, each core being taken from a random location inside a 10 x 10 m segment of the 50 x 20 m sampling plot. To avoid sampling disturbance near the contaminant sampling zones, ecology cores have not been taken within 1.5–2 m of the contaminant sampling lanes. The locations of the lanes are evident, because the contaminant sampling is always undertaken before the ecology coring.

Ecology sampling has now been undertaken at many sites on at least 3 occasions. Because the cores are taken from locations spread across the site, and these locations vary randomly from one sampling to the next, it would be prudent to assume that much of the site, outside of the contaminant sampling zones, has now been disturbed by sampling activities. How well the site recovers in the two years between samplings is unknown However, benthic ecology sampling is routinely conducted at more frequent intervals in other programmes (e.g. the Central Waitemata Harbour benthic ecology programme) with no evidence of disturbance-related issues.

Again, this disturbance is of little relevance to the contaminant sampling, because the contaminant samples are taken from zones excluded from other sampling. However, as the surrounding site area becomes increasingly mixed by ecology sampling, the surficial sediments in the contaminant sampling zone might be affected by lateral transfer of disturbed sediment into the chemistry sampling lanes.

Whether this has had, or is likely to have, any significant effect on sediment chemistry is currently unknown. However, a limited comparison of sampling methods, which took samples from the sampling lanes (the normal “blueprint” method) and from locations adjacent to ecological sample cores (taken at 10 random locations across each site), showed similar metals concentrations were obtained by the two approaches (see section 5.2.1). This also suggests that site disturbance from repeated samplings may not be a major contributor to any variations in monitoring results, however these sites had only previously been sampled once or twice. In general, sites with the greatest potential to be affected are those with wet soft mud, which is greatly disturbed during sampling.

5.1.2 Missing pegs and sampling lane position uncertainties

From a practical point of view, the blueprint method has worked well. The only practical issue that has arisen relates to missing marker pegs. At the SoE sites, four pegs mark the positions of the sampling lanes, one at each end of the two lanes. For RDP sites, only one peg is used to mark the corner of the sampling plot. The locations of the other corners of the plot, and the sampling lanes, are measured from this peg by pacing out the required dimensions (usually for a 50 x 20 m plot, with sampling lanes 10 m apart), with site axis directions defined by a combination of notable landmarks and compass bearings.

Sometimes, marker pegs are missing. The position of a single missing peg at a SoE site is usually obvious, and it is replaced in the correct position (within probably 1 or 2 m) using a combination of GPS and distance from the nearest remaining peg. When two or more pegs are missing (a rare occurrence), replacing them in the correct position is subject to greater uncertainty.

For RDP sites, peg replacement could be up to a few meters from the original location, because of GPS accuracy limitations (although often buried remnants from a broken peg can be located within a meter of the GPS coordinates). At RDP sites, the position of the sampling lanes is therefore less clearly defined than at SoE sites, and it is quite possible that some of the sampling zone may have been disturbed by previous sampling (2–5 years earlier).

This reliance on accurate location of the sampling lanes to avoid sampling areas disturbed by previous sampling is a potential weakness of the current method, especially for RDP sites. Addition of four pegs (as per SoE sites) would greatly reduce this potential problem, and should be considered for the future.

Adding a marker code to each peg (e.g. 1–4 cuts near the top of the pegs to denote which corner each peg is located) might help clarify which peg(s) are missing, and hence reduce the uncertainty when replacing missing pegs.

5.1.3 The effects of variation in sampling depth on metals concentrations

The blueprint method specifies sampling the top 2 cm of sediment. This depth represents a compromise between sampling a thinner surface layer (which would represent more recently deposited sediment, but which would be prone to greater compositional variability because of short-term effects of storms, erosion etc) and sampling over a greater depth interval (which is likely to better represent longer term accumulation, but would be slower to change concentrations over time due to contaminants being mixed into a greater mass of sediment).

Whether small variations in sampling depth around the 2 cm would contribute significantly to variability in chemical contaminant concentrations is not clearly known. It probably varies with location, depending on a variety of factors, including the depositional characteristics of the site, mixing depths (bioturbation and physical processes), contaminant loads, and historical contaminant concentrations.

An indication of potential effects can be obtained from a study undertaken in Henderson Creek estuary (Diffuse Sources 2003). Shallow cores (0–10 cm) were analysed for Cu, Pb, and Zn at four depths – 0-1 cm, 1–2 cm, 4–5 cm, and 9–10 cm – from channel banks and inside the mangrove fringe at five sites along the estuary. The results showed that changes in metals concentrations with depth varied among the sites, but overall there was little change in concentration with depth (Figure 5-1 - taken from Diffuse Sources 2003).

Averaged across the five main channel bank sites, differences between the surface (0–1 cm) and bottom (9–10 cm) of the cores were -1.4 mg/kg (3.8%) for Cu, +1 mg/kg (+2.2%) for Pb, and -25 mg/kg (-11%) for Zn. These concentration differences were not significant (t-test, $p=0.196$ for Cu, 0.800 for Pb, and 0.083 for Zn).Changes in concentration with depth in the mangrove fringe areas were also small, Zn levels on average being 7.8 mg/kg (4.5%) lower in the 9–10 cm depth than in the top 1 cm of sediment, Pb 0.2 mg/kg (0.5%) lower, and Cu 2.6 mg/kg (7.9%) lower. These concentration differences were not significant (t-test, $p=0.058$ for Cu, 0.854 for Pb, and 0.308 for Zn).

At individual sites, consistent changes were limited to one channel bank site (Site 1, Figure 5-1) where a 25% decrease in Zn occurred over the 10 cm depth profile. Also notable was Pb in the 9–10 cm sample from the site closest to the north-western motorway, which had a markedly higher Pb level than other samples (58.5 mg/kg compared with 44.5 mg/kg at the surface). This could reflect historically higher Pb discharges from the motorway.

Overall, the results indicate that metals concentrations may vary with depth at individual sites, but the effect of sampling over 3 cm rather than 2 cm is likely to be minor.

