Literature Review on Performance Testing Approaches for Gross Pollutant Traps

Luis Neumann and Ashok Sharma

Report to Stormwater Industry Association

Commercial-in-confidence
Australia is founding its future on science and innovation. Its national science agency, CSIRO, is a powerhouse of ideas, technologies and skills.

CSIRO initiated the National Research Flagships to address Australia’s major research challenges and opportunities. They apply large scale, long term, multidisciplinary science and aim for widespread adoption of solutions. The Flagship Collaboration Fund supports the best and brightest researchers to address these complex challenges through partnerships between CSIRO, universities, research agencies and industry.

The Water for a Healthy Country Flagship aims to provide Australia with solutions for water resource management, creating economic gains of $3 billion per annum by 2030, while protecting or restoring our major water ecosystems.

For more information about Water for a Healthy Country Flagship or the National Research Flagship Initiative visit www.csiro.au/org/HealthyCountry.html


Copyright and Disclaimer
© 2010 CSIRO To the extent permitted by law, all rights are reserved and no part of this publication covered by copyright may be reproduced or copied in any form or by any means except with the written permission of CSIRO.

Important Disclaimer:
CSIRO advises that the information contained in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice. To the extent permitted by law, CSIRO (including its employees and consultants) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.

Cover Photograph:
From Author’s personal file.
Description: Culvert and channel flow.
Photographer: Ashok Sharma
© 2010 Ashok Sharma
CONTENTS

Acknowledgments ........................................................................................................ iii

Executive Summary ........................................................................................................ iv

1. Introduction ...................................................................................................................... 1

2. Gross Pollutants and Gross Pollutant Traps .................................................................... 3

3. Literature Review .............................................................................................................. 5
   3.1. Field Studies ................................................................................................................ 5
       3.1.1. Gross Solids ........................................................................................................ 5
       3.1.2. Suspended Sediment ........................................................................................ 6
       3.1.3. Nutrients and Heavy Metals ............................................................................ 7
       3.1.4. Hydraulic Measurements .................................................................................. 8
   3.2. Laboratory .................................................................................................................. 8
       3.2.1. Gross Solids ........................................................................................................ 9
       3.2.2. Suspended Sediment ........................................................................................ 9
   3.3. Modelling .................................................................................................................. 12
   3.4. Summary ................................................................................................................... 14

4. Testing Standards .............................................................................................................. 16
   4.1. The BMP database ..................................................................................................... 16
   4.2. ASCE Guideline for Monitoring Stormwater Gross Pollutants ..................................... 21
   4.3. ASCE/EWRI ............................................................................................................... 22
   4.4. Technology Acceptance Reciprocity Partnership (TARP) ............................................. 26
   4.5. Technology Assessment Protocol – Ecology (TAPE) ..................................................... 27
   4.6. Environmental Technology Verification (ETV) .............................................................. 28
   4.7. Summary ................................................................................................................... 29

5. Survey on Perceptions of GPT performance ................................................................. 31
   5.1. Methodology ............................................................................................................... 31
   5.2. Results ....................................................................................................................... 32
       5.2.1. Respondents Background ................................................................................. 32
       5.2.2. Maintenance ....................................................................................................... 33
   5.3. Performance ............................................................................................................... 36
   5.4. Other Issues .............................................................................................................. 38

6. Conclusions .................................................................................................................... 40

7. Recommendations .......................................................................................................... 42

References ......................................................................................................................... 45

Glossary ............................................................................................................................... 49

Appendix A – Survey Questionnaire ................................................................................ 50

LIST OF FIGURES

Figure 5.1 Typical cleaning frequency for each GPT group according to council employees ................................................................. 35

Figure 5.2 Typical cleaning frequency for each GPT group according to cleaning contractors ............................................................ 35
LIST OF TABLES

Table 4.1. Summary of methods for BMP water quality monitoring data analysis (adapted from (GS Consultants 2002)) ................................................................. 19
Table 4.2. Comparison of sediment removal efficiency for a pond using different estimation methods. (adapted from (GS Consultants 2002)) ............................................. 21
Table 5.1. Distribution of respondents per state .................................................................. 33
Table 5.2. Number of devices that the respondents are involved in cleaning and maintaining. ................................................................................................................. 33
Table 5.3. Number of respondents that have received training from GPT manufacturer and have read the Maintenance Manual ................................................................. 34
Table 5.4. Number of devices that the respondents are involved in cleaning and maintaining. ................................................................................................................. 34
Table 5.5. Response count for some issues associated with GPTs ........................................ 38

Figure 5.3 Perception of performance for each GPT group according to council employees ................................................................................................................................. 37
Figure 5.4 Perception of performance for each GPT group according to cleaning contractors ................................................................................................................................. 37
ACKNOWLEDGMENTS

CSIRO acknowledges the Stormwater Industry Association for their partial funding of this work.
EXECUTIVE SUMMARY

In the last 20 years, there has been increasing recognition in Australia and overseas of the need to manage urban stormwater not only in its traditional quantity context, but also in terms of a quality context (Engineers Australia 2006). This report focuses on gross pollutants traps that are constantly evolving and have limited data published on their performance (Victorian Stormwater Committee 1999), while detailed field monitoring is also very scarce (Wong et al. 2000).

The Stormwater Industry Association (SIA) has commissioned CSIRO to prepare a literature review to summarise the existing knowledge on gross pollutant traps in terms of testing and performance.

Following discussions with the SIA review panel, for the purposes of this review, the focus was on devices designed to trap materials ranging from gross pollutants (> 5 mm) to coarse sediment (larger than 0.5 mm). Hence, devices which are designed to capture only gross pollutants and incidentally capture some sediment (e.g. floating debris traps, channel nets, side entry pits, and grates) are not considered as the focus is mainly on pipe/culvert drainage systems, with open channel applications not explicitly considered. As part of this project, CSIRO has contacted several Australian manufacturers to provide information on their testing procedures. The purpose of the consultation was to determine if there was a series of procedures that were common and could be adopted to form the basis for an Australian testing protocol. An analyses of the procedures revealed that the testing conditions and procedures are highly heterogeneous. The gross pollutant and sediment size and compositions used by different manufacturers are different, as well as test locations, procedures, data analyses and flow rates. No particular details of any of the testing regimes are included in this report as all information was provided to CSIRO under confidentiality agreements that prevent the inclusion of such information.

As part of the literature review, a consultation was undertaken with specialist cleaning contractors and local councils (GPT users) to assess their perceptions on the performance of these systems. From an initial list of 25 council personnel and cleaning contractors supplied by the SIA, 21 took part in the survey, with 19 full responses and two partial responses. The range of experience among the respondents varied between 3 to 25 years, and the average for all respondents was 11.5 years. To facilitate data collection, the survey aggregated the various GPTs into the following groups:

1. Difference in specific gravity traps (systems which use gravity to separate pollutants that float and that settle without the use of screens by incorporating baffles/booms in (a series of) chambers)
2. Direct screening (devices that incorporate screens in various orientations to the flow and which are not self cleansing)
3. Vortex type devices (devices that direct flow to produce vortices/hydrodynamic separation, but do not have a screen)
4. Continuous deflective separation devices (devices that combine a vortex/hydrodynamic separation with a non-blocking screening system)
5. Others (e.g. inclined screens, devices combining different groups, etc)

Several field studies agree qualitatively on the assessment that GPTs are highly efficient in capturing gross pollutants (Allison et al. 1998; Engineers Australia 2006; Birch and Matthai 2009). However, there are uncertainties on the performance of GPTs for suspended sediments and other pollutants. In a review study on the performance of
The literature review on performance testing approaches for gross pollutant traps in stormwater treatment devices in the United States, the authors concluded that due to the inconsistency of study methods, lack of associated design information and reporting protocols, comparison of different systems is very difficult or impossible (Strecker et al. 2001). The survey results also indicate a perception that some systems are not performing well. The survey, although based on a relatively small number of respondents, revealed a wide range of opinions in several issues, particularly in their perception of system performance.

At present, there are no standard methods or guidelines for the testing of GPTs or other stormwater treatment devices in Australia, but in the United States, several different guidelines and protocols for GPTs and stormwater treatment devices currently exist, both for laboratory and field testing. The protocols in use in the US are not consistent in their requirements and the inconsistency has led to mixed performance results. However, most US protocols generally agree on the need for data quality assurance and control, statistical significance of results and a preference for independent third party studies. It is important to note that the focus on most protocols in the United States is on suspended sediment.

Survey respondents revealed a wide range of opinions on several issues, particularly in their perception of systems performance. The responses disagree even on the performance of the different GPT groups, with council employees and contractors also diverging in their opinions. What the large variation between the answers of the respondents highlights is the vastly different perception of systems performance in the field. It is unclear whether the different perceptions are actually linked to the actual performances of these devices. Nonetheless the variety of opinions could indicate that the systems are actually performing at various degrees of efficiency, which reinforces the need for standardization in testing and maintenance protocols to minimize the differences in efficiency between the systems and guarantee that they meet the operational objectives.

Based on the review of published studies on the performance of GPTs in the field and laboratory, the survey of GPT users and the review of the protocols used in the US, several issues must be assessed during the development of an Australian protocol, which should be developed by an industry-wide review panel. It is suggested that the following issues should be considered in developing testing protocols:

- Third party assessment
- Testing at non-ideal conditions
- Use of different flow rates and concentrations to assess performance
- Appropriate scale up of test results
- Device classification
- Testing for intended purpose
- Establish quality control and assurance protocols
- Possible use of a tiered system performance classification system
- Define standard gross pollutant and suspended sediment size and composition for testing
- Guidance in the protocol on recommended or accepted methods for removal efficiency calculation
- Guidance on appropriate sampling methods for gross pollutants and sediments
- Additional research on performance of GPTs in nutrient removal and scale up between field and laboratory testing
1. INTRODUCTION

In the last 20 years, there has been increasing recognition in Australia and overseas of the need to manage urban stormwater not only in its traditional quantity context, but also in terms of a quality context (Fletcher et al. 2004; Engineers Australia 2006). The traditional approach to stormwater works focuses on the conveyance of stormwater safely and economically from urban areas to receiving waters. Increased awareness of the need to mitigate the environmental impacts of urbanisation, and more recently, interest in alternative water sources is driving a transition in the urban water industry from disparate management of functional areas (water supply, wastewater and stormwater) to integrated urban water management (Engineers Australia 2006). As part of the integrated management, there have been several advances in the treatment processes for urban stormwater aiming to reduce pollutant loads before discharge to waterways. In recent years, the focus in stormwater treatment has moved from concentration targets to load reduction targets.

However, as noted by Fletcher et al (2004, page i):

"With this increased attention, has come a rapid development in knowledge of key issues affecting stormwater management:

I. The impact of urbanisation on hydrology.
II. Expected stormwater quality emanating from catchments of different land uses.
III. The performance of a range of stormwater treatment measures in reducing stormwater pollution.
IV. Maintenance and operation of stormwater treatment measures.
V. Expected lifecycle costs of stormwater treatment measures.

This knowledge has been developed by a wide range of researchers and industry practitioners. Consequently, it is difficult for any one organisation to use this information effectively. In addition, despite the recent advances, there are still many knowledge gaps which impact upon the ability of stormwater managers to prioritise, optimise and evaluate their strategies."

This literature review focuses specifically on gross pollutants traps. Gross pollutant traps are constantly evolving and performance data is scarce (Victorian Stormwater Committee 1999), while detailed field monitoring is also very limited (Wong et al. 2000). In a review study on the performance of stormwater treatment devices in the United States, the authors concluded that due to the inconsistency of study methods, lack of associated design information and reporting protocols, comparison of different systems is very difficult or impossible (Strecker et al. 2001). These studies often utilize different methods for data collection, analysis and reporting, resulting in significant differences in the range of treatment efficiency for similar devices. Thus it is difficult to apply the limited information to develop protocols for performance assessment.

The performance, operation and maintenance requirements of Gross Pollutant Traps (GPTs) are determined by individual design characteristics, catchment characteristics and stormwater composition. Design, function and the degree of sophistication of GPTs varies widely, ranging from simpler capture baskets to highly engineered models. The selection of GPTs for a particular site takes several factors in consideration, such as (Wong et al. 2000; Engineers Australia 2006)

- Location and layout
- Design, bypass flow and operation at above design flow
- Pollutant removal efficiency
- Pollutant load
- Maintenance costs and life cycle assessment

The combination of a large number of devices, a lack of reporting protocols and standard methods and only a small number of detailed monitoring studies has resulted in a large uncertainty in GPT selection. Local government, which is largely responsible for the implementation and management of stormwater infrastructure, is dependent on in-house expertise and manufacturer’s advice in selecting appropriate stormwater treatment strategies.

Given the current limited state of knowledge and increasing use of stormwater as a resource, interest in the adequate management and treatment of stormwater will continue to grow. Independent discussions with local government revealed interest in the documentation and development of guidelines and frameworks to assist in the system design, product selection and evaluation to ensure adequate stormwater treatment and management.

As such, the Stormwater Industry Association (SIA) has commissioned CSIRO to prepare a literature review to summarise existing knowledge of gross pollutant traps in terms of testing, design and performance to assist in identifying knowledge gaps. As part of the literature review, a consultation was undertaken with a limited number of specialist cleaning contractors and local councils (GPT users) to assess their perceptions on the performance of these systems.

The literature review includes a series of recommendations to assist in the next steps of the project, which involve the development of a project brief and research proposal to identify potential methods to test gross pollutant trap performance, including modelling, field monitoring and validation and development of a laboratory based validation protocol.

Chapter 2 presents the definitions of gross pollutants used in this report and the choice of GPTs reviewed. Chapter 3 reviews literature from Australia and around the world. Chapter 4 discusses the standard test protocols currently available. Chapter 5 presents the result of an Australia survey with councils and GPT contractors on perceptions of GPT performance. The information in these chapters is used to draw conclusion and recommendations in Chapter 6 and Chapter 7 respectively.
2. GROSS POLLUTANTS AND GROSS POLLUTANT TRAPS

There are different definitions of what constitutes gross pollutants and what could be expected to be captured by a gross pollutant trap. Moreover, there are several proprietary and non-proprietary devices that could be defined as a gross pollutant trap. For the purposes of this review, discussions with the SIA panel defined the pollutants to be considered and also which type of devices are to be included. This chapter presents a brief discussion on this selection, but it is by no means a discussion on gross pollutant types, sources and all the different GPT devices available. For a more comprehensive view on pollutant generation, influence of land uses and operational principles of GPTs, the reader is referred to the Australian Runoff Quality (Engineers Australia 2006) and the Urban Stormwater Best-Practice Environmental Management Guidelines (Victorian Stormwater Committee 1999). In addition to these references, there are Australian guidelines on the selection of GPTs, where the performance information for GPT selection is based on manufacturer supplied information. (JDA 2006) or based on estimations of expected performance (Victorian Stormwater Committee 1999; WBM 2003).

Initially, GPTs were designed to capture mainly litter and debris, but they have evolved and now are capable of not only capturing those pollutants, but also capture Suspended Sediment (SS) and hydrocarbons. As a benefit of capturing sediment, GPTs also can potentially remove particulate bound nutrients (nitrogen and phosphorus) as well as heavy metals and other pollutants that are particle bound.

As recently as 2006, the Australian Runoff Quality (Engineers Australia 2006) observed that “gross pollutants comprise the larger particles of natural material and artificial litter that may be transported by runoff. Definitions of this parameter in the industry are not yet uniform. Some authorities prefer to include the larger grades of sediment with gross pollutants, thus recognizing the design intent of some gross pollutant traps. Others prefer to exclude sediment and emphasize the distinctive properties – larger size and lower density – which are the distinguishing features of the remaining material.”

The definition of gross pollutants usually follows the early works of Essery (1994), which used samplers with a size aperture of 5 mm, and which measured gross pollutant loads retention in a trap with a screen size of 5 mm. As such, in Allison et al. (1998), gross pollutants are defined as the material that would be retained by a five-millimetre mesh screen, practically excluding all sediment except that attached to litter and other large debris.

As mentioned above, some GPTs also capture sediment in addition to capturing gross pollutants. The degree of capture varies for different devices and conditions, but also according to the Particle Size Distribution (PSD) of the suspended sediment. The sediment can be classified into 3 different size ranges, as coarse (particles between 5 and 0.5 mm), medium (size between 0.5 and 0.062 mm) and fine (particles smaller than 0.062 mm) sediment (Victorian Stormwater Committee 1999).