An exception is likely to occur at sites with thin layers of soft sediment overlying material of a different nature (e.g. basement clay or weathered rock, shell hash, or sand). In these situations, higher variability is to be expected because it is difficult, if not impossible, to avoid including some underlying material in the shallow surface sample taken with a scoop. Sampling notes need to record this to aid data interpretation.

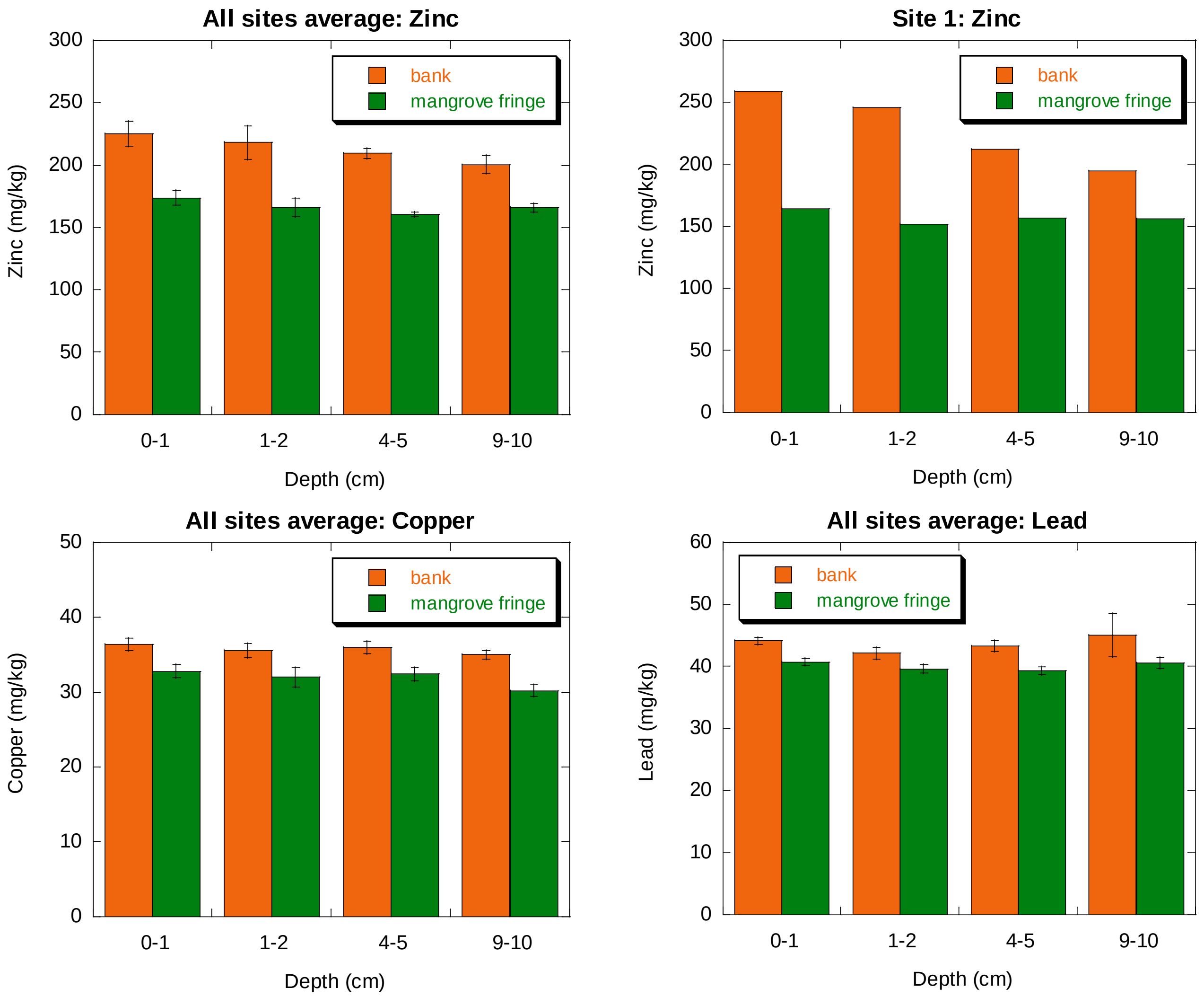


Figure 5-1 Extractable copper, lead, and zinc concentrations (2 M HCl, <63µm fraction) in sediment cores taken on the channel banks and just inside the mangrove fringe at five sites in Henderson Creek estuary (Diffuse Sources 2003).
The “all site average” plots show means ± standard errors in means (n=5).

5.2 Alternative sampling approaches

The blueprint sampling protocol has been used for sediment contaminant monitoring at all SoE and RDP sites to date. Two variations to the protocol have also been trialled or routinely used, and these may provide alternatives to the blueprint method should the need arise. These involve:

- Sampling from adjacent to the 10 ecology core locations at each site – this was trialled at four RDP sites in 2007 (Diffuse Sources 2008); and
- Sampling from cores taken from three random locations within each site – this is the method employed in the UWH benthic ecology programme (Lundquist et al. 2010).

An outline of the performance of these approaches is given below.

5.2.1 Sampling contaminants from ecology core locations

An alternative sediment contaminant sampling approach was trialled in 2006 and in 2007 at RDP sites. Of interest was whether equivalent metals concentrations would be obtained from sampling from the benthic ecology core locations and from the contaminant lanes (the blueprint protocol). It was considered that this approach might have some advantages over the blueprint method in avoiding potentially disturbed sampling zones adjacent to the lane ways, and might provide more rigorous testing of relationships between sediment ecology and contaminants. Sampling from the ecology core locations, rather than separately from the lane ways, could also reduce sampling time and cost, as only a single pass over the site would be required to collect both contaminant and ecology samples.

Initial trial conducted in 2006

As an initial assessment of this alternative sampling approach, composite samples made up from sub-samples taken from adjacent to each of the 10 biological cores at each of 19 sites were analysed for metals. The metals concentrations in these samples were compared with those from the same sites obtained using the conventional RDP sediment contaminant sampling protocol.