However, following discussions with the SIA review panel, for the purposes of this review, the focus will be on devices designed to trap materials ranging from gross pollutants (> 5 mm) to coarse sediment (larger than 0.5 mm). Where appropriate, the performance, testing and results will be reported for other pollutants such as medium and fine sediment, nutrients and hydrocarbons.

As there are several definitions of pollutants captured by GPTs, there are an even wider range of GPTs, with different operating principles, sizes and associated performance. Following the decision to focus on gross pollutants and coarse sediment, only devices that are designed to capture both types of pollutants are considered in this review. Hence, devices which are designed to capture only gross pollutants and incidentally capture some sediment (e.g. floating debris traps, channel nets, side entry pits, grates) and devices which are not primarily designed to capture gross pollutant and coarse sediment (e.g. settling
ponds) are also not part of the review. Hence, the review is focused mainly on pipe/culvert drainage systems, with open channels application not explicitly considered.

Thus, the review and the associated survey are focused on the following groups of GPTs:

1. Difference in specific gravity traps (systems which use gravity to separate pollutants that float and that settle without the use of screens by incorporating baffles/booms in (a series of) chambers)
2. Direct screening (devices that incorporate screens in various orientations to the flow and which are not self cleansing)
3. Vortex type devices (devices that direct flow to produce vortices/hydrodynamic separation, but do not have a screen)
4. Continuous deflective separation (devices that combine a vortex/hydrodynamic separation with a non-blocking screening system)
5. Others (e.g. inclined screens, devices combining different groups, etc)

The authors and the SIA panel recognize that some devices could be considered to fall into more than one category and that some devices may not be easily categorized. The distribution into groups is mainly to facilitate data collection in the survey and further discussion on recommended testing procedures. Further steps of the project would require a refined definition of these groups.
3. LITERATURE REVIEW

There are generally three broad methodologies for the assessment of GPTs: field scale, laboratory scale and modelling (Engineers Australia 2006). Field scale testing, while closely reproducing operational conditions can be time consuming, expensive and its results can be of limited transferability due to catchment and climate characteristics. Laboratory testing on the other hand is cheaper, but there are still uncertainties in terms of the scale up of performance and the representation of gross pollutants, their characteristics and the transferability of results to different devices. The simulation of GPTs performance is an ongoing development and is therefore of limited application to guiding GPT selection. In the past, most testing by manufacturers was done for hydraulic behaviour, usually under ideal conditions and with no interaction of the flow with litter and gross solids. Additionally field performance testing was non-existent, therefore, any design improvement was based on limited observations (Wong et al. 2000). As a result of this, there is limited information available for Australian manufacturers’ GPT performance testing, and there is not an established Australian standard.

As part of this project, CSIRO contacted several Australian manufacturers to provide information on their testing procedures. The purpose of the consultation was to determine if there was a series of procedures that were common and could be adopted to form the basis for an Australian testing protocol. An analyses of the procedures revealed that the testing conditions and procedures are highly heterogeneous. The gross pollutant and sediment size and compositions used by different manufacturers are different, as well as test locations, procedures, data analyses and flow rates. No particular details of any of the testing regimes are included in this report as all information was provided to CSIRO under confidentiality agreements that prevent the inclusion of such information.

3.1. Field Studies

As noted in Australian Runoff Quality (Engineers Australia 2006), the amount of gross pollutants captured is commonly used to represent its performance. However, due to the variation of loads between events, as well as the occurrence of bypass flow, the use of the amount of gross pollutants captured in the sump as a measure of efficiency is erroneous. The captured amount must be compared with bypass flow and loads by collecting downstream data to properly assess performance. The device also needs to be assessed over several events to determine conditions under which by pass flow occurs, and the Australian Runoff Quality (Engineers Australia 2006) recommends at least six and preferably ten events of varying intensity and duration, with at least one event bypassing the unit. The high cost and logistics associated with field testing resulted in a limited number of studies available.

3.1.1. Gross Solids

In a landmark study, Allison et al. (1998) monitored the performance of a Continuous Deflective Separation (CDS) device located in Coburg, Victoria over 13 runoff events during 1996. The authors concluded that the CDS unit was very efficient, practically capturing all gross pollutant transported in the runoff. The study used a definition of gross pollutants as material larger than 5 mm, which was larger than the mesh of the CDS device (4.7 mm), so gross pollutants that entered the device were necessarily captured. The performance was assumed to be at least 99% trapping efficiency, but this conclusion was based on the fact that only 1% of the flows during the monitored period bypassed the unit and assuming that these flows had the same load as the captured flows. As no bypass flows were sampled, the actual efficiency may be smaller. The analysis of the trapped material in the unit found that 65% to 85% of the dry mass of gross pollutants was organic material, followed by plastic and paper. The authors estimated that assuming the monitored year was a typical year, the annual gross pollutant loads would be 30 kg/ha/yr (dry) or 0.4 m3/ha/yr. Similarly high estimates of removal efficiency are reported for other Australian studies in Brisbane (Greenway et al. 2002; Water & Environment 2004) and Sydney (Birch and Matthai 2009) using CDS units and other devices.
The Brisbane City Council (BCC) held a monitoring program between July 1998 and June 2003 as part of its program of implementation and monitoring of stormwater quality improvement devices, (Water & Environment 2004). During this program, one open GPT with trash rack (at Aspley), three CDS units (at Eagle Junction, Calamvale and Boondall) and three Ecosol units (Manly, Nudgee and Calamvale) were monitored, with nutrients, sediment, water quality and hydraulics measured although actual parameters measured varied between devices and stages of the program.

The monitored trash rack captured 165 kg/ha/year of material, including sediment trapped within the wet basin prior to the trash rack. For the trash rack alone, the amount of material captured was approximately 65 kg/ha/year, with urban litter being less than 1% of the captured material (by weight). Sediment (61%) and vegetation (39%) comprised the remaining material captured. For the CDS and Ecosol units, litter once more did not account for more than 10% of material (by weight) with the exception of the Manly site where litter represented 20% of the captured material. The large amount of litter was attributed to the higher proportion of commercial land use compared to the other sites, suggesting GPTs are not justifiable in largely residential areas. The estimated load captured by the different devices was 160 and 355 kg/ha/year for the CDS units at Eagle Junction and Calamvale respectively, and 1620, 290 and 328 kg/ha/year for the Ecosol units at Manly, Nudgee and Calamvale respectively. The high value of the load at Manly was attributed to the high intensity of the rainfall at the location being able to mobilise large quantities of sediment, but the report suggested that the result was based on limited data and therefore needs to be treated with caution. The material captured for all units was generally composed of vegetation and sediment, with some units (CDS at Kalinga and Ecosol at Calamvale) recorded a higher proportion of vegetation, while other units had a higher proportion of sediment (CDS at Calamvale, Ecosol at Nudgee).

Birch and Mattai (2009) measured suspended sediment, nutrients and heavy metals for the inflow and outflow off a CDS unit located in Chiswick, Sydney during 6 storm events. The authors indicated a high efficiency in gross pollutant trapping due to the elevated levels of debris in the sump (between 61 and 111% of the sump capacity) and the absence of gross pollutants downstream of the trap, although no prior measurements were carried out to validate downstream impact.

3.1.2. Suspended Sediment

Allison et al. (1998) investigated the performance of a CDS unit in reducing Total Suspended Solids (TSS) during wet weather and dry weather flows, and observing that the unit captures sediment much smaller than its screen opening, as also noted by several other authors ((Walker and Cooperative Research Centre for Catchment 1999; Kim et al. 2007; Sansalone and Pathapati 2009). In fact, 90% of the captured sediment was smaller than the screen aperture. Whilst, the inflow concentrations of TSS vary greatly, the outflow concentrations were reduced to a baseline level ranging between 150 and 200 mg/l. The observation of the dry-weather events however showed a different outcome, with no discernible trend in TSS concentrations. A further study by Walker et al (1999) analysing data collected over an extended period on the same CDS unit in Coburg concluded that the CDS unit was effective in reducing 70% of the sediments for concentrations above a background of 75 mg/l. For events with concentrations below 75 mg/l and high flow rates, the authors reported increased concentrations in the outflow when compared to the inflow, probably due to resuspension. This behaviour is also found in other field studies for devices based on different technologies (Water & Environment 2004; Birch and Matthai 2009) and laboratory studies (Kim et al. 2007; Pathapati et al. 2008). During dry weather flows however, the study by Walker et al (1999) showed no significant effect on the TSS concentration, with erratic performance including increases in concentration.

A more recent study by Kim and Sansalone (2008) measured flow rates, Particle Size Distributions (PSD) and concentration for the inflow and outflow runoff from a paved urban area flowing into a screened hydrodynamic separator. The reported efficiencies were calculated using Event Mean Concentrations (EMCs) and also mass based loads. EMC based efficiencies for the coarser sediment fraction (> 75 μm) varied between 75 and 94%,
while for the entire sample they varied between 9.8 and 67.2%. The large difference is due to the lower EMC efficiencies for the fractions smaller than 75 μm, as the inflow has much coarser sediment than the outflow. The same event which resulted in a 9% efficiency using EMCs had a 49% efficiency on a mass basis, as much of the large sediment was removed, thus affecting the load but not the concentration dominated by the smallest fractions. It also indicates that much of the removal is due to volumetric storage rather than sedimentation. The authors also hypothesized that higher efficiencies at higher concentrations are due to the fact that higher concentrations occur at higher runoff rates and the higher flows also carry coarser sediment which settles easy. While that may be true at field sites, Fenner and Tyack (1997) reported no changes in efficiency for hydrodynamic separation of polystyrene beads, although their particles size range (400-2690 μm) and concentrations (3000-31600 mg/l) were much larger than most studies involving sediment in urban runoff. The higher efficiency in capturing larger particle sizes with decrease performance for smaller sizes in also found in other studies, such as field measurements using CDS and Ecosol units (Water & Environment 2004), laboratory testing (Kim et al. 2007; Ismail and Nikraz 2009) and modelling studies (Egarr et al. 2004; Dickenson and Sansalone 2009).

Comparison of the performance of devices in terms of sediment removal is difficult due to differences in sampling methods and PSD encountered in field measurement. Use of different PSDs result in different relationships between Total Suspended Solids (TSS) and Suspended Solids Concentration (SSC) (Clark and Pitt 2008; Clark and Siu 2008), which makes comparison without knowledge of the full PSD difficult. Using different procedures to measure SSC and TSS, Clark and Liu (2008) demonstrated that for two PSDs, one of fine silica (median particle size 100 μm) and another of silica and sand (median particle size 500 μm), SSC and TSS measurement methods yielded significantly different results. The TSS method reproduced the concentration accurately for the tested range between 50 and 500 mg/l. On the other hand, the TSS methods had a tendency to underestimate the concentration. The under estimation was greater for the silica and sand mixture with a larger median particle size. Moreover, different pipetting locations within the sample bottle also significantly affected the SSC measurements, leading to over or under estimation of the concentration. In a separate study, Clark and Pitt (2008) did a similar experiment using sediments with a size distribution similar to the ones recommended under the NJCAT (NJDEP 2004) protocols. The results were similar, but for this PSD, the TSS method reported smaller under estimations of the concentration.

3.1.3. Nutrients and Heavy Metals

Allison et al. (1998) also investigated the performance of a CDS unit in reducing Total Phosphorus (TP) and Total Nitrogen (TN) during wet weather and dry weather flows. During wet weather flows, inflow concentrations of TSS vary greatly, the outflow concentrations were reduced to a baseline level ranging between 150 and 200 mg/l. A similar trend in reducing TP and TN was also observed. In all cases, the highest concentration in the inflow occurred at the beginning of an event. The observation of the dry-weather events however showed a different outcome, with no discernible trends and no reduction in TP and TN concentrations. A more detailed study by Walker et al (Walker and Cooperative Research Centre for Catchment 1999) on the same CDS unit found a contrasting behaviour for TP. The average TP concentration increased around 23% during dry weather flows, attributed to increase in the redox potential in the wet sump leading to leaching of adsorbed phosphorus, but in wet weather flow, the unit removed around 30% of TP, mainly material adsorbed to sediment. In terms of TN, the wet weather flow results show an erratic behaviour, while the dry weather results show a reduction in TN of around 13%. The BCC study held between July 1998 to June 2003 (Water & Environment 2004) to monitor one GPT with trash rack (at Aspley), three CDS units (at Eagle Junction, Calamvale and Boondall) and three Ecosol units (Manly, Nudgee and Calamvale) also collected water quality samples, although parameters measured varied between devices and stages of the program. Water quality monitoring was undertaken at the CDS units at Calamvale and Kalinga Park and the Manly Ecosol unit, with limited data collected for the trash rack and the Kalinga Park Ecosol unit. For the units with limited data, no clear trends were observed,
while for the Manly Ecosol unit dry weather sampling was affected by tidal inundation. The dry weather sampling indicated that the CDS unit at Calamvale exported phosphorus and nitrogen, with a similar conclusion for the Kalinga unit although based on more limited data. For wet weather, there were no identifiable trends in terms on reduction of phosphorus for the CDS unit in Calamvale, with the CDS unit in Kalinga Park and Ecosol unit indicating an increase in phosphorus, suggesting both units were releasing phosphorus in storm events. None of the units showed trends in terms on nitrogen capture during wet weather flows, but the data for the CDS unit in Calamvale indicated generation of bioavailable form of nitrogen within the unit. The report notes that, these conclusions are based on a limited amount of data and should be used cautiously.

The results of these studies highlight the highly variable range of results reported for the performance of GPTs in removing nutrients. The high variability and general low efficiency is also found in several other studies, such as field studies by Birch and Mattai (2009), Cook et al (2003) and West et al. (West et al. 2004) or the laboratory study of Nnadi et al (2005) and Al-Hamdan et al (2007).

Birch and Mattai (2009) sampled locations upstream and downstream of a CDS unit located in Chiswick, Sydney during 6 storm events. The authors reported low removal rates for nutrients and heavy metals, with 4% reduction for TP, 6% for TN, 10% for Total Kjeldahl Nitrogen (TKN), other forms such as NOx had -4% efficiencies and all constituents reported negative efficiencies for at least one vent (i.e. generation). This net increase was attributed to resuspension of fine trapped sediment due to high flow rates, with an overall TSS removal of 28%. The results for the concentrations and efficiencies in the study are calculating as a simple average without weighting based on flow rates and therefore subjecting the results to some bias. The study also found low average removals of Cr (5%), Cu, Fe (0%), Mn (10%), with other metal such as Pb (-49%), Ni (-8%) and Zn (-21%) showing an increase in concentration resulting in negative efficiencies.

3.1.4. Hydraulic Measurements

Although the Australian Runoff Quality (Engineers Australia 2006) recognizes that hydraulic analysis of the drainage systems where a GPT is located should be performed, including headloss through the device and diversions under high flow conditions, most of the literature studies do not investigate hydraulic performance. Allison et al. (1998) investigated the hydraulics associated with a installed CDS unit by measuring water levels upstream and downstream of the unit. By assuming that the downstream levels were representative of the scenario where the CDS unit was not present, the authors calculated the afflux caused by the unit for various flow rates. The results show a large difference in water levels between upstream and downstream, particularly if the flow was large enough to flow over the bypass weir. Both afflux and estimated head losses increase with increased flow rates.

The BCC study (Water & Environment 2004) investigated the hydraulic efficiency of the CDS unit in Calamvale over the 18-month monitoring period, with 89% of the flows passing through the unit, and if one only considers storm events that fall within the design flow rate, 100% of the flow was captured by the unit. A similar result is found for the Ecosol unit in Manly, with 95% of flow passing through the system, including flows from events that exceeded the design flow capacity. In both cases, the units capture the recommended 95% of the runoff volume (Wong et al. 2000).

3.2. Laboratory

Due to the high cost of performing field studies and the inherent intra-variability between constituent concentrations and runoff in different catchment contexts, several studies evaluate the performance of GPTs using laboratory studies. In these studies, the device being tested can be a full scale unit or a scaled down version of a device. The Australian Runoff Quality (Engineers Australia 2006) notes that physical hydraulic scale models of GPTs can be useful for optimising the hydraulics of a device during the development stages.
However, even a relatively simple scale up of hydraulics needs to carefully choose the dimensionless number used for scale up, as it usually not possible to match all dimensionless numbers between the model and the field unit (Gulliver et al. 2008; Gulliver et al. 2009). In addition to careful consideration on the use of Froude and Reynold numbers, increased friction factors are necessary to obtain a similar relationship of wall shear stress to flow velocity (Gulliver et al. 2009). Although not straightforward, scale modelling of a GPT in a blocked or full condition for determining hydraulic headloss characteristics is still a valid approach (Engineers Australia 2006).