The results showed that in almost all cases the results from the ecology composites were very similar to the blueprint protocol results. The most notable exceptions were for:

- weak acid extractable Pb at Hillsborough, although the 2006 data were fairly low and variable at this site, and
- total recoverable Cu at Hobson Bay Awatea, where the ecology core composite sample gave a markedly higher result (approximately double) than the blueprint protocol result. This was considered an anomaly because the Pb and Zn results agreed well in the two sets of samples from this site.

The results from this initial trial (Diffuse Sources 2007) indicated that sampling for metals from the ecology core locations would be an acceptable option. The variability obtained using samples from the ecology coring locations was, however, unknown, because only composite samples were analysed. Future sampling would incorporate variability assessment by taking 3 replicates at each core location, producing 3 replicate composites for analysis (each composite containing 10 sub-samples from adjacent to each ecology core).

A further trial of this approach, using replicated sampling, was therefore undertaken in 2007.

Trial with replicated sampling in 2007

In 2007, four RDP sites were sampled by the two methods. Three replicates (0–2 cm) were sampled from adjacent to each of the 10 ecology cores. Each replicate was therefore a composite sample made up from 10 sub-samples (analogous to the blueprint method, which produced 5 composites (or replicates), each comprised of 10 sub-samples from along the sampling lanes). The metals concentrations in these samples were compared with those from the same sites obtained using the conventional blueprint sediment contaminant sampling protocol.

The results showed that both sampling methods gave very similar results (Figure 5-2). Only the <63 μm metals from Hobson Bay Whakataka were significantly different (slightly higher in the ecology core sampling than in the blueprint method samples; t-test, p = 023–0.089). The Whakataka site sampled at that time was a larger, 100 x 100 m, site. Benthic ecology results indicated that textural heterogeneity (a gradient) was likely to be present across the site. While this may have contributed to the difference between the two sets of metals concentrations, it was not conclusive because total recoverable metals (in the <500 μm fraction) were not significantly different (while <63 μm metals were different). This was not consistent with a textural effect.

Variability of the two methods was similar, although markedly higher variability was recorded for extractable metals (<63 µm) at Pollen Island, and for total Cu (<500 µm) at Whakataka, using the ecology plot sampling method.

The results from this small trial supported those from the 2006 trial and indicated that sampling for metals from the ecology core locations would be an acceptable option for future sediment contaminant monitoring. Undertaking further trials at a wider range of sites would be worthwhile to more comprehensively characterise the comparative data (both agreement between site means or medians, and data variability).

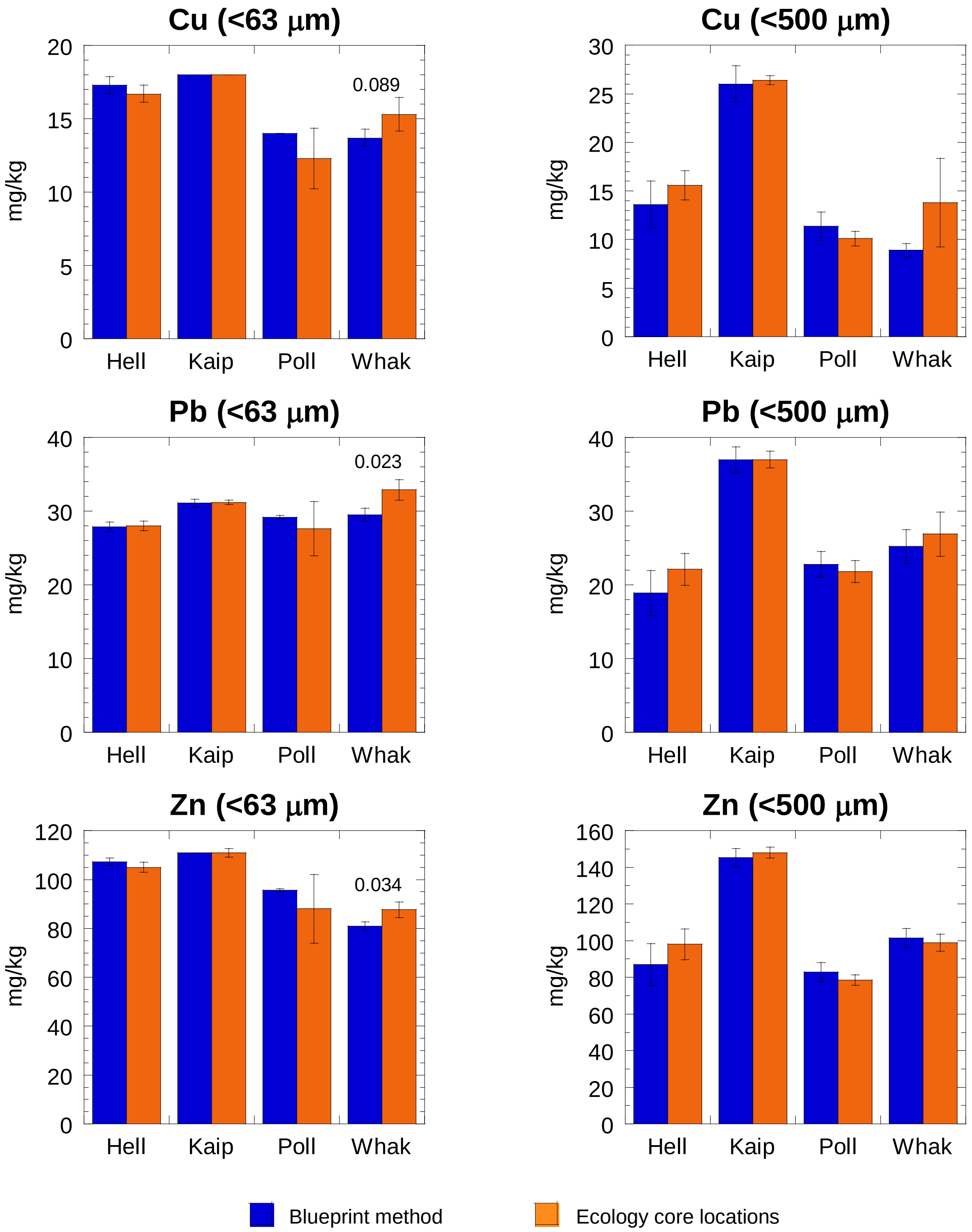


Figure 5-2 Comparison of extractable (<63 µm fraction) and total recoverable (<500 µm fraction) metals from samples taken using the "blueprint" protocol (ARC 2004) and samples taken from adjacent to the 10 benthic ecology cores at four RDP sites in 2007 (taken from Diffuse Sources 2008).
Data are means ± s.d. (n=3 for each method). Only the <63 µm metals at Hobson Bay Whakataka (Whak) were significantly different (t-test $p<0.1$; p values shown above bars).

5.2.2 Upper Waitemata harbour ecology programme method

Sediment chemistry samples are collected as part of the Upper Waitemata Harbour (UWH) benthic ecology programme (Lundquist et al. 2010). The sampling method employed involves collection of at least three replicate cores (5 cm diameter, 0-2 cm depth) from three random locations within each sampling plot. This generates 3 replicate samples, each sample made up from at least 3 sub-samples. Cores from muddy sites are taken at high tide from a boat, and therefore completely avoid the site disturbance that occurs in the blueprint sampling method.

There are no known studies that have directly compared the results from this sampling approach with the standard blueprint method. However, the variability in metals data obtained using this method at the UWH sites can be compared with that from the SoE and RDP programmes, both of which have been sampled using the blueprint method.

A comparison of variability in metals and PAH data acquired in the UWH, RDP, and SoE programmes has been given in section 2.3.5.

While the sites in this comparison are different (so the effects of two variables – site and sampling method, rather than just sampling method – are being compared), it provides a useful way of assessing whether the variability found at the UWH sites is markedly different from that in the rest of the AC monitoring programme. By grouping the results from a range of sites, the effects of differences in individual site characteristics are reduced.

Generally, there was no great difference in data variability between the programmes. The similarity of the variability across the SoE/RDP and UWH programmes, when sample processing and analysis methods were essentially the same in each (for 2008–2011 data), indicates that the difference in sampling procedure between the UWH and the RDP/SoE programmes does not markedly affect the results. However, to conclusively test this, sampling by the two methods at the same sites, on the same occasions would be required.

5.3 Summary

Sampling is currently conducted using two methods – the blueprint method (ARC 2004) and the UWH method (Lundquist et al. 2010). There are no known definitive comparisons of results obtained by these two methods. However, the comparisons described above of variability and concentrations of metals obtained, using the blueprint and ecology core location approaches (including the UWH method) indicates that there is likely to be little difference between them.

The UWH coring approach has a key advantage over the blueprint method – there is no site disturbance (apart from the corer itself). This should reduce any possible artefacts associated with physical mixing of the site during sampling, and prolong the time that the site will provide reliable samples. Because of this potential advantage, the UWH method is probably the better method for sampling muddy, soft sediment sites. For firm/sandy sites, there is unlikely to be any advantage over sampling on foot.

A disadvantage of the UWH method is that the physical nature of the site (on the sampling occasion) cannot be assessed, as sampling is not conducted at low tide when the sediment is exposed. Any changes in physical characteristics of the site contributing to changes in PSD, contaminants, and ecology may not be picked up. In addition, site accessibility (by boat) may be an issue at some small upper estuary sites.

Conducting a trial to directly compare the results from the two sampling methods is recommended, to assess whether it would be suitable for wider use at muddy sites in the RSCMP.

Apart from potential effects of site disturbance, the blueprint method has worked well. Minor modifications, such as coding the site marker pegs (to define which of any pegs are missing), and marking RDP sites with four pegs (as per SoE sites), would be small, but worthwhile, improvements.

6.0 Recommendations for future monitoring

6.1 Sites

Continued sampling at a site is considered worthwhile if it fulfils all of the following criteria: physical stability and consistency across the site (i.e. absence of erosion or major changes in sediment characteristics), clear from mangrove encroachment (now and in the next 10 years or so), and that the site locations provide relevant information for Council needs (e.g. spatial coverage for State of Environment reporting, targeting key project areas such as stormwater network discharges, future urban development zones etc).

All sites monitored to date have been strategically located to provide information that is useful to Auckland Council. As long as they meet the other criteria, continued monitoring of the sites sampled to date is recommended.

Reassessment of the physical characteristics (in particular site erosion and mangrove encroachment) of each site should be made after each monitoring round to evaluate the need for re-siting (or possibly removal from the programme). This is the current approach, and should continue.

There are currently relatively few clean reference sites in the RSCMP/UWH programmes. Reference sites are useful as reality checks for trend monitoring (e.g. elevated metals concentrations at Big Muddy in 2007 signalled a need for data reassessment). Incorporating around three more reference sites (particularly muddy ones) is recommended.

6.2 Sampling protocols

The blueprint and UWH sampling methods appear to give comparable results for sediment contaminants. The UWH method has a significant advantage of minimal site disturbance for muddy sites. Trialling the UWH method at a wider range of sites to obtain more definitive confirmation of the equivalence of the two sampling methods is recommended.

The frequency of sampling should be kept the same as currently used, to build on the baseline data set acquired to date. As discussed in Mills et al. (2012), the current data set is in its infancy for trend monitoring, so acquisition of at least two more sets of data from all sites before changing sampling frequency is considered essential. The uncertainty surrounding the quality of the extractable metals results from SoE sites for 2003–2007 also necessitates further sampling to improve the robustness of the current data set.

6.3 Changing the number of replicates analysed

As discussed in this review, blind duplicate analysis results have shown that the sediment contaminant data are generally too variable for reliable comparisons between sites, or over time, to be made from analysis of single samples. Analysis of replicate samples is therefore required.