The scale-up of the efficiency of gross pollutants and sediment removal is even more problematic than the hydraulic performance, as modelling of "scaled" pollutants is poorly understood (Wong et al. 2000). Often, physical hydraulic scale modelling results are unsuitable for determining efficiency as it is not possible to derive comparably scaled gross pollutants (Engineers Australia 2006). The use of scale down pollutant is often difficult as it does not cover the range of shapes and specific gravities of life size gross pollutants, as such they may not reproduce the behaviour of gross pollutants. If one uses large material, the trapping efficiency is overestimated (Engineers Australia 2006). If using laboratory data, it is also imperative to consider duration and pollutant loading rate to evaluate potential deterioration of efficiency.

### 3.2.1. Gross Solids

Al-Hamdan et al (2007) measured the performance of three full scale proprietary devices, a CDS unit (PMSU3019), a Stormceptor (STC4800) and a Baysaver (1K) in removing litter, TSS, nutrient and heavy metals using a facility designed to test full scale devices. Each device was tested for 20, 40, 60, 80 and 100% of the flow capacities in 40 minute tests. In each of the tests, a mixture of organics (3 gallon mixture of leaves and sticks), sediment (street swepped), metal cans (5), plastic bottles (3), shopping bags (1) and cigarette butts (1 cup) were added to the inflow. For all the litter categories, the CDS unit had the highest capture, collecting 100% of the metal, plastic and cigarette boxes. The other two devices also had a reasonable performance, with efficiencies higher than 75% with the exception of the Baysaver reporting efficiencies of 70% (organics), 60% (plastic bags) and 70% (cigarette butts) and the Stormceptor reporting 6.7% (plastic bottles) and 60% (shopping bags). However, such results must be considered carefully as they are based on one measurement with no replicates and using a very small number of samples (e.g. 1 plastic bag per flow rate). Also, comparisons are based on volumetric bases without drying the collected gross pollutants, and the authors recognized that some material did change volume due to water adsorption.

On a separate study, Ismail et al (2009) used scaled pollutant to measure the trapping efficiency of a scale model of a Versa Trap G. The authors used a mixture of “leaves, twigs, cigarettes, film caps” with different sizes and shapes, all with smaller sizes than expected in the field or cut down to size, although no formal scale ratios are presented. The test condition was a screen 50% blocked to simulate operating conditions, and the tested results showed efficiencies of 92 and 94%. Once more, these results have limited transferability as the methodology in the choice of pollutants and their size is somewhat ill-defined, as well as based on limited testing. Other studies in the literature also used debris cut down to scale (Quinn et al. 2005) or simply plastic chips to simulate gross pollutants (Armitage and Rooseboom 2000), with similar issues in methodology.

### 3.2.2. Suspended Sediment

As a general observation, most studies observed that sediment retention in GPTs increases with sediment sizes (Al-Hamdan et al. 2007; Ismail and Nikraz 2009; Wilson et al. 2009; Yu et al. 2010) and that retention efficiency is also larger for lower flow rates ((Schwarz and Wells 1999; Wilson et al. 2009)). In general, efficiencies for particles larger than 500 μm is greater than 90%, even though different studies used different soils or report different
fractionation for the soils. Another interesting aspect observed in different studies is that sediment is not only collected in the spots intended by design such as sumps primary or secondary chambers, but also inlet and outlet pipes in laboratory studies (Schwarz and Wells 1999; Al-Hamdan et al. 2007) and field studies where analysis of the sediments in the screen area and the volute area (between screen and wall) showed that most coarse sediment was captured in the screen area, while the fines were captured in the volute section (Kim and Sansalone 2008). It is debatable whether sediment deposit in outlet pipes should be considered as captured sediment, as large events can potentially move this sediment downstream (Schwarz and Wells 1999).

In their study of three full scale proprietary devices, Al-Hamdan et al. (2007) measured removal of suspended sediment in a CDS (PMSU3019), Stormceptor (STC4800) and Baysaver (1K) devices. In the CDS device, there was a significant difference in the size distribution of sediments collected, with coarser fractions found in the inlet, followed by finer sediments in the sump and an even finer distribution after the screen. A similar scenario was observed in the Baysaver device, with the primary sump containing coarser sediments than the storage sump, while no results were reported for the Stormceptor device. An analysis of the sediments captured by all three devices show very similar results, with the Stormceptor having a slightly larger fraction of fines, indicating a better performance in capturing smaller sediment. In terms of loads, the reductions were 48%, 51% and 67% for the Baysaver, CDS and Stormceptor devices, respectively. However, these results are subject to errors as not all sediment was recovered as it was trapped within litter or within units. In a more thorough study, Schwarz and Wells (1999) used a CDS unit to investigate the separation of sediment using different flow rates (7.9 and 17 l/s), screen aperture sizes (1200 and 2400 μm) and different sand particles. Disassembly of the units showed 16% of the material being deposited in the outlet past the screen, which could be remobilised easily by a larger/ longer event. Not surprisingly, the sediment capture rate decreased with the particle size, and at the smallest sizes, the overall capture was higher using the small aperture screen. For the particle sizes larger than 425 μm, the removal rates exceeded 90% (7.9 l/s) and 80% (17 l/s).

Recognizing the difficulty in comparing studies based on different experimental conditions, efficiency criteria and diversity of devices, Wilson et al. (2009) tested 6 different devices using similar experimental conditions to create a performance function. The authors tested the four hydrodynamic devices in the field (V2B1 Model 4, Vortechs Model 2000, Stormceptor STC4800, CDS PMSU20_15) in controlled conditions and two full scale devices in the laboratory (BaySaver Model 1k and Model 3 ecoStorm). For all units, the tests were run in triplicates, using four flow rates between 15 and 100% of the design flow rate. In each test, a composite sand mixture composed of three discrete fractions with median sizes of 107 μm (range 89-125 μm), 303 μm (range 251-355 μm), and 545 μm (range 500-589 μm) and a concentration of 200 mg/l. The authors correlated the performance using a variation on the Peclet number, the ratio of convection to diffusion, defined as

\[ Pe = \frac{w_s h d}{Q} \]

where \( w_s \) is the settling velocity of sediment, \( h \) is the height of mixing in the chamber, \( d \) is the diameter and \( Q \) is the flow rate. The settling velocity follows an equation proposed by Cheng (1997), shown to be superior to Stokes law to describe settling of sediment (Fentie et al. 2004). Not surprisingly, devices performed better at higher Peclet numbers, which represent either low flow rates or higher settling velocities (higher particle size). The efficiency function can be written as:

\[ \eta = \left( \frac{1}{R^b} + \frac{1}{(a \cdot Pe)^b} \right)^{-1/(b+1)} \]

Where \( \eta \) is the removal efficiency and \( a, b \) and \( R \) are fitting parameters, the initial slope at \( Pe=0 \), the curvature at \( Pe=R/a \) and the efficiency as \( Pe \) approaches infinity respectively.
For all devices, efficiency could be predicted with a Nash Sutcliffe (Nash and Sutcliffe 1970) value between 0.87 and 0.99. For one of the devices, the authors were able to use the function calculated for a full scale device to predict the performance of a device with a false floor and obtained a Nash Sutcliffe coefficient of 0.96 between the prediction and the measurements. The total load removal efficiency of the devices was almost 100% for particles larger than 250 μm and between 30-70% for particles smaller than 250 μm. It should be noted that the tests were performed at relatively high concentrations, and as demonstrated by Walker (1999), the performance may be significantly different at concentration below 75 mg/l. The suitability of this approach to predict performance and scale up should be tested using different range of concentrations.

Another issue with performance prediction is that for the tests performed by Wilson et al. (2009), the devices were dewatered and all sediment was removed before the start of a new replicate or a test at a different flow rate. This neglects the influence of resuspension and screen blockage in the performance of a device. A blocked screen would modify the dimension of the sump, reducing the Peclet number for a given flow rate and particle size, so the performance function could still be valid. However, it is unclear how to consider the potential resuspension of sediment as noted in field studies (Water & Environment 2004; Birch and Matthai 2009) and laboratory studies (Kim et al. 2007; Pathapati et al. 2008).

The effects of resuspension can be estimated using scour tests with the sumps loaded with sediment. Pathapi et al. (2008) used an eccentric hydrodynamic separator based on swirling flow and a 2400 μm screen to measure resuspension and sediment trapping efficiency of a deposited sandy soil with a narrow particle size distribution (94.5% mass between 75 and 150 μm). The efficiency results using a concentration of 200 mg/l ranged from 60 to 90% efficiency for flow between 10 and 100% of the design flow rate (32 l/s), with an arithmetic mean overall efficiency of 82%. The scour tests used the sump filled with the sediment mixture at 50% and 100% capacity and flow rates of 100 and 125 % of the design flow. The outflow concentrations ranged between 103 mg/l (50% load) to 117 mg/l (100% load) at design flow and 175 mg/l (50% load) to 190 mg/l (100% load) at 125% of the design flow. Unfortunately, the authors do not report the particle size distribution of the sediment in the outflow. While the increase in the scouring with flow rates is also reported in another study by Saddoris et al. (2009), in their tested swirling flow device (Environment 21 V2B1 Model 4), but it was not observed in a device based on gravity separation (Stormceptor STC 1200). The test involved flow rates between 50 and 100 l/s, using pre load of the units with either US Silica F110 (median size 50 μm) or AGSCO 70-100 (median size 200 μm). For the swirling flow devices, the concentration of sediments in the outflow increased with flow rates and was higher for the SilicaF110 due to its smaller particle sizes. For the gravity based device, the scour test did not show evidence of resuspension. In order to further explain these results, the authors built a 1:10 scale model of a swirling flow device and used an Acoustic Doppler Velocimeter to measure flow velocities inside the device. The results showed a complex flow pattern and in particular, locations within the device that had their higher flow velocity near the bed, which can cause resuspension.

Another aspect also overlooked in some studies is the measurement method used to derive efficiencies, depending either on the method of calculation (see discussion in the following chapter) or based on different methodologies (Osei et al. 2008). Osei et al. (2008) used two protocols for testing hydrodynamic separators in the laboratory, the direct test method and the indirect method. In the direct method, the PSD of the sediment is measured prior to injection and is compared to the PSD of recovered sediment. The sediment captured within the device is also collected and measured. In the indirect method, the PSD of the inflow is sampled from the inlet pipe and the outflow is usually based on grab samples. Results show that for the same sediment flow rate, differences of over 40% in measured removal efficiency are possible, with the indirect method over estimating efficiency. The over estimation is caused by biased sampling at the inlet due to poor mixing and the position of sampling port within the flow.
3.3. Modelling

The performance of GPTs can also be simulated using computer modelling based on different approaches, such as Computational Fluid Dynamics (CFD) (Egarr et al. 2004; Kim et al. 2007; Ismail and Nikraz 2009; Madhani et al. 2009; Pathapati and Sansalone 2009; Sansalone and Pathapati 2009), non-dimensional numbers (Wilson et al. 2009) and load estimation (Wong and Walker 2002). Other models such as MUSIC (Wong et al. 2002), use transfer functions based on concentration reduction between inflow and outflow. In this case however, the models rely on *a priori* information about the performance of the GPT, and the quality of its results depends on the reliability of the transfer function. The transfer functions however, rely on the manufacturers testing or other similar data, so testing conditions can be quite different to modelled conditions and as such, comparison between devices using this approach is not appropriate.

In an evaluation of field based performance of GPTs in New South Wales, Wong and Walker (2002) compared performance using an index based on simulated and measured loads captured by GPTs. The authors used the load generation data from Allison et al (1998), which calculates a generated load according to rainfall and included an adjustment factor to account for land use. The performance of different systems was then compared using an index calculated by dividing actual captured loads by modelled loads. The authors recognized that the study was limited by the applicability of the load generation equation derived from Melbourne to Sydney catchments and possible inadequate representation of the rainfall series. Moreover, the load generation is based on a single study and more data would be required to validate its use on different catchments. In any case, such a model can be useful to compare units operating in similar conditions, but as it requires maintenance data and is dependent on local generation of loads not explicitly represented in the load equation, it is of limited use to predict performance of GPTs.

While Computational Fluid Dynamics (CFD) models for simulating performance of hydraulic structures is a well-established field and has been extensively used to simulate the performance of swirl concentrators for sewage separation, their use in stormwater is not as widespread (Engineers Australia 2006). CFD can model three dimensional flows across various structures, blocked screens and also track solids of different specific gravity. As such, it has potential to investigate the performance of GPTs in trapping solids through the process of settling and screening (Engineers Australia 2006). However, model development must be accompanied by field or laboratory measurement to allow proper validation of the model under different conditions. It is generally recognized that the modelling of “scaled” gross pollutants is poorly understood (Wong et al. 2000).

CFD can be used not only to determine performance of the separation of sediments, but also the velocity distribution and recirculation zones in blocked litter traps (Madhani et al. 2009) or head loss in hydrodynamic devices (Ismail and Nikraz 2009). Comparison of simulations using the Fluent package and Acoustic Doppler Velocimeter measurements in a scale 1:2 model of a Litterbank trap with fully blocked zones yielded good qualitative comparison of the velocity fields in the chamber and bypass channel. The results also indicate promise for the determination of recirculation zones and near wall velocities which are likely to influence gross pollutant retention (Madhani et al. 2009). On a separate study, Ismail and Nikraz (2009) modelled a Rocla VersaTrap type G (VTG) to compare the head loss of the unit when the screen is 0, 25, 50, 75 and 100% blocked. The experimental results showed that the head loss increases with flow rates, and the calculated headloss coefficients values ranged from 1.89 for 0% up to 2.41 for 75% blocked condition, but at the 100% condition, the headloss coefficient went significantly up to 3.64. The headloss curves also showed a distinct profile for the scenario with 100% blocked screen. However, CFD modelling for the unblocked and 100% blocked screens showed a reasonable agreement between model and measurements, in terms of flow rate-head loss curves. The modelled headloss was larger than the measured head loss, with the difference increasing as the flow rates increase. Nonetheless, CFD is capable of simulating the hydraulic performance of GPTs.
The use of CFD modelling must be accompanied by experimental measurements for model validation as the model is reasonably sensitive to not only physical parameters as expected but also to model parameters. Egarr et al (2004) modelled separation of particles in a 4 m hydrodynamic vortex separator using CFD particle tracking (Fluent software) to evaluate the model sensitivity to physical parameters (temperature, particle position, etc) and model variables (choice of dispersion models, length scales, etc). The authors demonstrated that the CFD model was able to predict the settling of cellulose acetate spheres ranging from 1000 μm to 8000 μm in a quiescent column. Then a sensitivity analysis for the modelled hydrodynamic separator showed that both model and physical variables can significantly influence model performance, although no validation against experiments were performed for any of the efficiency calculations. The use of a turbulence dispersion model for flow modelling resulted in around 10% increase in efficiency compared to a test when a turbulence model was not used, while a reduction in the model length scales of three orders of magnitude resulted in around 7% increase in efficiency. As for physical variables, changes in viscosity due to temperature variation between 0.2°C and 25°C resulted in efficiency changes from negligible up to more than 25% for different sizes, as particles have increased settling velocity due to temperature increase (through decreased viscosity). However, the most significant factor was the injection position of particles in the inlet pipe. For particles placed near the top of the pipe, the efficiency was only 58%, whereas for particles placed at the bottom section, the efficiency was 98%. However, such difference in behaviour is not present in other studies using different particles (Pathapati and Sansalone 2009; Sansalone et al. 2009).

The different particle sizes used in laboratory and field performance tests is also another opportunity for CFD modelling to compare or optimize studies, as models can predict performance of particular particle sizes. Pathapi and Sansalone (2009) used CFD to model the performance of a hydrodynamic screened separator in treating various flow rates and the PSD from the NJCAT protocol (NJDEP 2004) for flow rates varying between 0.16 and 29 l/s. The model was able to reproduce the mass based efficiencies within a 10% relative difference. The authors also simulated the behaviour of particles of 450, 75 and 25 μm for a flow rate of 3.8 l/s. The trajectories of the largest particles (450 μm) showed almost instantaneous settling in the chamber, while some of the 75 μm particles and most of the 25 μm particles escaped in the outflow. This is in agreement with another study by Kim et al (2007) using 424, 42.9 and 10.5 μm and a much larger flow rate of 37 l/s, with all 500 μm particles being retained but most of the smallest particles not being captured by the device.