Currently three replicates per site are analysed per sampling round, with two additional replicates being stored for analysis if unusual results are obtained from replicates 1–3. The GWRC monitoring data showed lower variability than the Auckland Council data (section 2.8.1). One possible reason for this is the analysis of a greater number of replicates per site (5 reps rather than 3) in the GWRC programme. Increasing the number of replicates will reduce the uncertainty in the annual mean (or median), which is advantageous for comparing differences between two (or more) sets of results. However, how much improvement in trend determination power would result from an increase in the number of replicates analysed each year is not clear (to us) because the trend analysis test (Mann Kendall test) can either use annual medians or all the data. The trend analysis option used for recent reporting of Auckland Council data (Mills et al. 2012) used the “all data” option. Expert statistical advice is required to optimise this aspect of the programme.

6.4 Contaminant analysis methods

Current analytical methods used for primary contaminants (metals and PAH) have satisfactory sensitivity (detection limits; DL). Very few <DL values are obtained, which simplifies data analysis. Secondary contaminants (organochlorines) have, in the past, suffered from the use of unsatisfactory analytical methods, with DLs that were too high to yield useful data. This was discussed in Mills (2014a). It is important that future surveys employ analytical protocols with sufficiently low DLs to obtain reliable quantitative data – DLs well below any sediment quality guidelines (5–10 times lower is recommended) enables contaminant status to be evaluated properly and are more likely to provide a robust baseline for trend monitoring.

As discussed in this report, analytical variability is probably a major contributor to overall monitoring data variability. Continued investigations into key source(s) of analytical variability, in order to reduce it as much as possible, are worthwhile. The introduction of the BRS for QA is part of this process. As the BRS results’ database grows, the contribution of analytical variability and bias (or drift) in variations in the field monitoring data will become more apparent.

Regular checks of data accuracy (or bias) using reference materials should continue. This is especially important for checking data consistency over time for trend monitoring and to assess changes that may occur if analytical providers (labs) are changed.

Increasing the number of potential analytical providers is desirable for on-going security of the monitoring programmes. Assessing the data quality from providers is required to ensure that when changes are made, there are no unexpected changes in data quality (e.g. shifts in analyte concentrations, increases in DLs, or increases in variability). Providers need to be able to clearly document the continued performance of their analytical methods (e.g. including detection limits, accuracy, precision, and QA methods), and to provide detailed method descriptions (unexpected variations in method details encountered during this review highlight the need for this). Auckland Council should consider conducting inter-laboratory comparisons to validate the methods and performance of potential labs.

This review of QA, and detailed analysis of monitoring data (Mills et al. 2012, Mills 2014a), have reinforced our view that “quality is remembered long after the cost has been forgotten” (and that poor quality is also long-remembered, but for the wrong reasons!). The cost of checking, interpreting, reanalysing, and reporting data generally far outweighs differences in the analysis cost between low end and high end providers. Auckland Council requirements for data quality are demanding compared with many screening level applications, and the laboratory methods best suited to reliable status and, in particular, trend monitoring are invariably the high quality, often costly, methods. In the end, using these methods, and obtaining a higher level of confidence in the results that comes from them, is (in our opinion) money well spent.

6.5 PSD analysis methods

The PSD method currently used in the RSCMP, UWH, and related ecological monitoring programmes, is the sieve/pipette/weight method. This provides measures of the proportion (% by weight) of the total sediment in each of 6 particle size ranges.

Advantages of this method include that the results are easy to understand, it is and has been widely used over a range of programmes/investigations including benthic ecology monitoring, and it does not rely on sophisticated instrumentation to perform. Preliminary results from 2011 and 2012 for the BRS QA samples indicate that the precision of this method is good, at least for the combined <63 µm fraction (i.e. silt plus clay fractions). However, the accuracy of the sieve/pipette method is not yet known and this should be investigated. The BRS, which are now used as part of the overall RSCMP monitoring QA protocol, provide a secondary standard for checking within-year and year-to-year variability. However, certified reference materials (or standards) are required to assess absolute accuracy, and should become part of the QA programme for PSD analysis. This is important to ensure consistent results should analytical providers change or more than one provider contributes PSD data to the monitoring programmes.

Currently, PSD is analysed on a single composite sample at each site. There is therefore no measure of variability in the PSD data, although the replicated BRS analysis data provide an indication of this. The lack of replication at each monitoring site limits the usefulness of composite sample data for detecting trends over time, which is a key component of the RSCMP.

Variation in sediment muddiness is an especially important indicator at sandy sites, where benthic fauna may be sensitive to deposition of fine sediment. Replicated measurement of PSD at these sites is worth considering, so changes over time can be assessed with greater certainty. At muddy sites, change in muddiness is likely to be less significant, and therefore replicated PSD analysis may not be as important.

Recommendations:

- Continue to use the sieve/pipette PSD analysis method. Report PSD results for reference standard(s) to provide on-going validation of the accuracy of the method.
- Identify sediment texture thresholds (e.g. % mud content below which changes in muddiness are considered to be ecologically important) and/or key sites where changes in sediment texture are important to track over time (e.g. for ecological, catchment management or other reasons). Undertake replicated PSD analysis (minimum of n=3 each sampling round) at these sites, and at selected reference sites (where no change is anticipated).

6.6 Quality assurance

As this report has shown, quality assurance assessment has been undertaken by a variety of approaches since 1998 when monitoring began. This has made evaluating and concisely reporting the information difficult, and the results have sometimes been ambiguous.

It would be advantageous to standardise on a common QA protocol across the programme components (RSCMP, UWH, and any new initiatives that include sediment contaminant monitoring data). It is recommended that such a QA programme be established, and clearly documented. Data providers to AC should comply with the QA protocols to demonstrate the validity of their data.

Measures to include are those discussed in this report – within batch blind duplicates, analysis of BRS and/or CRM samples (or similar methods to verify accuracy and year-to-year consistency), and analytical method blanks. In addition, definition of data acceptance criteria is required. This could follow criteria used in overseas programmes (see section 2.8.2, 2.8.3, and 2.8.4) where measurement quality objectives are defined for duplicate agreement, acceptable variation from previous concentrations, and agreement with standard reference material concentrations.

The extent of the QA requirements will vary with different analytes – e.g. it would be unrealistic to expect the same level and frequency of QA checks for highly expensive analyses (such as organochlorines) as those required for primary contaminants such as metals, which are used for detailed trend analysis.

The current RSCMP QA approach includes all these elements and could serve as a basis for a formalised protocol. The major issue with the existing QA approach is cost. The analysis of QA samples (including repeats of unusual results, CRMs, WB duplicates, and BRS analyses) and the analysis and reporting of these data represents a significant cost to AC. For example, the 2012 RSCMP monitoring included nearly as many QA check samples for sediment contaminants as new monitoring samples[8]. Much of this QA should be the responsibility of laboratory providers, and therefore collaboration with the laboratories is required to ensure the QA requirements are clearly understood and formally agreed to.

Quality assurance for PSD is also required. The SoE programme undertook a QA programme based on repeated analysis of archived sediments. The use of the BRS continues this approach, with the advantage of not exhausting the supply of archived sample for many years (approximately 20 monitoring years of BRS samples have been archived). However, the BRS is a secondary standard and does not enable the accuracy of the PSD results to be directly verified. Analysis and reporting of PSD reference standard sample(s) is required to provide on-going validation of the accuracy of the PSD results (equivalent to the sediment contaminant CRMs).

---

[8] In 2012, 66 field samples were analysed (22 sites x 3 replicates per site). Quality assurance for sediment chemistry included analysis of CRMs (5 samples), WB replicates (6 samples), BRS sample analyses (6 samples of frozen and freeze dried BRS, from each of the BRS “sites” - a total of 24 BRS samples), and checks of unusual 2011 data (25 samples). A total of 60 QA samples were analysed. For PSD, 3 BRS replicates were analysed.

A key issue is what action is triggered if some (or all) quality criteria are not met. It is possible that sample reanalysis cannot be undertaken, due to lack of sample and cost. The data may have to be accepted, and possibly flagged as being deficient in some way. In addition, simply reanalysing the samples may not result in acceptability criteria being met. It is therefore critical that AC is confident that contracted labs are capable of producing the quality required, and this can be demonstrated before undertaking any important analyses. Confirming in advance with the labs the details of acceptance criteria and how non-conforming results will be handled (e.g. reanalysis, penalty clauses in contracts etc) is required to ensure a workable system is achieved.

Reanalysis of archived SoE samples (section 3.0) indicated that unusual variations in extractable metals results may have been largely due to analysis. Once confident about capabilities, AC should consider reanalysing historical archived samples from more SoE sites to bolster the robustness of the trend data base.

6.7 Analytes

The core suite of metals and PAH routinely analysed to date are still considered appropriate for on-going monitoring. Other metals and metalloids (e.g. Cd, Cr, Ni, Hg, As etc), as well as organic contaminants such as organochlorine pesticides and PCBs have also been analysed as required for improving spatial coverage and for establishing baseline concentrations. New analytes, in particular emerging contaminants, are also beginning to be analysed as analytical methods become available. This on-going, considered, development of the sediment contaminant data base is a sound approach.

Analysis of all the sediment contaminant data has revealed that the analyses that we are least confident about are the weak acid extractable metals (2 M HCl, <63 µm fraction). This analysis has proven to be the most prone to unusual or variable results (e.g. the elevated data from 2003–2007 in the former SoE programme – section 3.0). If the reasons for the sometimes unusual or more variable results encountered in the past cannot be ascertained and remedied, it is recommended that extractable metals analysis be dropped from the programme. Trends in metal concentrations would then need to be assessed from total recoverable metals data, which trend records indicate have been more consistent over time.

Despite the inconsistency in some of the extractable metals data, trend analysis results indicate that a greater number of significant trends were found for extractable Pb and Zn than for total recoverable (Mills et al. 2012). The extractable metals approach (or at least the analysis of the finer <63 µm fraction) would therefore appear worth further attention to improve its reliability. One approach to improve the reliability of the extractable metals data for trend assessment would be to analyse sets of archived samples (as described in section 3.0). This could be limited to key sites where detailed measurement of trends is critical (e.g. those associated with catchment land use change or contaminant management).

Analysis of the QA data (WB, BB duplicates and within-year variability in site replicates – see section 2.3) suggested that greatest variability in metals analysis results has generally occurred for sediments with relatively low mud content (below ~20%). This applied to both extractable and total recoverable metals. Higher variability in extractable metals in sediments with low mud content might be expected due to the small amount of <63 µm fraction material present in these samples. Higher variability in total recoverable metals in coarser textured samples might be associated with greater variations in sediment texture between replicates or the generally lower metals concentrations in sandier samples. For monitoring purposes, the (sometimes) higher variability inherent in the analysis of sandier sediments needs to be acknowledged and realistic acceptance criteria established for variability in the analytical results. Trend assessment at sites with coarser textured sediment may also need to take into account the possibility of more variable sediment contaminant data – trends may take longer to emerge from the noise if analytical data are more variable.

With regard to inherently more variable analyses, such as PAH and organochlorines, assessment of trends is probably limited by analytical variability (although this needs to be confirmed by further work on the statistical detail associated with sample number estimation – see section 2.7). Depending on the outcome of the sample number estimation studies, future monitoring of PAH (and other organic contaminants) may best be targeted at general status assessment and measurement of large changes over time (or between sites), rather than for more sensitive quantification of small trends.

7.0 Concluding remarks

This review has brought together a number of strands of information to provide a picture of data quality and operational issues affecting the AC's sediment contaminant monitoring programmes.

The evolution and growth in the marine sediment contaminant (and ecology) monitoring programmes has meant that, until recently, there has been no consistent quality management protocol or plan covering all the monitoring data. Each component of the overall programme has previously undertaken its own QA assessments, influenced by monitoring objectives, budget constraints, target analytes, operational limitations, and identification of QA issues as they have arisen over time. While a considerable body of useful QA data has been accumulated, some useful and innovative quality assurance initiatives, such as the reanalysis of archived samples conducted in the SoE programme, have sometimes been marred by small but important ambiguities in record keeping and/or reporting.