Dickenson and Sansalone (Dickenson and Sansalone 2009) recognized that different variations in PSD used for testing hydrodynamic devices vary widely between protocols used in the US, requiring repeated testing for different agencies. The authors hypothesized that it is possible to model the behaviour of systems by testing a particular PSD with high resolution monitoring and use the results to calibrate a CFD model to be used to model the performance for different frameworks. Their results indicated that a minimum number of 16 discrete sizes are required for the PSD tested, and that the inflow EMCs could be adequately predicted by the CFD model. The authors also presented efficiency removal curves, however no experimental results where shown to validate the model. It is also not clear if different number of sizes are needed for different PSDs with different coarse-fines ratio. In a separate study, Sansalone and Pathapi (2009) monitored a hydrodynamic separator in Louisiana, US and used the measurements to validate a CFD model of the device. The CFD model showed good agreement between the experimental and modelled mass of sediment removed, with a maximum relative percent difference of 10.4%. In all cases, the CFD model overestimated the removal of sediment by the device. The predicted and measured PSD of the outflow were also in good agreement. The authors also tested the use of steady-state flow rates represented by the median, mean and peak flow and an event based EMC to model the performance of the hydrodynamic separator. The results showed significant under and over predictions of performance, particularly when using the median or mean flow, indicating that adequate modelling of such devices requires unsteady flow if it is to represent field conditions.
While most of the modelling efforts are focused on the use of CFD, Mosheim et al (2009) developed a much simpler model. The model links a rainfall runoff model based on curve numbers and a hydrodynamic separator performance model based on the Peclet number (Wilson et al. 2009). The performance models estimates efficiency based on the Peclet number and its dependencies of flow rates (from the rainfall runoff model) and an assumed particle distribution. The authors used the model to estimate the performance of four devices and assuming a particle size distribution. However, they did not provide any model validation to verify the validity of the model.

3.4. Summary

Even though the Australian Runoff Quality (Engineers Australia 2006) highlights the need for monitoring of flow and the amount of gross pollutants upstream and downstream of the GPT, most field studies still rely on collected material. This can be attributed to the high cost and logistics associated with field sampling upstream and downstream of the units. All field studies however agree qualitatively on the assessment that the GPTs tested were highly efficient in capturing gross pollutants (Allison et al. 1998; Engineers Australia 2006; Birch and Matthai 2009). To overcome the high cost of field studies and the inherent intra-variability between constituent concentrations and even runoff for a giving rainfall, several studies evaluate the performance of GPTs in the laboratory. These studies are also somewhat also limited, either by using a limited number of replicates and a very small load to determine performance (Nnadi et al. 2005; Al-Hamdan et al. 2007) or by using scale down versions of pollutants (Armitage and Rooseboom 2000; Quinn et al. 2005; Ismail and Nikraz 2009). The studies using scaled gross pollutants do not attempt any validation in terms of performance in a full scale model, so the use of scale down pollutants remains poorly understood and the determination of performance based on this type of study are still questionable.

However, laboratory testing is very useful for the determination of hydraulic behaviour, provided that attention is given to the choice of the dimensionless numbers used for scale up, as it is usually not possible to match all dimensionless numbers between the model and the field unit (Gulliver et al. 2008; Gulliver et al. 2009). In addition to careful consideration on the use of Froude and Reynold numbers, increase of friction factors are necessary to obtain a similar relationship of wall shear stress to flow velocity (Gulliver et al. 2009). Laboratory and field measurements of head loss showed acceptable results (Allison et al. 1998; Ismail and Nikraz 2009), although results indicate that for fully blocked screens, head loss condition profiles and coefficients become significantly different than for low flows or partially blocked screens (Ismail and Nikraz 2009).

The reported removal efficiency of suspended sediments by GPTs is highly variable due to differences in methodologies across several studies. Indirect and direct sampling methodologies, use of Total Suspended Solids (TSS) and Suspended Solids Concentration (SSC) to measure sediments, and calculations based on different methods such as EMCs or load based make comparison of results difficult. Recognizing the difficulty in comparing studies, Wilson et al (Wilson et al. 2009) tested 6 different devices using similar experimental conditions to create a performance function based on a Peclet number. The authors were able to use the function calculated for a full scale device to predict the performance of a device with a false floor and obtain a Nash Sutcliffe coefficient of 0.96 between the prediction and the measurements. However, the tests were performed at relatively high concentrations, so the suitability of this approach to predict performance and scale up should be tested using different range of concentrations.

The efficiency is also directly linked to the flow rates and concentration used, as several studies found that at high flow rates, increased concentrations are found in the outflow when compared to the inflow, either in the field (Water & Environment 2004; Birch and Matthai 2009) or in the laboratory (Phipps et al. 2004). Resuspension of sediments at high flow rates may be due to the fact that sediment is not only collected in the spots intended by the design, such as sumps primary or secondary chambers, but also inlet and outlet pipes (Schwarz and Wells 1999; Al-Hamdan et al. 2007). It is debatable whether sediment deposit in outlet pipes should be considered as captured sediment, as large events can potentially move this
Sediment downstream (Schwarz and Wells 1999). Sediment also can be made available through scouring at high flow rates in some swirling devices and the scouring rate increases with the increase in the amount of sediment pre-captured (Pathapati et al. 2008).

The number of studies investigating the performance of nutrients removal is small compared to the studies of gross pollutants and suspended sediment. The results on the various devices (although some based on the same technology) show a highly variable range of results for the performance of GPTs in removing nutrients (Allison et al. 1998; Walker and Cooperative Research Centre for Catchment 1999; Cook et al. 2003; Nnadi et al. 2005; Al-Hamdan et al. 2007; Birch and Matthai 2009). The different devices and testing conditions make it difficult to determine the performance of GPTs in removing phosphorus and nitrogen, with studies also indicating the possibility of release of nutrients in dry-weather flows if redox conditions are favourable.

An option to avoid the relatively higher costs of field and laboratory testing is the use of Computational Fluid Dynamic (CFD) models for simulating hydraulic and pollutant removal performance. Studies have show that CFD can reproduce velocity fields (Madhani et al. 2009), flow rate versus head loss profiles (Ismail and Nikraz 2009) and removal of sediments (Sansalone and Pathapati 2009), although in some cases CFD overestimates the head loss or the efficiency. However, CFD modelling must be accompanied by experiments to validate the initial model, as predictions are reasonably sensitive to some model parameters (Egarr et al. 2004), and it unclear what degree of discretization should be applied to the PSD to adequately reproduce its behaviour or how to properly represent gross pollutants in a CFD model.
4. TESTING STANDARDS

To the best of our knowledge, there are no Australian standard methods or guidelines for the testing of GPTs performance. In countries such as Canada, Japan, Korean and Bangladesh there are established Environmental Technology Verification (ETV) protocols but where they may be applied to GPTs, there are no specific guidelines for testing.

More recently, Modra and Drapper (2010) described a series of tests equivalent to an ETV protocol on a product prototype. The study included the development a synthetic stormwater with target concentrations of TSS, TP and TN for performance testing of devices, however, the authors recognized the difficulties in consistently producing a synthetic stormwater that reflected Australian conditions.

The situation is different in the US, where several different guidelines and protocols currently exist, both for laboratory and field testing. In 1996, collaboration between the US Environmental Protection Agency (EPA) and the American Society of Civil Engineers (ASCE), established a program to create a database of Best Management Practices (BMP). In this instance, the BMPs under evaluation include non-structural and structural devices, with the structural including: ponds, GPTs, filters, etc. The database provided guidance of the selection and testing of BMPs. More recently, the ASCE and the Environmental Water Resources Institute (EWRI) established a committee to develop certification guidelines that are specific to manufactured devices (Guo et al. 2008). The committee included engineers, scientists, regulators, manufacturers, and GPT owners. The aim of the committee is to produce a set of guidelines for laboratory testing, field monitoring, scaling up of results, data evaluation, reporting and maintenance. In addition, other protocols used in the US include the Environmental Technology Verification (ETV) protocol (ETV Verification Protocol 2002), the TARP (Technology Acceptance Reciprocity Partnership 2003), NJDEP (New Jersey Department of Environmental Protection 2003), and TAPE (Ecology 2002). A summary of the main issues raised in these guidelines follows, but for complete details, the reader is referred to the original documents.

4.1. The BMP database

As reported by Clary et al (2002), for the BMP database project “the long-term goal is to gather sufficient technical design and performance information to improve BMP selection and design so that local stormwater problems can be effectively addressed.” As part of the project, tasks included collection and evaluation of existing BMP design and performance, creation of a national database (http://www.bmpdatabase.org/), development of performance evaluation protocols and monitoring guidelines and reporting. The monitoring guidelines developed were not part of the original project, but were developed as it became apparent that much of the available data was of limited value.

It is important to note that the BMP database and its guidelines were designed to be applicable to a large number of BMP, not only Gross Pollutant traps, and at present, there is no information on the database in terms of gross pollutants. Therefore, some of the issues discussed in the guideline may not be highly relevant for GPTs.

For inclusion in the database, the studies must be field based, performed on devices post constructions, and in most cases, include calculations of Event Mean Concentrations (EMCs). Studies from vendors or manufacturers are accepted in the database if they meet three basic criteria: data collection is performed by a third party, the database data collection and reporting guidelines were followed and the data collected complies to quality assurance and control. For studies by third parties involving proprietary devices, data summaries and results are sent to appropriate manufacturer prior to inclusion in the database. The manufacturers have 30 days to provide written comments and discuss concerns, which then may be included in the study. The database management team has the final decision on the inclusion of the study and the inclusion of manufacturers’ comments. Mention of names, manufacturers and product results does not constitute endorsement by the database management.
The database developed a standard set of data for reporting, and although it collects different data sets for different BMP types, for every BMP the following categories are collected (Clary et al. 2002):

- test site location characteristics
- sponsoring and testing agencies
- watershed characteristics
- BMP design and cost data
- monitoring locations and instrumentation
- monitoring costs
- precipitation data
- flow data
- water quality data

The recommendation on the parameters to be measured is further discussed by Urbonas (1995) and also in the Urban Stormwater BMP Performance Monitoring (GS Consultants 2002). The database recognizes that methods of data collection could be varied, using different monitoring (such as manual sampling, automatic sampling, different constituent detection levels) and reporting techniques. This is invariably the case as the methods selected depend on local conditions, budgets and expertise. The database also include guidelines on developing a monitoring plan and using statistical information to determine the required number of events to be monitored according to the concentration reduction to be measured and the respective confidence interval. As such, the Urban Stormwater BMP Performance Monitoring outlines and discusses several available techniques, recommended methods and equipment. A review of those methods and recommendations is however beyond the scope of this report.

In general, all of the discussions involving the preparation of the database note the importance of collecting and reporting data that is quality assured and statistically significant (Urbonas 1995; Clary et al. 2002; GS Consultants 2002). Efficiency measures should follow standard methods as much of the disparity between reported efficiencies can be attributed to comparisons between non-equivalent methods. As noted by Strecker et al. (2001), the inconsistency of study methods, lack of associated design information and reporting protocols, makes the comparison of different systems very difficult or impossible. This issue is also noted by Wong (2000), who pointed out that efficiency measures often report volume or weight of solids removed from the trap, but provide no additional information on what has passed downstream so mass balance calculations are not possible. Allison et al. (1998), also reports on several studies not collecting data on material that has escaped traps. Strecker et al. (2001) described an example of the influence of efficiency calculation by comparing the same dataset using statistical characterization of flows, total load in the inflow and outflow, and percent removal by storm (assigned and equal weight). While the two former techniques resulted in calculated efficiencies of 66 and 67%, the former resulted in an efficiency of 48%. Thus, the authors recommended statistical characterization of inflows and outflows as the recommended method, and if enough data was collected, load reduction was also a valid method. Statistical validation is also necessary due to the inter storm variability in loads. As an example, Strecker et al. (2001) estimated the number of samples required to detect a 5, 20 and 50% change in concentration, at a 80% probability that the difference is significant at a 5% level of significance. While for differences of 50% a small number of samples (2-6) could be sufficient, between 61 and 442 samples would be required to detect a 5% reduction in the same data.

As a result, the Urban Stormwater BMP Performance Monitoring (GS Consultants 2002) includes a detailed discussion on concentration, loads and the calculation of efficiency, which is summarized below.
Concentrations, Loads and Event Mean Concentration (EMC)

The three most common forms of representing concentrations are: the measurement of concentrations of pollutants at a few points in time, an estimation of the total load over a specified duration, and the event mean concentration (EMC).

Concentrations measured can be useful for the generation of a pollutograph (i.e., plot of concentration versus time) to investigate intra-event temporal variations in runoff pollutant concentration, but the large number of samples can lead to high monitoring costs.

Loads are typically calculated by the physical or mathematical combination of a number of individual concentration measurements, associated with a certain flow volume, from which loads can be derived. The specific method applied varies according to the sampling and flow measurement techniques used and can include timed samples, flow weighted samples, or some combination of both. As dry weather flows can also contribute substantially to long-term loading, in some instances wet weather and dry weather loads may have to be measured separately.

EMC is a statistical parameter used to represent the flow-proportional average concentration during a storm event, and defines as total constituent mass divided by the total runoff volume. EMCs and flow measurement data can also be used to estimate total loads for an event. Under most circumstances, the EMC provides the most useful means for quantifying the level of pollution resulting from a runoff event, and as such, was the main focus of the National Stormwater BMP Database Project.

BMP Efficiency

The efficiency of stormwater BMPs can be evaluated in a number of ways and the use of different methods causes much of the disparity between reported efficiencies (Strecker et al. 2001).

The BMP database project recognizes that percent removal is an easy-to-understand method, but states that it has too many shortcomings. A percent removal is primarily a function of inflow quality, where higher inflow loads often result in higher efficiencies than for similar condition and lower inflow concentrations. As such, the percentage removal reflects the inflow quality, and test at higher concentrations will lead to better efficiencies. Also, high percent removals can still occur even when high (and therefore unacceptable) concentrations are found in the outflow. Any measure of BMP performance should be universally interpretable regardless of influent concentration, device design and number of samples collected.

Table 4.1 list several methods and the recommendation provided by the Urban Stormwater BMP Performance Monitoring manual, and a brief description of some methods is provided below. All methods provide a single number to characterize the efficiency of the BMP, however they are not designed to determine if the differences in inflow and outflow water quality measures are statistically significant.
Table 4.1. Summary of methods for BMP water quality monitoring data analysis (adapted from (GS Consultants 2002))

<table>
<thead>
<tr>
<th>Category</th>
<th>Method Name</th>
<th>Recommendation</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Historical Methods</td>
<td>Efficiency Ratio (ER)</td>
<td>Not recommended as a stand-alone assessment of BMP performance. More meaningful when statistical approach is used.</td>
<td>Most commonly used method to date. Most researchers assume this is the meaning of “percent removal”. Typical approach does not consider statistical significance of result.</td>
</tr>
<tr>
<td></td>
<td>Summation of Loads (SOL)</td>
<td>Not recommended as a stand-alone assessment of BMP performance. More meaningful when statistical approach is used.</td>
<td>Utilizes total loads over entire study. May be dominated by a small number of large events. Results are typically similar to ER method. Typical approach does not consider statistical significance of result.</td>
</tr>
<tr>
<td></td>
<td>Regression of loads (ROL) or Linear regression</td>
<td>Do not use</td>
<td>Very rarely are assumptions of the method valid. Cannot be universally applied to monitoring data.</td>
</tr>
<tr>
<td></td>
<td>Mean Concentration</td>
<td>Do not use</td>
<td>Difficult to “track” slug of water through BMP without extensive tracer data and hydraulic studies. Results are only for one portion of the pollutograph</td>
</tr>
<tr>
<td></td>
<td>Efficiency of Individual Storm Loads</td>
<td>Do not use</td>
<td>Storage of pollutants is not taken into account. Gives equal weight to all storm event efficiencies</td>
</tr>
<tr>
<td>Alternative Methods</td>
<td>Percent Removal Exceeding Irreducible Concentration or Relative to WQ Standards/Criteria</td>
<td>Not recommended. May be useful in some circumstances</td>
<td>Typically only applicable for individual events to demonstrate compliance with standards.</td>
</tr>
<tr>
<td></td>
<td>Relative Efficiency</td>
<td>Not recommended. May be useful in some circumstances</td>
<td>Typically only applicable for individual events to demonstrate how well a BMP performs relative to how well it would perform if it operate at the highest possible efficiency</td>
</tr>
<tr>
<td></td>
<td>“Lines of Comparative Performance©”</td>
<td>Do not use</td>
<td>Spurious self-correlation. Method is not valid.</td>
</tr>
<tr>
<td></td>
<td>Multi-Variate and Non-Linear Models</td>
<td>Possible future use</td>
<td>Additional development of methodology based on more complete data sets than are currently available.</td>
</tr>
<tr>
<td>Recommended Method</td>
<td>Effluent Probability Method</td>
<td>Recommended Method</td>
<td>Provides a statistical view of influent and effluent quality. This is the method recommended in this guidance manual.</td>
</tr>
</tbody>
</table>

Efficiency ratio

The efficiency ratio (ER) is defined in terms of the average event mean concentration (EMC) as:

$$ ER = \frac{\text{average inflow EMC} - \text{average outflow EMC}}{\text{average inflow EMC}} $$
The average EMC for a series of events is calculated as the sum of all event concentrations divided by the number of events. It may be more appropriate however to calculate, a logarithmic average, as the sum of the log of EMC divided by the number of events, as this transformation allows for normalization of the data for statistical purposes. The log-transformed data has been shown to provide a better estimation of the population mean (Athayde et al. 1983). The assumption inherent in this method is that an equal weight can be applied to all events, irrespective of the concentration, and as such, reduces the potential impacts of smaller/"cleaner" storm events on actual performance calculations. However, it fails to take into account the fact that some BMPs may not have outflow EMCs that are normally distributed and treat to a (relatively) constant level. It is recommended that this should be complemented by an appropriate statistical test to indicate if the differences in mean EMCs are statistically significant.

**Summation of loads**

The summation of loads method defines the efficiency as the ratio of the inflow loads to the summation of outflow loads, where the loads are calculated using the EMC and the total volume of each event:

\[
SOL = 1 - \frac{\text{total inflow loads}}{\text{total outflow loads}}
\]

This method assumes that monitoring data accurately represents the actual entire total loads in and out of the BMP for a long period to eliminate the influence of any temporary storage. It also assumes that events or dry weather flows that were not monitored had a similar performance. As for the EMC method, it is recommended that this should be complemented by an appropriate statistical test to indicate if the differences in mean EMCs are statistically significant.

**Linear regression**

The use of linear regression involves the use of least squares linear regression of inflow loads and outflow loads, with the intercept constrained to zero. It is rare that all of the assumptions for this method are valid, in particular the requirements for linear regression such as evenly space data on the x-axis, lack of bias on the residuals and slope significantly different than zero.

**Mean concentration**

This method defines efficiency as one minus the ratio of the average outlet to average inlet concentrations. It does not require that concentrations be flow weighted and it may have some value for evaluating grab samples where no flow weighted data or total volume is available. The main assumption in this method is that the flows from which the samples were taken are indicative of the overall event, which may be difficult to meet given the intra variability of events. As disadvantages, individual samples are weighted equally, and biases are likely to occur, and such variation in sampling can make comparisons difficult.

**Efficiency of Individual Storm Loads**

Efficiency of Individual Storm Loads (ISL) method calculates efficiency as the mean value of the efficiencies for individual storms. The efficiency for each storm is calculated based on the total inflow and outflow loads. The main assumption is that event size or other characteristics do not affect the overall efficiency. The method also ignores event sizes as all events are given the same weight irrespective of size.

As an example, the guidelines provide a comparison of sediment removal efficiencies in a stormwater pond, calculated using four different methods (Victorian Stormwater Committee 1999), as shown in Table 4.2. The results clearly showed the variation in efficiencies solely based on the calculation method.
Table 4.2. Comparison of sediment removal efficiency for a pond using different estimation methods. (adapted from [GS Consultants 2002])

<table>
<thead>
<tr>
<th>Year</th>
<th>Efficiency Ratio (%)</th>
<th>Summation of Loads (%)</th>
<th>Linear Regression (%)</th>
<th>Efficiency of Individual Storms (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>59</td>
<td>71</td>
<td>79</td>
<td>29</td>
</tr>
<tr>
<td>1993-1994</td>
<td>64</td>
<td>66</td>
<td>82</td>
<td>-2</td>
</tr>
<tr>
<td>1994-1995</td>
<td>95</td>
<td>94</td>
<td>95</td>
<td>89</td>
</tr>
</tbody>
</table>

Effluent Probability Method (recommended method)

The method recommended in the Urban Stormwater BMP Performance Monitoring ([GS Consultants 2002]) is the effluent probability, where the first step is to determine if the inflow and outflow EMCs are statistically different from one another. If they are, the next step is to examine either a cumulative distribution function of inflows and outflows quality or a standard parallel probability plot. The authors of the manual strongly recommended that the stormwater industry accept this approach as a standard “rating curve” for BMP evaluation studies.

4.2. ASCE Guideline for Monitoring Stormwater Gross Pollutants

Most of the protocols developed in the US have been adapted from wastewater treatment, and consequently have focussed on dissolved and suspended pollutant, however, the ASCE has also developed the Guidelines for Monitoring Stormwater Gross Pollutants ([ASCE 2010]).

The guidelines breaks gross solids into 3 categories: Litter, including all anthropogenic litter greater than 4750 μm in size; organic debris such as leaves, twigs, grass clipping, also larger than 4750 μm; and coarse sediments, larger than 75 μm and including litter and organic debris smaller than 4750 μm. The limit of 4750 μm (4.75 mm) was chosen as the lower limit for organic debris and litter as it is not possible to separate smaller particles from the coarse sediment fraction. Similarly, 75 μm is used as the lower size as it is the limit that can be representatively collected and analysed using auto-samplers.

Moreover, the standard approaches of EMCs, total loads or percent removals is challenging for the monitoring of gross solids in the influent and effluent, compared to monitoring fine sediments. At present, it is only possible to measure the volume or mass of the material actually removed, with no proven methods for comparing upstream and downstream loadings since gross solids are measured as a total mass rather than a concentration. The variation of gross pollutant accumulation between events is even greater than the variation for fines, so yearly accumulation data is more useful than short time comparisons. In addition, some gross pollutant traps retain the collected material in a wet sump (wet systems) where measurement of a dry weight is more difficult, the accumulated water may dissolve some of the accumulated solids and some pollutants may leach into the accumulated water. In dry systems, where material is collected above standing water, testing is somewhat simpler and it may be possible to collect nearly 100% of the pollutants in the screens or filters ([ASCE 2010]).

The guidelines recognize that different monitoring programs invariably have different goals as well as different budgets available. As such, the guidelines adopt a three tier level, where each level adds extra requirements to the previous level. A common denominator in all levels is to develop and obtain approval of an appropriate quality assurance plan.

- Level 1 monitoring programs are aimed at demonstrating basic effectiveness in the removal of gross solids, requiring only data to provide minimal performance data and to quantify the mass or volume of gross solids removed. It is designed to demonstrate pollution removal and it is not necessary to include water column monitoring in order
to control costs. Statistical validation of results and analysis of large number of storms is not typically required.

- Level 2 programs include water quality sampling, collection of flow data, and a more detailed analysis of gross solids, and are designed to provide manufacturers or agencies with screening performances. The sampling scheme in the guidelines should be incorporated into a TARP-type program (see below).

- Level 3 programs are complex and expensive studies, used to develop primary data for research programs or for development of new devices.

For all levels of monitoring, the litter and organic debris should be separated from the coarse sediment and smaller debris, while on levels 2 and 3 the smaller debris should be further separated from the coarse sediment. The litter material can either be collected by skimming the surface of the traps, or depending on the device, the material is collected during clean out. Once the material is collected and dried, it can be separated into different categories.

The collection of sediment and debris is performed using various techniques, depending on the device and the monitoring levels. These include grab samplers, core samplers and use of vacuum trucks followed by segregation and dewatering of the samples. For a more detailed description of the sampling guidelines, the reader is referred to the guideline for monitoring stormwater gross pollutants (ASCE 2010).

4.3. ASCE/EWRI

In 2007, ASCE and EWRI established a committee to develop certification guidelines that are specific to manufactured devices for stormwater treatment (Guo et al. 2008). The initial scoping meeting decided to focus on the physical separation of non-dissolved particles based on scientific limitations, time constraints, and existing regulatory requirements. Six committees were formed to address laboratory testing, field monitoring, scale-up, data evaluation, data reporting and maintenance. The committee also identified the US protocols; most of which are described in this report.

Field Testing

The field testing program was focused on field verification of the separation of Particulate Matter (PM) from runoff, and it did not include gross pollutants. It noted that there are several protocols available, looking at TSS, Suspended Sediment Concentration (SSC) or some other indexes to measure PM (Sansalone 2008). The committee recognized the need for protocols as field testing is complex due to variation in catchment, runoff characteristics, testing methodologies and the non-stationary nature of systems performance (Sansalone et al. 2009). The use of protocols would allow for reduction of bias and a defensible level of statistical confidence in the analysis and performance comparison of different devices.

One of the biggest challenges is the selection of the methodology to measure the PM, as tests range from TSS to SCC and also include turbidity, index testing and full PSD characterization. As both the fine and coarse sediment fractions are important for field testing, field testing requires the transport of a varied PSD so the performance is evaluated on the entire PSD (Sansalone et al. 2009). In addition, representative hydrologic measurements and sampling are difficult due to temporal variations.

As such, the field protocols must have as components a geometric component, which is essentially a detailed as constructed data set in standard form, and a data component, where the hydrology, hydraulics, PSD and other relevant site conditions are described. Sansalone et al (2009) recommend three different levels of data components required, with a Level 1 being requirements for event based investigations. A Level 2 requirement includes more detailed measurements, with more input of temporal data and is useful for monitoring and modelling. The Level 3 are for research datasets or other studies for specific pollutants beyond PM such as nutrients or metals.

As an example, a level 1 protocol would require:
- A minimum of 15 events monitored, with 12 for calibration and 3 verification events for modelling. The events should be of at least 5 mm rainfall and include between 25 and 75% of the designed flow rate.

- Hydrologic data including catchment boundaries, historic and event rainfall record, storage depth. Rainfall record intervals should be no greater than the time of concentration for the catchment.

- Hydraulic data such as flow depths and velocity for the entirety of each event, temperature (for the calculation of density and viscosity), a measurement of the residence time distribution for a low and a high flow event using a conservative tracer and a calibration curve for all the hydraulic measurements.

- Chemistry: Constituent species and phases should be analysed no latter than 24h after collection, and should use duplicate composite samples. The committee recognizes the limitation of co-automatic samplers to collect a representative PM, particularly for sediment larger than 75 μm. However, it does not recommend a method to address the issue.

- Particulate: The guidelines describe methods for measurement of PSD, its mass distribution, sediment depth and measurement of specific gravity.

**Data Reporting**

As part of the ASCE/EWRI project, Roseen et al (2008; 2009) reviewed the data reporting guidelines for existing protocols in the US. The authors stated that consistent data reporting guidelines are needed as there is a wide range of factors that can affect the testing results. These factors include the testing environment and methodologies, statistical analysis, data handling and reporting. The current third party verification protocols are not consistent on their requirements. The authors argued that the current protocols have had mixed results despite being in place for over seven years partially due to the inconsistency in reporting methodologies and guidelines.

The specification of a uniform testing environment test in protocols can minimise the potential introduction of biases which are difficult to quantify. Such an environment also provides an accelerated review process and aids the acceptance of stormwater technologies. To allow interpretation by non-experts results should be presented in a clear manner, with summary data indicating whether systems meet performance standards. At the same time however, the report must contain sufficient information (in appendices of experimental data), to indicate the level of statistical confidence (Roseen et al. 2008).

They recommend a framework largely based on the TAPE protocol (see below), including:

- Basic summary
- Definitions
- Site conditions such as land type, land use, slopes, etc
- Technology Description including model, size and sizing methodology, by-pass design and maintenance procedures
- Test Methods and Procedures including inflow and outflow quality, methods and quality assurance
- Testing and Sampling Event Characteristics
- Performance Results and Discussion including EMCs for inflow and outflow, statistical evaluation and performance measures
- Conclusions, Performance Claims, and Limitations

Roseen et al (2009) also recommends that if the testing is not conducted by an independent third party, it must be subjected to an independent review including observation of at least one monitoring event.
Data Analyses

The data analyses committee conclusion reinforces the need for standardized procedures to collect and report performance data so a verification program can be transferred to different regions (Kayhanian et al. 2009). The committee reviewed existing protocols such as TARP and TAPE to make recommendations on the appropriate sampling and data analyses methodology.

The use of flow-weighted composite samples was considered to be the most representative sampling regimen, preferably using automated sampling. The composite sample needs to cover the entire hydrograph to account for temporal variation of runoff and concentrations. The use of grab samples was not recommended as they can introduce bias and provide limited information in terms of total load. However, the authors recognized that grab samples are needed for collecting samples of coarse sediment and litter where automatic samplers are impractical.

The recommendations for data analyses largely follows the recommendation presented in the Urban Stormwater BMP Performance Monitoring (GS Consultants 2002), noting there are several method in use. Kayhanian et al (2009) also recommend the use of the probability method as the most comprehensive way to evaluate performance. The first step is to determine if the data is normally or log normally distributed, followed by parametric tests to determine if the inflow and outflow values are significantly different from each other. The cumulative distribution functions of inflows and outflows quality are plotted in a parallel probability plot. The analysis of the plots shows the differences in EMC for inflow and outflow and their probability. An advantage of this plot is that it highlights the different performance probability at high and low concentrations, when devices performed well at high concentration but poorly at low concentrations. However, the method does not immediately demonstrate the direct link between specific inflow and outflow samples.

Laboratory Testing

The ASCE committee for laboratory testing guidelines was mainly focused on sediments and excluded procedures to evaluate removal of trash, litter and hydrocarbons. In their review of several of the current protocols, Bannerman et al (2009) noted the wide range of flow rates, particle size distributions and other factors in the various protocols. The authors decided to develop performance functions for tested devices instead of specifying flow rates and other factors. This approach would allow evaluation of devices at different locations and conditions the requirement of further testing. The guidelines at present are not available, so no conclusion can be drawn on the suitability of the performance functions recommended and the proposed scale up and transferability of results.

Scale-up

The progress of the scale-up committee is reported in two papers by Gulliver et al (Gulliver et al. 2008; Gulliver et al. 2009). The authors identify the need to address the settling of negatively buoyant particles, the rise of positively buoyant particles (such as oils and litter) and the possible re-entrainment after separation. It was recognized that testing is possible on model or full scale up devices, and scale up to different sizes is possible if accurate scaling procedure can be developed and verified. In their initial discussion, the authors identified a set of dimensionless numbers to be used on the scale-up process.

- Reynolds’s number: The Reynolds number should be higher than a critical Reynolds number to ensure turbulent flow, so:

\[ Re = \frac{Q}{\nu d} \geq Re_c \]

where Q is the discharge through the GPT, \( \nu \) is the kinematic viscosity, d is the smallest of the (important) dimensions in the flow field, either diameter or flow depth.

- Weber number: The Weber number, a ratio of the fluid's inertia compared to its surface tension, also should be higher than a critical value to ensure the surface tension does not affect the flow field:
\[ We = \frac{\rho Q^2}{\sigma d^3} \geq We_c \]

where \( \rho \) is the fluid density and \( \sigma \) is the fluid surface tension.

- Hazen Number: a measure of similarity of settling in a quiescent flow field, defined as:
  \[ Ha = \frac{Q}{A_{ws}} \]
  where \( A \) is the plan area of the GPT.
- Froude Number: a measure of similarity of fluid flow driven by gravity, defined as:
  \[ Fr = \frac{Q}{\sqrt{g d^5}} \]
  where \( g \) is the acceleration of gravity.
- Euler Number: used to scale-up pressure loss, defined as:
  \[ Eu = \frac{\Delta P d^4}{\rho Q^2} \]
  where \( \Delta P \) is the pressure drop.

In addition to the dimensionless number, the friction factor is also increased to obtain a similar relationship of wall shear stress to flow velocity. The friction factor similarity would also be important for removal efficiency, because it affects the turbulence and particle settling.