The variety of different QA assessments conducted in the past have complicated data reporting and analysis, and have not helped us present clear statements about data quality and the consequent implications for data use (in particular temporal trend assessment). Since 2009 most QA procedures have been standardized and additional QA measures have been added. This now allows for comprehensive evaluation and reporting of QA data. However, clear objectives for the QA data still need to be defined and a formalised QA plan implemented.

The key recommendations of this review are therefore to develop a quality assurance plan that clearly defines and embeds data quality objectives and protocols, and to ensure all future AC sediment contaminant monitoring complies with the requirements set out in the plan.

Another key issue arising from this review is the need to involve data users in the future refinement of the monitoring programme, to ensure their needs are met (as far as possible) and that they are aware of, and comfortable with, the limitations of the data produced. Further work on better defining the trend detection capabilities of the programme is one area that stands out as being important in this regard.

The review has highlighted how uncertainty can be created when changes in data providers (especially labs) occur. There is a need for a greater number of potential analytical providers that are capable of producing the analyses required by AC to run its monitoring programme, and for these providers to be able to clearly demonstrate that their capabilities meet AC's quality requirements.

Finally, it is important not to be over critical of the quality of the data gathered to date. A considerable body of highly valuable data has been gathered over the past 15 years, and this has provided a detailed understanding of contaminant status and trends in Auckland's estuaries (Mills et al. 2012, Mills 2014a). The monitoring and associated analyses have been undertaken by professionals, generally using the best knowledge and most appropriate techniques available to them at the time. The AC's monitoring objectives are demanding, and it is in this rigorous context that some issues with the data and areas for further improvement have been identified and acted on. This process of continuous improvement should continue into the future. Hopefully this review will contribute to this process and consolidate the introduction and bedding in of improved monitoring approaches and methods that will add value to AC's future monitoring efforts.

8.0 Acknowledgements

The authors would like to acknowledge the very helpful inputs to this review made by:

- Graham McBride (NIWA), for review of the section on sample number estimation for trend detection (section 2.7)
- Auckland Council staff members Marcus Cameron and Mike Hudson for thorough reviews and useful comments on draft versions.

9.0 References

AGNMI (2012). Report of analysis. Reference Material: AGAL 10. Hawkesbury River Sediment. Australian Government National Measurement Institute, January 2012.

ANZECC (2000). Australian and New Zealand Guidelines for Fresh and Marine Water Quality. National Water Quality Management Strategy. Australian and New Zealand Environment and Conservation Council. Agriculture and Resource Management Councils of Australia and New Zealand. Canberra, Australia.

ARC (2004). Blueprint for monitoring urban receiving environments. ARC Technical Publication No. 168, August 2004.

Diffuse Sources (2003). Project Twin Streams. Estuarine Monitoring Programme. Report prepared for Waitakere City Council, by Diffuse Sources Ltd, June 2003.

Diffuse Sources (2005). Regional Discharges Project sediment chemistry and ecological sampling: October 2004. RDP annual monitoring report prepared by Diffuse Sources Ltd for Auckland Regional Council.

Diffuse Sources (2006). Regional Discharges Project Estuarine Monitoring 2005. RDP annual monitoring report prepared by Diffuse Sources Ltd for Auckland Regional Council. February 2006.

Diffuse Sources (2007). Regional Discharges Project. Estuarine Sediment Monitoring 2006. Report to ARC by Diffuse Sources Ltd, March 2007.

Diffuse Sources (2008). Regional Discharges Project. Estuarine Sediment Monitoring 2007. Report to ARC by Diffuse Sources Ltd, March 2008.

Diffuse Sources (2009). Regional Discharges Project. Estuarine Sediment Monitoring 2008. Report to ARC by Diffuse Sources Ltd, March 2009.

Diffuse Sources (2010). Regional Discharges Project. Estuarine Sediment Monitoring 2009. Report to ARC by Diffuse Sources Ltd, March 2010.

Diffuse Sources (2011). Regional Discharges Project. Estuarine Sediment Monitoring 2010. Report to ARC by Diffuse Sources Ltd, March 2011.

Diffuse Sources (2012). Regional Sediment Chemistry Monitoring 2011. Report to Auckland Council by Diffuse Sources Ltd, September 2012.

Hailes, S.; Hewitt, J and Townsend, M (2010). Kaipara Ecological Monitoring Programme: Report on data collected between October 2009 and February 2010. Prepared by NIWA for Auckland Regional Council. Auckland Regional Council Technical Report 2010/058.

Halliday, J and Cummings, V (2012). Mahurangi Estuary Ecological Monitoring Programme: Report on Data Collected from July 1994 to January 2011. Prepared by NIWA for Auckland Council. Auckland Council Technical Report 2012/003.

Hewitt, J and Simpson, J (2012). Assessment of the Estuarine Ecological Monitoring Programme to 2012. Prepared by NIWA for Auckland Council. Auckland Council technical report 2012/040.

Hewitt, J. E., Lohrer, A M and Townsend, M (2012). Health of estuarine soft-sediment habitats: continued testing and refinement of state of the environment indicators. Prepared by NIWA for Auckland Council. Auckland Council technical report TR2012/012

Kelly (2007). Marine Receiving Environment Stormwater Contaminants: Status Report 2007. Auckland Regional Council Technical Publication No. 333, June 2007.

Kimbrough, K L., Johnson, W E., Lauenstein, G G., Christensen, J D and Apeti, D A (2008). An Assessment of Two Decades of Contaminant Monitoring in the Nation’s Coastal Zone. Silver Spring, MD. NOAA Technical Memorandum NOS NCCOS 74. 105 pp.

KML (2003). Marine sediment monitoring programme 2002 results. Report prepared by Kingett Mitchell Ltd, for Auckland Regional Council. March 2003.