According to a study by Fenner and Tyack (1997) investigating the performance of hydrodynamic separators using plastic beads in two separators of 1.6 m and 0.3 m diameter. The Froude number was the best predictor at low removal efficiencies, with the Hazen number the best predictor at higher removal efficiencies. The authors proposed a scaling formula incorporating the effect of both numbers, but it needs to be adjusted for the specific gravity of particles and it also combines flow and particle scaling in one equation, and as such, it is difficult to be used in practice (Gulliver et al. 2008; Gulliver et al. 2009).

The final recommendations of the scale-up group can be summarised as:
- The model tested should be large so surface tension does not affect the performance of the model, i.e., it does comply with the Weber number criteria.
- Reynolds’s number should exceed 2000 for open channel and 8000 for closed conduits.
- Similar friction factors are recommended, and may involve adding roughness to the model.
- Scaling in Froude number is possible using the appropriate dimensions in the flow field \( L \) (with \( m \) designating model parameters and \( p \) prototype parameters):
  \[ \frac{Q_p}{Q_m} = \left( \frac{L_p}{L_m} \right)^{5/2} \]
- The scale up follows a form similar to the Hazen number and the Peclet number proposed by Wilson (2009) and can be used for positively or neutrally buoyant particles:
  \[ \frac{Q_p}{Q_m} = \frac{w_{s,p}}{w_{s,m}} \left( \frac{L_p}{L_m} \right)^2 \]

This relationship can be satisfied while also satisfying the Froude number criteria above. However, model-prototype test are still required to validate this approach (Gulliver et al. 2009). The recommended settling velocity formula is:
\[ w_s = \frac{g R d_p^2}{18v + \sqrt{0.75 \cdot C \cdot g \cdot R}} \]

where \( R \) is the specific gravity of the particle, \( d_p \) is the particle diameter and \( C \) is a constant equal to 0.4 for spheres and 1 for typical sand.

- Potential scour should be evaluated using the dimensionless Shields stress, the assumed bottom stress where sediment begins to move, divided sediment characteristics, resulting in the following scale-up relationship:

\[ Q_p = \frac{d_p f_m R_p}{d_{p,m} f_p R_m L_p} \]

4.4. Technology Acceptance Reciprocity Partnership (TARP)

A process started by six states in the US (California, Illinois, Massachusetts, New Jersey, New York, and Pennsylvania) in 1996 led to the creation of the Technology Acceptance Reciprocity Partnership (TARP), to establish common pathways for the reciprocal state approval and permitting of environmental technologies (MSSSMU 2000). The participating states recognized that such reciprocity accelerates technology adoption at reduced costs for both states and manufacturers. However, this agreement did not supersede individual state requirements.

The six states identified three tiers of data requirements for specific technologies, beginning with general requirements for all BMPs, through to detailed studies for regulatory approaches. In particular, the second tier provides vendor guidance for comprehensive performance testing for a particular class of BMP, as well as providing information on additional criteria required by specific partner states. Some of the features of the TARP (Technology Acceptance Reciprocity Partnership 2003) field protocols are:

- Performance claims and specific conditions must be stated as tests performed under optimal conditions may not represent typical stormwater conditions. It is preferred that claim states performance goals and the range of conditions that they are applicable. Where applicable, effect of bypass flow on process efficiency and system performance should be quantified.
- Data must be Quality Assured (QA) and Quality Controlled (QC) to ensure scientific validity and compliance to test plan;
- Provide test site plan and description.
- Third party studies preferred, but provision is made for data auditing if not collected by a third party.
- Sample at least 50% of average annual rainfall, 380 mm spread over at least 15 events at the test location to ensure a large enough sample size and varied representative conditions. This information also establishes performance over time and maintenance schedules.
- Sample to be collected as flow-weighted composite samples covering at least 70% of the total storm flow, including as much of the first 20% of the storm as possible.
- Parameters to be monitored must be based on vendor’s claims and all parameters for which performance is claimed must be tested.
- The mean particle size should be \( \leq 100 \) microns with a recommended distribution of 55% sand, 40% silt, 5% clay, considered typical of sediments found in stormwater. The mean influent concentration should be 100-300 mg/l. Scour tests should also be
performed at 50% and 100% sediment loading and 125% of flow capacity, to
determine loss of sediment.

- The efficiency ratio method is the preferred method for the calculation of loads, but
  the summation of loads should also be calculated where appropriate. Statistical
testing should be performed on performance claim data to ensure that data are
reliable, significant, and within confidence limits.

For the laboratory protocols, the principles outlined in the field protocols are also applied, but
some differences and additional tests are required:

- It is preferred that tests are performed on full scale units, but the difficulty of testing all
  models/sizes is recognized. If models smaller than the smallest mode available are
  tested, hydrodynamic and scale-up information need to be included to allow
  extrapolation to larger units. They should include velocity, settling times, unit
  dimensions, etc.

- Units should be tested at a range of flow rates, at least using 25%, 50%, 75%, 100%,
  and 125% of the design flow as units tend to operate better at lower flow rates. Each
  test should be performed 3 times, so at least 15 tests are run.

- The calculation efficiency is obtained by a weighted method, where the removal
  efficiency for each operation rate is multiplied by a weight. The weights are 0.25, 0.3,
  0.2, 0.15 and 0.1 for flow rates of 25%, 50%, 75%, 100%, and 125% of the design
  flow, respectively.

While the protocols are agreed by the states with different modification, their final certification
and use may vary in different states. The New Jersey Stormwater Best Management
Practices Manual (NJDEP 2004) provides for two levels of certification, interim and final. An
interim certification may be granted if the manufacturer submits a verification report through
the New Jersey Corporation for Advanced Technology (NJCAT). The interim certification is
valid for a limited period and subject to an agreement to submit to a TARP II tier field protocol
or equivalent.

The Massachusetts Stormwater Handbook (MDEP 2008), only provides guidance and allows
approval of devices that have been certified by the TARP protocol. However, it also
recommends using the UMass Stormwater Technologies Clearinghouse database
(www.mastep.net). This database rates devices using a rating system:

- 0: Unrated. Data review not yet conducted.
- 1: There is sufficient TARP-compliant or similar reliable field or laboratory data on this
  technology to be able to evaluate pollution removal efficiency claims.
- 2: Sound field or laboratory performance studies exist, with some caveats.
- 3: Performance studies with some scientific merit exists, but with significant caveats.
- 4: Insufficient reliable data available to evaluate the technology.

The database recognizes that systems rated 1 not necessarily perform better than systems
rated 2, 3 or 4, only that they have been tested in a credible manner.

4.5. Technology Assessment Protocol – Ecology (TAPE)

As part of the Washington State Department of Ecology water quality program, the State of
Washington in US has established the Technology Assessment Protocol – Ecology (TAPE)
(WSDEWQP 2008). The testing protocol is designed for short detention, flow-based BMPs
and it recognizes that it may not be suitable for all stormwater treatment devices. A candidate
can request a review to identify the suitability of the protocol to its particular device.

The protocol recognizes as participants in the process a review committee, the ecology
department and any party proposing the technology. The review committee reviews all
submissions, advice the ecology department on emerging technology and reviews the
protocols as new information becomes available. The ecology department is responsible for facilitation of the project, granting level designations (see below) and publishing information. The party responsible for the technology being evaluated is responsible to prepare the application, data quality assurance plan and evaluation reports.

The use of an independent third party (e.g. a consultant or experienced agency) to oversee the process is recommended, and for reports involving field data, third party involvement is mandatory at the onset of the field program, unless the party requesting the test does not have a direct financial interest in the outcome of testing. The third party must complete the data validation report attesting that it complies with the approved quality assurance protocol, prepare a summary comparing the field results and the manufacturers claims and provide a recommendation on the technology level designation.

The protocol recognizes three levels of certification:

- **Pilot Use Level Designation (PULD):** Technologies that have limited performance data, the PULD allows limited use to enable field testing, and may be granted based solely on laboratory data. This designation is valid for a limited time, during which field testing must be completed and only allows for a maximum installation of five (5) devices. Local governments may allow installation of PULD devices provided the proponent tests all installed devices to obtain a general designation. The department of ecology has to approve the installations after reviewing the test and data quality assurance protocols. An application at this level must include a description of the technology, the hydraulic capacity and performance testing with field or laboratory data to support the claim and data analysis. The data may have been collected under other consistent protocols.

- **Conditional Use Level Designation (CULD):** For technologies that have considerable amount of data collected under other protocols reasonably consistent but that not necessarily fully meet the TAPE protocol. The designation is also only valid for a limited time before completion of field testing and limited to ten (10) installations. In addition to the data required in the previous designation, a third party review must be submitted.

- **General Use Level Designation (GULD):** For technologies that have a general acceptance for the treatment device. Devices with this designation may be used anywhere in the estate, subject to department of ecology conditions. This designation required submission of a technical evaluation report following the TAPE protocols.

The pollutant removal effectiveness should demonstrate that the advertised treatment performance goals are achieved based on annual averages for the entire discharge volume (treated plus bypassed). As opposed to other protocols with a fixed number of events to be sampled, the TAPE establishes a number of sampling events required ranging from 12 to 35. The number is dependent on the different needs to establish statistical significance. The protocol allows for the use of data from several sites provided that they are from similar land uses and the pollutant concentration variability is comparable. The sampling must be automatic flow-weighted composite sampling, discrete flow composite sampling or a combination of both.

The TAPE protocol recommends several methods for the calculation of efficiencies, including the efficiency ratio, the summation of loads and the calculation of efficiency for individual storm loads.

### 4.6. Environmental Technology Verification (ETV)

The US EPA has a voluntary Environmental Technology Verification (ETV) which uses the verification protocol for stormwater source area treatment technologies (ETV Verification Protocol 2002). Its basic methodology is not dissimilar to the TARP field protocol or the BMP database and is intended to verify the performance of a product against the vendor’s claims. The basic components of the test plan under the ETV are describing roles and
responsibilities of involved organizations, technology and site description, sampling and analysis plan, data management, reporting and quality assurance plans.

The ETV process involves three different phases, planning, verification and data assessment and reporting. In the planning phase, all procedures involved in the testing, personnel and involved organization are determined to produce a verification test plan. In the verification test phase, the required tests are performed by the designated field testing organization and data is collected. The last phase includes data analysis and preparation of a public verification report and verification statement. In the ETV program, certification is managed by a company called NSF international who acts as a verification partner by selecting organization responsible for field testing, developing and implementing test plans, verifying data assurance and approving reports.

Once more, the minimum number of events is set at 15 events, but not only composite samples are required, but each composite should include at least 5 sub-samples, two collected in the hydrograph rising limb, two on the falling limb and one near the peak. The minimum runoff depth for an event to be valid is 5 mm. The selection of constituents is based on manufacturer’s claims for each sub-group: sediment/particulates, nutrients, heavy metals, petroleum hydrocarbons and bacteria. All sampling (unless required by special characteristics of pollutants) is to be conducted by automatic samplers, at locations where mixing is maximized to ensure representation. In terms of efficiency, the protocol requires reporting of EMC and efficiency calculated as the summation of loads and the efficiency ratio methods as described in the BMP database section.

4.7. Summary

The current third party verification protocols used in the United States are not consistent on their requirements, and it has been suggested that current protocols have had mixed results due to the inconsistency in reporting methodologies and guidelines (Roseen et al. 2008; Roseen et al. 2009). The use of protocols allows for reduction of bias and a defensible level of statistical confidence in the analysis and comparison of the performance of the different devices. Another reason for standard testing is that it accelerates technology adoption at reduced costs for both states and manufactures (MSSSMU 2000), avoiding the need for regional or local protocols.

The test protocols cannot be too flexible due to the potential introduction of biases which are difficult to quantify and which can be minimized by the use of uniform testing environment. The results have to be presented in a clear manner, so that no high scientific training is needed to interpret the results with summary data indicating whether systems meet performance standards.

Despite their different requirements in terms of testing and methods, all protocols basically agree on the following:

- Data must be quality Assured (QA) and quality controlled (QC) to ensure scientific validity and compliance to test plan;
- All data must be subject to appropriate statistical tests to verify that the performance measures are statistically significant. Reports must contain sufficient information to indicate the level of statistical confidence. Most protocols have a set number of measurements required, with an exception being the TAPE protocol, which establishes a number of sampling events required ranging from 12 to 35. The number is dependent on the different needs to establish statistical significance. The protocols allow for the use of data from several sites provided that they are from similar land uses and the pollutant concentration variability is comparable
- Performance claims and specific conditions must be stated as tests performed under optimal conditions may not represent typical stormwater conditions. It is preferred that claim states performance goals and the range of conditions that they are applicable. Testing is only necessary for performances that the manufacturer would like to make
claims about, i.e., removal of nutrients only needs to be evaluated if the manufacturer wants to claim nutrient reduction by its device

- Third party studies are preferred, and all protocols required involvement of a third party in at least one part of the process. Levels of requirements varied from requirement of testing being entirely conducted by a third party to cases where a third party is only involved in some tests and data quality assurance.
- Provide test site plan and description, data collection plan, methodologies and water quality data

As the efficiency of stormwater BMPs can be evaluated in a number of ways, with different methods causing much of the disparity between reported efficiencies (Strecker et al. 2001), the protocols make recommendation on the efficiency calculation methods. As percent removal is primarily a function of inflow quality, with higher inflow loads resulting in higher efficiencies than for similar condition and lower inflow concentrations, the method is not recommended. In general, use of EMCs or summation of loads is acceptable, and it is recommended or required that they are complemented by an appropriate statistical test to indicate that the differences are statistically significant. Kayhanian et al (2009) also recommend the use of the probability method as the most comprehensive way to evaluate performance. The first step is to determine if the data is normally or log normally distributed, followed by parametric tests to determine if the inflow and outflow values are significantly different from each other. The cumulative distribution functions of inflows and outflows quality are plotted in a parallel probability plot. The analysis of the plots shows the differences in EMC for inflow and outflow and their probability. An advantage of this plot is that it highlights the different performance probability at high and low concentrations, when devices performed well at high concentration but poorly at low concentrations.

Moreover, the ASCE Guideline for Monitoring Stormwater Gross Pollutants (ASCE 2010) recognizes that for gross pollutants approaches of using EMCs, total loads or percent removal is not as straightforward as for sediments as sampling of the inflow and outflow mass of gross solids is challenging. At present, it is only possible to measure the volume or mass of the material actually removed, with no proven methods for comparing upstream and downstream loadings since gross solids are measured as a total mass rather than a concentration.

It is also important to note that some protocols are used to simply verify manufacturers’ claims, but others also provide different levels of certification according to the level of testing that the devices have been subjected to. In some cases, laboratory or limited field data is enough for the first level of certification, with more extensive data collection and analysis required for higher levels of certification.
5. SURVEY ON PERCEPTIONS OF GPT PERFORMANCE

An anonymous online survey was conducted across Australia as a preliminary investigation on the perceptions of council personnel and cleaning contractors involved in the maintenance of GPTs. Participants were invited to provide their opinions of the performance of different groups of GPT, maintenance frequencies, perceptions of performance and any other comments.

From an initial list of 25 council personnel and cleaning contractors supplied by the SIA, 21 took part in the survey, with 19 full responses and two partial responses. Survey results have been aggregated to the national level, so idiosyncrasies of different states and territories are not considered.

Survey respondents revealed a wide range of opinions on several issues, particularly in their perception of GPT performance. The responses reveal disagreement on the performance of the different GPT groups, particularly between council employees and contractors.

5.1. Methodology

An online survey was created to evaluate the perceptions of council personnel and cleaning contractors on the performance and maintenance of GPTs. The survey comprised two sections:

(a) Part 1 asked respondents about their background:
   - State they were based
   - Position they held at council/contractor company
   - Years of experience
   - Current association with any GPT manufacturer

(b) Part 2 asked the respondents a range of question regarding the perceived performance and maintenance requirements of their different devices, including:
   - Ownership of cleaning equipment and use of standard reporting forms
   - Approximate number of devices each respondent dealt with
   - Training provision by GPT manufacturers
   - Cleaning frequency
   - Level of performance of each GPT group
   - Preferred devices to work with
   - Views on a number of operational issues such as blocking, presence of moving parts, access, flooding, bypass, bacterial breakdown, etc.

As mentioned previously, there is a wide range of GPTs, with different operating principles, sizes and associated performance, and following advice from the SIA panel, the review was focused on devices which are designed to capture gross pollutants and coarse sediment.

To facilitate data collection, the survey aggregated the various GPTs in the following groups:

1. Difference in specific gravity traps (systems which use gravity to separate pollutants that float and that settle without the use of screens by incorporating baffles/booms in (a series of) chambers)
2. Direct screening (devices that incorporates screens in various orientations to the flow and which are not self cleansing)
3. Vortex type devices (devices that direct flow to produce vortices/hydrodynamic separation, but do not have a screen)
4. Continuous deflective separation (devices that combine a vortex/hydrodynamic separation with a non-blocking screening system)

5. Others

The authors and the SIA panel recognize that some devices could be considered to fall into more than one category, but the distribution into groups is mainly to facilitate data collection in the survey. No list of devices or any classification was provided to respondents, and some respondents raised the issue that some of their devices could fit into more than one category. Any further developments in terms of a testing protocol have to address the classification of devices based on different operating principles.

The survey was distributed online using Survey-Monkey software (http://www.surveymonkey.net/) and a copy of the full survey questionnaire is reproduced in Appendix A.

5.2. Results

5.2.1. Respondents Background

The 21 respondents that participated in the survey had a wide range of experience, varying between 3 and 25 years of involvement with GPTs, with the average number of years working with GPTs being 11.5 years. The respondents can also be divided among council (12 respondents) and cleaning contractors (9); two council employees and one cleaning contractor indicated a current association with a manufacturer. The distribution across states is shown in Table 5.1, with majority of respondents from NSW.

The roles of the respondents in their organization were varied, with the following reported roles:

- Manager (11)
- Stormwater Asset Planner (1)
- Director (2)
- Business Development (1)
- Stormwater Engineer (3)
- Supervisor (1)
- Waterways Rehabilitation Officer (1)
- Other (1)
Table 5.1. Distribution of respondents per state

<table>
<thead>
<tr>
<th>State</th>
<th>Respondents*</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT</td>
<td>1</td>
<td>4.8</td>
</tr>
<tr>
<td>NSW</td>
<td>11</td>
<td>52.7</td>
</tr>
<tr>
<td>NT</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>QLD</td>
<td>3</td>
<td>14.3</td>
</tr>
<tr>
<td>TAS</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SA</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>VIC</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>WA</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*One respondent (contractor) indicated that he works across 3 states

5.2.2. Maintenance

In total, the respondents estimate that they are responsible for the maintenance of 1356 devices across the 5 GPT groups, with the “other group” devices reported as Baramy traps, Nettech devices, nets and booms, sand filters and wetlands (although some of these devices do not fit the definition adopted in the report). The numbers of devices in each group is shown in Table 5.2. As it can be seen, all cleaning and contractors are involved in maintenance of direct screening devices, and at least 50% of councils and 75% of contractors are involved in the maintenance of all other groups (except others). It is interesting to note that councils report a significant proportion of direct screening devices, while cleaning contractors are more involved in cleaning vortex and continuous deflective separation devices.

Table 5.2. Number of devices that the respondents are involved in cleaning and maintaining.

<table>
<thead>
<tr>
<th>GPT Group</th>
<th>Council Respondents with device %</th>
<th>Contractors Respondents with device %</th>
<th>Council Number of devices %</th>
<th>Contractors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>6</td>
<td>63.6</td>
<td>115</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>8</td>
<td>100</td>
<td>497</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>6</td>
<td>54.5</td>
<td>36</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>5</td>
<td>72.7</td>
<td>141</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>2</td>
<td>27.3</td>
<td>45</td>
</tr>
</tbody>
</table>

Total 834  Total 522

1 - % of total respondents, 2 - % of total devices

All respondents reported that they fill a standard format for reporting, with 90% of councils and 87.5% of contractors filling a standard internal form, and 3 contractors reporting filling one internal and one external report form. Not surprisingly, all contractors use “in house” equipment during the cleaning operation, while 45% of council respondents indicate the use of “in house” equipment, with the remaining 55% using outsourced equipment. Table 5.3 shows the level of training supplied by the GPT manufactures and whether the respondents have read the Operation and Maintenance Manual for each device they perform maintenance on. Even though the percentages are slightly different, most of the council personnel and contractors have received training, with about half of those receiving training for all devices, while the other half only received training for some devices. Also, while the majority of the respondents have read the maintenance manual, a reasonable number had not.
Table 5.3. Number of respondents that have received training from GPT manufacturer and have read the Maintenance Manual

<table>
<thead>
<tr>
<th>Answer</th>
<th>Council</th>
<th>%</th>
<th>Contractors</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you received training from GPT supplier on how to clean the device?</td>
<td>No</td>
<td>2</td>
<td>18.2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Yes, for all types</td>
<td>4</td>
<td>36.4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Yes, for some types</td>
<td>5</td>
<td>45.5</td>
<td>3</td>
</tr>
<tr>
<td>Have you read the Operation and Maintenance Manual for each device you work with?</td>
<td>Yes</td>
<td>7</td>
<td>63.6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>4</td>
<td>36.4</td>
<td>2</td>
</tr>
</tbody>
</table>

1 - % of total respondents

The final set of question on maintenance enquired on what is cleaning frequency for the different GPT groups and what is the basis for determining the cleaning frequency. As shown in Table 5.4, 73% of the council respondents clean their devices within a specific period of time. However, half of those added that the cleaning was only triggered if necessary. The necessity was determined by different triggers, such as inspection following >10 mm rainfall in previous 48hrs, or % full as per monthly/fortnightly monitoring report. Almost all the contractors considered time and budget as influencing maintenance frequency, with two also reporting rainfall (> 5 mm) and % full based on regular inspection as triggers for maintenance. Only one contractor reported using a specific period of time as the basis for the cleaning frequency. In general, responses would indicate that half of the respondents perform maintenance based on different triggers, while the other half is either based on time (councils) or time and budget (contractors).

Table 5.4. Number of devices that the respondents are involved in cleaning and maintaining.

<table>
<thead>
<tr>
<th>Cleaning Frequency basis</th>
<th>Council</th>
<th>%</th>
<th>Contractors</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific period of time</td>
<td>8</td>
<td>72.7</td>
<td>1</td>
<td>16.7</td>
</tr>
<tr>
<td>Budget</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Both time and budget</td>
<td>3</td>
<td>27.3</td>
<td>5</td>
<td>83.3</td>
</tr>
</tbody>
</table>

1 - % of total respondents

The typical cleaning frequency for both council and contractors is also show in Figure 5.1 and Figure 5.2. It is clear from the figures that council maintenance intervals are comparatively shorter for all GPT groups, with the majority of council responses indicating maintenance periods smaller or approximately equal to 3 months. In comparison, most contractors report cleaning periods between 3-6 months or even longer than 6 months. As most contractors clean to time and budget constraints as opposed to council which seem to be solely tied to specific periods of time, these two different approaches are probably the explanation for the different patterns in Figure 5.1 and Figure 5.2.
Figure 5.1 Typical cleaning frequency for each GPT group according to council employees

Figure 5.2 Typical cleaning frequency for each GPT group according to cleaning contractors
5.3. Performance

The survey also asked the respondents for their perception on the performance of the different groups of GPTs, as well as a series of open questions:

- In your opinion which is the easiest type of device to clean and why?
- In your opinion which is the hardest type of device to clean and why?
- Which devices suffer most from blockage?
- For the “group” of devices you thought were the most effective, is there a reason for this?

An analysis of the open responses for these questions revealed a wide range of opinions among the respondents regarding the performance of the different GPT groups. For instance, some devices which were cited as the easiest to clean by some respondents (35%) in the first question were also listed as the hardest device by other respondents (18%). Most GPT groups were cited both as the easiest and the hardest device to clean by different respondents.

The question on which GPT group is most affected by blockage revealed a wide range of responses, with some respondents indicated none while others responded all GPT groups being affected by blockages. Most GPT groups were indicated by at least one respondent as being affected by blockage, including the self cleaning devices. The only group councils and contractors cite as having a high number of devices affected by blockage are the direct screening devices and trash racks by council, possibly as they are a significant proportion of the assets as shown in Table 5.2.

What the variation between the answers of the respondents highlights is the different perception of the performance of the different systems in the field. Nonetheless the variety of opinions could indicate that the systems are actually performing at various degrees of efficiency, which reinforces the need for standardization in testing and maintenance protocols in an attempt to minimize the differences in efficiency between the systems and guarantee that they meet the operational objectives.

The performance perceptions are also shown in Figure 5.3 and Figure 5.4, with council respondents rating the performance of all groups of devices somewhat higher than the contractors. Even though they differ on their assessment of the overall performance, both groups indicate the self cleaning devices represented by the continuous deflective devices and vortex devices as the best performing. The specific density traps also have a similar assessment by council employees and contractors, with a more widespread opinion between very good and poor (council) and good to poor (contractors). Moreover, it should be noted that once more there is some divergence between opinions on the performance of the different groups. It is unclear whether these differences are related to real performance, and if so, if they are caused by the device design, design of the treatment train and appropriateness of the device, installation or maintenance issues.
Figure 5.3 Perception of performance for each GPT group according to council employees

Figure 5.4 Perception of performance for each GPT group according to cleaning contractors
5.4. Other Issues

The final two questions of the survey were open questions that asked respondents to identify GPT issues, which are summarised in Table 5.5. The last question of the survey also asked for any other comments the respondents would like to make.

Table 5.5. Response count for some issues associated with GPTs

<table>
<thead>
<tr>
<th>Issue</th>
<th>Response Count</th>
<th>Comments summary (# responses in brackets)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Screens blocking</td>
<td>7</td>
<td>Regular inspection need to avoid blockage (6), nets have tendency to block (1)</td>
</tr>
<tr>
<td>2. Moving parts</td>
<td>5</td>
<td>To be avoided (3), collapsible inlet weirs need regular maintenance (2)</td>
</tr>
<tr>
<td>3. Baskets</td>
<td>7</td>
<td>Easy to damage (1), difficult to clean (1), prone to resuspension (1), silt up quickly (1), need regular maintenance (1), are good if self draining (1) and short life span (1)</td>
</tr>
<tr>
<td>4. Lids</td>
<td>7</td>
<td>Lids need to be light to facilitate access (2), internal bars prone to rust (1), need large diameter (2), need regular cleaning (1) and not an issue (1)</td>
</tr>
<tr>
<td>5. Dewatering</td>
<td>6</td>
<td>Can be expensive and difficult (2), water quality checking and appropriate disposal needed (1) and not an issue (3)</td>
</tr>
<tr>
<td>6. Access</td>
<td>5</td>
<td>Critical consideration (3), however position for large vehicles is overlooked at times (1), need for OHS compliance (1)</td>
</tr>
<tr>
<td>7. Confined spaces</td>
<td>7</td>
<td>Difficult and preferably avoided by design (3). Maintenance personnel trained for confined spaces (3), not an issue (1)</td>
</tr>
<tr>
<td>8. Ease of screen cleaning</td>
<td>7</td>
<td>Important (5), easiness depends on device (2)</td>
</tr>
<tr>
<td>9. Regular clean vs. annual clean</td>
<td>8</td>
<td>All agree that regular, frequent cleaning is essential (8)</td>
</tr>
<tr>
<td>11. Bacterial breakdown/ odour</td>
<td>6</td>
<td>Needs to be monitored (1), depends on frequency of cleaning, location and season (3), not an issue (1)</td>
</tr>
<tr>
<td>14. Evidence of bypass or flooding</td>
<td>4</td>
<td>There is evidence of bypass, disagreement on the frequency</td>
</tr>
</tbody>
</table>

As observed in the other question where additional input was requested, there was a variety of issues that were raised, with some issues where most respondents are in agreement, where for other issues the responses are varied. It is important to note that for these question also there was a limited number of answers compared to the other open questions.

Three issues that have a good agreement between respondents are the occurrence of odour and bacteria breakdown the need for regular cleaning and that regular cleaning is need to avoid blockage. Most of the responses (5/6) on the question of odour and bacterial breakdown reported that it is an issue that requires attention, as it can occur at particular location at certain season, and it can be minimized by frequent cleaning. This support the finding of some studies in the literature review which found evidence of nutrient leaching due to bacterial breakdown.

When asked on regular clean versus annual clean, only one respondent named it as not an issue, with remaining 8 respondents naming it as very important. This is in direct contrast to some of the frequency reported by contractors, where the frequency is relatively low. Most of
the responses regarding screen blocking (6/7) agree that regular cleaning is essential to reduce the frequency of screen blocking.

In the final question for open comment, only 6 participants responded, with responses including:

- Many sites have GPTs retro-fitted to existing drains and creeks. Many could have been better designed with the input of the cleaning contractor as the thinking or practicality allowing for clean-out is poorly thought through on quite a number of occasions.

- There should be more consultation with the groups that do the maintenance of the devices to see what problems may occur with the different devices before they are installed.

- Our objectives are to use the most efficient/economical means of removal of litter/large pollutants, and sediment to a lesser extent.

- A regular (scheduled) maintenance program, an effective & efficient work method for extracting waste from/cleaning GPTs, utilising the most appropriate plant and equipment, and having properly trained staff are all essential for successfully maintaining GPTs and enhancing their performance/effectiveness.
6. CONCLUSIONS

Useful guidance for GPT selection is limited by the large number of devices, a lack of reporting protocols and standard methods and only a small number of detailed monitoring studies. Local government, who is responsible for the implementation and management of stormwater infrastructure, is dependent on in-house expertise and manufacturer’s advice in selecting appropriate stormwater treatment strategies.

Even though the Australian Runoff Quality (Engineers Australia 2006), highlights the need for monitoring of flow and the amount of gross pollutants upstream and downstream of the GPT, most field studies still rely on material collected during maintenance due to the high cost and logistics associated with field sampling upstream and downstream of the units. Several field studies agree qualitatively on the assessment that the GPTs tested were highly efficient in capturing gross pollutants (Allison et al. 1998; Engineers Australia 2006; Birch and Matthai 2009). Again, this is in contrast with the survey results which indicate a perception of some systems not performing well. It is possible that the case for this disparity is that the survey considers a wide range of devices; whereas the studies usually focus on one or two devices or technologies. Also, field studies have not captured system performance over a long period of time, and as such, may not be capturing decaying performance over time. Such possible change in performance is likely to be noticed by survey respondents who are involved in long term maintenance.

What the variation between the answers of the respondents highlights is the different perception of the performance of the different systems in the field. It is unclear whether the different perceptions are actually linked to the actual performances of these devices. Nonetheless the variety of opinions could indicate that the systems are actually performing at various degrees of efficiency, which reinforces the need for standardization in testing and maintenance protocols in an attempt to minimize the differences in efficiency between the systems and guarantee that they meet the operational objectives.

To overcome the high cost and the inherent intra-variability in field studies, evaluation of GPT performance is also tested in the laboratory or in modelling studies. A large number of laboratory studies are somewhat limited, either by using a limited number of replicates and a very small load to determine performance or by using scale down versions of pollutants. The use of scaled down gross pollutants is not well understood and unfortunately, laboratory studies do not attempt any validation in a full scale model. The same can be said for using modelling to simulate performance, where the use of CFD modelling must be accompanied by experiments to validate the initial model, as predictions are reasonably sensitive to some model parameters (Egarr et al. 2004), and it is unclear how to properly represent gross pollutants in a CFD model.

The reported removal efficiency of suspended sediments by GPTs is highly variable due to differences in methodologies across several studies, with use of indirect and direct sampling methodologies, or measurement of sediment concentration based on Total Suspended Solids (TSS) or Suspended Solids Concentration (SSC). In addition, calculations based on different methods such as EMCs or load based could impact the calculated efficiencies which make comparison of results difficult. Recent studies suggest that physically based performance function based on a Peclet number can potentially be used to upscale performance. However, the tests were performed at relatively high concentrations, so the suitability of this approach to predict performance and scale up should be tested using different range of concentrations as the efficiency is directly linked to the flow rates and concentration used in experiments and/or field testing. Sediment also can be remobilised through scouring at high flow rates in some swirling devices.

At present, there are no standard methods or guidelines for the testing of GPTs or other stormwater treatment devices in Australia, but in the US several different guidelines and protocols for GPTs and stormwater treatment devices currently exist, both for laboratory and field testing. The current US third party verification protocols are not consistent on their requirements, and it has been suggested that the protocols have had mixed results due to the inconsistency in reporting methodologies and appropriate guidelines. The establishment
of protocols allows for reduction of bias and a defensible level of statistical confidence in the analysis and comparison of the performance of the different devices. It also accelerates technology adoption at reduced costs for both states and manufactures and avoids the need for regional or local protocols.