Lauenstein, G. G and Cantillo, A Y (eds). (1993). National Status and Trends Program for Marine Environmental Quality. Sampling and Analytical Methods of the National Status and Trends Program. National Benthic Surveillance and Mussel Watch Projects 1984-1992. Volume I. Overview and Summary of Methods. NOAA Technical Memorandum NOS ORCA 71. National Oceanic Atmospheric Administration, Silver Spring, Maryland. July, 1993

Lauenstein, G. G and Cantillo A Y (eds). (1998). National Status and Trends Program for Marine Environmental Quality. Sampling and Analytical Methods of the National Status and Trends Program Mussel Watch Project: 1993-1996 Update. NOAA Technical Memorandum NOS ORCA 130. National Oceanic Atmospheric Administration, Silver Spring, Maryland. March, 1998.

Lohrer, A.M.; Hailes, S.; Cartner, K.; Wadhwa, S.; Edhouse, S. (2012). Wairoa Embayment: Benthic habitats, environmental status and potential future threats. Prepared by NIWA for Auckland Council. Auckland Council technical report 2012/028.

Lundquist, C.; Townsend, M.; Halliday, J.; Cartner, K.; Parkes, S. (2010). Upper Waitemata Harbour Ecological Monitoring Programme: 2005-2010. Prepared by the National Institute of Water and Atmospheric Research for Auckland Regional Council. Auckland Regional Council TR2010/059.

Mathieson T. J.; Olsen G. M.; Hawken J. L. (2002). Marine Sediment Monitoring Programme - 2001 Results. Report prepared by NIWA. Auckland Regional Council Technical Publication No. 192, July 2002.

McHugh, M. and Reed, J. (2006). Marine Sediment Monitoring Programme 2005 results prepared by M. McHugh and J. Reed, NIWA. Auckland Regional Council Technical Publication No. 316, October 2006.

Mills, G (2014a). Marine sediment contaminant monitoring: organic contaminant data review 2003-2010. Prepared by Diffuse Sources Ltd for Auckland Council. Auckland Council technical report, TR2014/001

Mills, G N (2014b). Marine sediment contaminant monitoring: 2013 data report. Prepared by Diffuse Sources Ltd for Auckland Council. Auckland Council technical report, TR2014/039

Mills, G.N. (2014c). Marine sediment contaminant monitoring: 2012 and 2013 organic contaminants data report. Prepared by Diffuse Sources Ltd for Auckland Council. Auckland Council technical report, TR2014/040

Mills, G.N.; Pickmere, S.E.; Olsen, G.M.; Williamson, R.B. (2000). Marine Sediment Monitoring Programme: 1999 Results. NIWA Client Report ARC00253. Auckland Regional Council Technical Publication No. 135, October 2000.

Mills, G.; Williamson, B.; Cameron, M and Vaughan, M. (2012). Marine sediment contaminants: Status and trends assessment 1998 to 2010. Prepared by Diffuse Sources Ltd for Auckland Council. Auckland Council technical report TR2012/041.

Milne, J.R.; Sorenson, P.G. and Kelly, S. (2009). Porirua Harbour subtidal sediment quality monitoring. Results from the November 2008 survey. Environmental Monitoring and Investigations Department, Greater Wellington Regional Council. July 2009.

Reed, J. and Gadd, J. (2009). Marine Sediment Monitoring Programme 2007 Results. Auckland Regional Council Technical Report No.098, July 2009.

Reed, J. and Webster, K. (2004). Marine Sediment Monitoring Programme – 2003 Results. Auckland Regional Council Technical Publication No. 246, September 2004.

SFEI (2006). 2004-2005 Annual Monitoring Results. The Regional Monitoring Program for Water Quality in the San Francisco Estuary (RMP). San Francisco Estuary Institute, Oakland, Ca.

SFEI (2011). 2009 Annual Monitoring Results. The Regional Monitoring Program for Water Quality in the San Francisco Estuary (RMP). San Francisco Estuary Institute, Oakland, CA.

Timperley, M. and Mathieson, T (2002). Marine Sediment Monitoring Programme 1998 □2001. Review of Results and Procedures. Report by M Timperley and T Mathieson, NIWA. Auckland Regional Council Technical Publication No. 193, July 2002.

Timperley, M. and Reed, J. (2008). Central Waitemata Harbour Contaminant Study. Predictions of Stormwater Contaminant Loads. Auckland Regional Council Technical Report TR2008/039.

Townsend, M.; Hailes, S.; Hewitt, J.E.; Chiaroni L.D. (2010). Ecological Communities and Habitats of Whangateau Harbour 2009. Prepared by the National Institute of Water and Atmospheric Research for Auckland Regional Council. Auckland Regional Council technical report 2010/057.

URS (2002). North Shore City Council Sediment Sampling Programme. Report prepared for North Shore City Council. May 2002.

USEPA (2004). The Incidence and Severity of Sediment Contamination in Surface Waters of the United States. National Sediment Quality Survey Second Edition. EPA-823-R-04-007, November 2004. United States Environmental Protection Agency Office of Science and Technology Standards and Health Protection Division Washington, DC 20460.

USEPA (2008). National Coastal Condition Report III. EPA/842-R-08-002. United States Environmental Protection Agency Office of Research and Development/Office of Water, Washington, DC 20460. December 2008.

USEPA (2010). National Coastal Condition Assessment Quality Assurance Project Plan. United States Environmental Protection Agency, Office of Water, Office of Wetlands, Oceans and Watersheds. Washington, D.C. EPA/841-R-09-004. July 2010.

Ward, R.C.; Loftis, J.C.; and McBride, G.B. (1990). Design of Water Quality Monitoring Systems. Appendix A. Sample size calculation formulae, A.2.2 Testing (p 220). van Norstrand Reinhold, New York, 1990.

Williamson, R.B; Costley, K.J.; Mills, G.N.; Pickmere, S.E. (1999). Marine Sediment Monitoring Programme. Design and 1998 Results. NIWA Client Report ARC90240. Auckland Regional Council Technical Publication No. 107, June 1999.