Based on the literature review, despite their differences in test requirements, methods, etc, all protocols currently in use in the US basically agree that data must be Quality Assured (QA) and Quality Controlled (QC) and subject to appropriate statistical tests to verify that the performance measures are statistically significant. Moreover, performance claims and specific conditions must be stated, as tests performed under optimal conditions may not represent typical stormwater conditions, so tests should provide detailed test methodologies and conditions. Finally, third party studies are preferred, and all protocols required involvement of a third party in at least one part of the process.

It is important to note that the focus of most protocols in the United States is on suspended sediment. Although the ASCE has published a guideline for monitoring gross pollutants, it recognizes that for gross pollutants approaches of using EMCs, total loads or percent removals are not as straightforward as for sediments as sampling of the inflow and outflow mass of gross solids is challenging. At present, it is only possible to measure the volume or mass of the material actually removed, with no proven methods for comparing upstream and downstream loadings since gross solids are measured as a total mass rather than a concentration.

The use of the protocols by various agencies is also different, with some agencies simply using them to verify manufacturers’ claims, but others also provide different levels of certification according to the level of testing that the devices have been subjected to. In some cases, laboratory or limited field data is enough for the first level of certification, up to extensive data collection and analysis for higher levels of certification.
7. RECOMMENDATIONS

Based on the review of published studies on the performance of GPT in the field and laboratory, the survey of GPT users and the review of the protocols used in the US, several issues must be discussed during the development of an Australian protocol. The protocol must be developed by an assembly of an industry-wide review panel. The development of any protocol will only be successful if it is accepted by manufacturers, councils and regulatory agencies, so all these parties must be part of the committee developing the guidelines. Also, it should include other third parties such as universities and research organizations, as they are likely to play a role in reviewing the studies. The committee developing definite guidelines must have representation from all states to ensure that development of localized versions is avoided, so the adoption of technology is accelerated and the cost of testing is kept to a minimum.

It is suggested that the following issues should be considered in developing the GPTs testing protocols:

- Third party assessment: Testing of devices could either involve an independent third party or be entirely conducted by an independent party.
- It is proposed that testing should include performance measurement at a set number of flow rates and different screen blockage conditions, even if systems are self-cleaning. Consideration should be given to non-ideal condition as lack of frequent maintenance was highlighted in the survey as a potential issue; also unusual loads due to large storm events may lead to temporary non-ideal conditions.
- Flow rates and concentrations: As previous studies demonstrated that low flow rates or high concentrations yield higher removal rates, it is recommended that the protocols include a minimum number of flow rates (as % of design flow rate) and the associated concentrations for gross pollutants and suspended sediment. The use of set flow rates will demonstrate performance for a range of flow rates to verify screen blocking and the onset of self cleaning flow patterns in some devices. Use of concentrations ranges between minimum and maximum concentrations typically found in the field would demonstrate performance across a varied range of concentrations, indicating the performance of a device at low levels and not only at higher concentrations when the performance is higher.
- Hydraulic testing: Methodologies for hydrologic testing and scale up are well understood and are part of testing. It is therefore recommended that the protocols only determine a standard set of (minimum) testing and the appropriate dimensionless groups for scale up of hydraulic laboratory scale tests.
- Scale of test model: As relationships between scaled up pollutants are not well understood, test on a full scale device (smallest in the range) is preferred. If the test is performed on a laboratory device, it is recommended that the scale-up issues are discussed as part of the test results.
- Device classification: As different devices have (sometimes markedly) different operating principles, particular aspects of the standard may not be applicable to a particular device. As such, the standard could divide the different devices and define different sets of criteria to allow a device to be classified in a determined group. As an example, devices which claim to be non-blocking could undergo testing to demonstrate that they are indeed non-blocking.
- Testing for intended purpose: It is proposed that testing should be only required for removal of gross pollutants (> 5 mm) and coarse sediment (> 500 μm). If a manufacturer wants to claim removal of other pollutants, the protocol should make allowance for the appropriate test for each pollutant to which a performance claim is intended to be made.
• Quality control and assurance: Quality assurance and quality control protocols are an important part of every study and the protocol could provide guidance on requirements to ensure data quality and statistical significance.

• Adoption of a tiered system: In recognition of the high cost of testing units, a tiered system can be considered. In such a tiered system, manufacturers could achieve different levels of certification according to their testing. In such a system, achievement of a certain certification does not necessarily indicate which devices perform better, but which device has been more widely tested. It is up to the adopting parties (i.e. councils) to decide on the level of recognition required for a device to be adopted. The tiered system (as an example) could be composed of three levels:
  
  o Level 1: Performance claims based on limited studies that are not reviewed by an independent third party. This would include all past testing (which did not involve an independent third party) and allow most manufacturers to attain level 1 based on prior testing upon the public release of reports detailing performance tests.
  
  o Level 2: Test performance based on either field and/or laboratory test performed by (or in conjunction with) an independent third party following an Australian protocol developed by the guideline committee, with appropriate data collection and quality control.
  
  o Level 3: In addition to level 2, this could involve long term monitoring of performance of the unit(s) in the field to evaluate their long term performance.

• Standard gross pollutant composition for testing: to allow for comparison of results between different devices, a standard gross pollutant sample could be developed. The sample standard composition of materials would represent typical compositions in field sites and include minimum and maximum sizes for the different materials. The determination of the standard would also include additional research required to determine validity of the use of scaled gross pollutants and in which conditions such assumptions are valid and appropriate.

• Standard particle size distribution for sediment testing: to allow a better comparison of sediment trapping performance between devices, it is recommended that the protocol adopt a standard sediment composition and grading typical of Australian condition. Even though the focus of the protocol is on coarse sediments, the standard size distribution should include fines as they will influence system performance. The efficiency of the devices then can be assessed against different fractions.

• Scour testing: As high flow rates have the potential to resuspend material, the potential for resuspension of captured pollutants could be incorporated in the protocol. Such an assessment would use different flow rates and different initial conditions within the device (e.g. pre loads of 50 and 100% of loading capacity) to indicate potential resuspension of material.

• Efficiency calculation method: The literature review has shown how different efficiency calculation methods can yield different efficiencies for the same dataset. Guidance in the protocol on recommended or accepted methods would ensure comparability within different studies.

• Sampling methods: Guidance is needed on appropriate sampling methods for gross pollutants and sediments. The discussion could focus on direct and indirect sampling, sample collection, composite sampling versus grab sampling, flow weighted concentrations and treatment. Adoption of a pre-defined set of methods is not recommended as they may not be suitable to new emerging technologies. Instead, any sampling protocol developed by a manufacturer should be reviewed and approved by an independent third party prior to testing to ensure they are appropriate.
- Guidance on modelling: The adoption of computer modelling to reduce costs and allow integration with other software is recommended, provided that guidance is also given for model calibration and validation, due to model sensitivity to a large range of parameters.

- Additional research: It is recommended that the committee developing the protocols also commissions research to investigate the following topics:
  - Performance of GPT in nutrient removal, particularly the conditions leading to the release of nutrients in dry weather flow conditions;
  - The development of suitable scale up functions of efficiency, based on dimensionless numbers such as Peclet numbers or other variations.
REFERENCES


Engineers Australia, 2006. Australian Runoff Quality, The Institute of Engineers Australia.


MDEP, 2008. Massachusetts Stormwater Handbook, Massachusetts Department of Environmental Protection


GLOSSARY

ASCE - American Society of Civil Engineers
BMP - Best Management Practices (BMP)
CFD - Computational Fluid Dynamics
CULD - Conditional Use Level Designation
EMC – Event Mean Concentration
ER – Efficiency Ratio
ETV - Environmental Technology Verification
EWRI - Environmental Water Resources Institute
GPT - Gross Pollutant Traps
GULD - General Use Level Designation
ISL - Efficiency of Individual Storm Loads
NJCAT - New Jersey Corporation for Advanced Technology
NJDEP - New Jersey Department of Environmental Protection
PM - Particulate Matter
PSD - Particle Size Distribution
PULD - Pilot Use Level Designation
ROL – Regression of Loads
SIA - Stormwater Industry Association
SOL – Summation of loads
SSC - Suspended Sediment Concentration
TAPE - Technology Assessment Protocol – Ecology
TARP - Technology Acceptance Reciprocity Partnership
TSS – Total Suspended Solids
USEPA - United States Environmental Protection Agency
APPENDIX A – SURVEY QUESTIONNAIRE

Dear GPT Practitioner

The Stormwater Industry Association (SIA) has engaged CSIRO to undertake some research as part of the first stage of development of a national Gross Pollutant Trap (GPT) Testing Standard. The project which commenced in August involves:

- the collection and review of data from all the main GPT proprietors and a review of the definitions and terminology used in the industry
- a review of national and international literature to summarise existing knowledge of GPTs, their function, design and performance and identify knowledge gaps
- an online survey of practitioners and major stakeholders involved with GPT performance and maintenance

The SIA is interested in using the information gathered through this study to develop improved guidelines for GPTs to encourage their uptake in stormwater management. Developing a standardised approach to evaluate and communicate the performance of a range of stormwater treatment devices is one that has been discussed within the industry for a number of years. Given the current state of limited knowledge and increasing use of stormwater as a resource, the interest on the adequate management and treatment of stormwater is likely to continue to grow. The function of GPTs in the stormwater treatment train and treatment objectives also has the potential to be expanded within this context.

As such, we would like to invite you to participate in an online survey aimed at gathering the input of practitioners and major stakeholders in GPT performance and maintenance. The questionnaire is comprised of 21 questions and requires a maximum of 20 min to be completed. Please note that additional comments/remarks can be made at the end of the survey.

The data being collected through the survey will be anonymous and individual comments will not be made available to any external party. No personal details will be collected from you. The information you provide to us via survey will be used to write a general report on views regarding GPT cleaning, maintenance and performance. The information will also be used to prepare manuscripts for academic publication although your personal information will not be identifiable at any stage of the writing process. A summary of the findings from the survey will be available through the SIA website.

Participation in this study is completely voluntary and you are free to withdraw at any time without prejudice or penalty. If you wish to withdraw your participation from the study, simply notify the researcher via the contact details below and you are free to leave. The survey information that you have provided up to that point will be deleted if requested and will not be included in the study unless you give us permission to use that information.

This study has been cleared in accordance with the ethical review processes of CSIRO and within the guidelines of the National Statement on Ethical Conduct in Human Research. If you have any questions concerning your participation in the study feel free to contact the researchers involved. Further information on your rights as a research participant in CSIRO research is also available at http://www.csiro.au/resources/human-research-ethics-brochure.html

Your participation is very valuable to us. Companies that clean GPTs have firsthand experience as to what works, why, and what are the problems with different devices. We thank you in advance for your valuable contribution and time input.

Kind regards,

Dr Luis Neumann
Research Scientist
Integrated Urban Water Systems Theme
CSIRO Land and Water
Phone: +61 3 9252 6224
luis.neumann@csiro.au
GPT Survey

Default Section

* Please identify your State/Territory

☐ ACT ☐ NSW ☐ NT ☐ QLD ☐ TAS ☐ SA ☐ VIC ☐ WA

What is your role?


How many years have you been involved with maintaining/cleaning GPTs?


Do you have a current association with any of GPT manufacturer?

☐ Yes
☐ No

* Are you:

☐ A council employee
☐ Cleaning contractor
Is your cleaning equipment typically in house or outsourced?
- In house
- Outsourced

Do you have a standard format for reporting on the cleaning?
- No
- Yes, standard internal form
- Yes, client supplied form

Which of the following “groups” of GPTs does your Council have?
- 1. Difference in specific gravity traps (systems which use gravity to separate pollutants that float and that settle without the use of screens by incorporating baffles/booms in a series of chambers)
- 2. Direct screening (devices that incorporate screens in various orientations to the flow and which are not self-cleaning)
- 3. Vortex type devices (devices that direct flow to produce vortices/hydrodynamic separation, but do not have a screen)
- 4. Continuous detective separation (devices that combine a vortex/hydrodynamic separation with a non-blocking screening system)
- 5. Other (please specify)

Have you ever receive instruction or training from GPT suppliers in how to clean each type of GPT?
- No
- Yes, for all types
- Yes, for some types

Have you read the Operation and Maintenance Manual for each device owned by your council?
- Yes
- No
GPT Survey

Approximately how many GPTs do you have for each of these “groups” listed below?

1. Difference in specific gravity traps
2. Direct screening
3. Vortex type devices
4. Continuous defective separation
5. Other (please specify)

What is the typical cleaning frequency?

<table>
<thead>
<tr>
<th></th>
<th>less than 3 months</th>
<th>approx 2 months</th>
<th>3 to 6 months</th>
<th>more than 6 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference in specific gravity traps</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Direct screening</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Vortex type devices</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Continuous defective separation</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Other (please specify)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

Would you like to provide further details on the previous two questions? E.g. number of different devices in each of the “groups”, different cleaning frequencies for different devices, etc. (optional)

What is the cleaning frequency typically based on:

- ○ Specific period of time
- ○ Budget
- ○ Both time and budget
- ○ Other (please specify)

In your opinion, are all the devices operating at a similar level of performance?

- ○ Yes
- ○ No
**GPT Survey**

In your opinion, what is the performance level for each GPT group?

<table>
<thead>
<tr>
<th></th>
<th>Very good</th>
<th>Good</th>
<th>Average</th>
<th>Poor</th>
<th>Very poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference in specific gravity traps</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct screening</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vortex type devices</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuous effective separation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (please specify)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In your opinion which is the easiest type of device to clean and why?

In your opinion which is the hardest type of device to clean and why?

Which devices suffer most from blockage?

For the “group” of devices you thought were the most effective, is there a reason for this?
Is your cleaning equipment typically in house or outsourced?
- [ ] In house
- [ ] Outsourced

Do you have a standard format for reporting on the cleaning?
- [ ] No
- [ ] Yes, standard internal form
- [ ] Yes, client supplied form

Which of the following “groups” of GPTs your company is involved in cleaning?
- [ ] 1. Difference in specific gravity traps (systems which use gravity to separate pollutants that float and that settle without the use of screens by incorporating baffles, booms in a series of chambers)
- [ ] 2. Direct screening (devices that incorporate screens in various orientations to the flow and which are not self-cleaning)
- [ ] 3. Vortex type devices (devices that direct flow to produce vortices/hydrodynamic separation, but do not have a screen)
- [ ] 4. Continuous deflected separation (devices that combine a vortex/hydrodynamic separation with a non-blocking screening system)
- [ ] 5. Other (please specify)

Have you ever receive instruction or training from GPT suppliers in how to clean each type of GPT?
- [ ] No
- [ ] Yes, for all types
- [ ] Yes, for some types

Have you read the Operation and Maintenance Manual for each device you clean?
- [ ] Yes
- [ ] No
GPT Survey

Approximately how many GPTs of “groups” listed below do you clean?

1. Difference in specific gravity traps
2. Direct screening
3. Vortex type devices
4. Continuous defective separation
5. Other (please specify)

What is the typical cleaning frequency?

<table>
<thead>
<tr>
<th></th>
<th>less than 3 months</th>
<th>approx 3 months</th>
<th>3 to 6 months</th>
<th>more than 6 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference in specific gravity traps</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct screening</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vortex type devices</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuous defective separation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (please specify)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Would you like to provide further details on the previous two questions? E.g. number of different devices in each of the “groups”, different cleaning frequencies for different devices, etc. (optional)

What is the cleaning frequency typically based on:

- Specific period of time
- Budget
- Both time and budget
- Other (please specify)

In your opinion, are all the devices operating at a similar level of performance?

- Yes
- No
In your opinion, what is the performance level for each GPT group?

<table>
<thead>
<tr>
<th>Type of Device</th>
<th>Very good</th>
<th>Good</th>
<th>Average</th>
<th>Poor</th>
<th>Very Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference in specific gravity traps</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct screening</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vortex type devices</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuous defective separation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (please specify)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In your opinion which is the easiest type of device to clean and why?

[Blank space for answer]

In your opinion which is the hardest type of device to clean and why?

[Blank space for answer]

Which devices suffer most from blockage?

[Blank space for answer]

For the “group” of devices you thought were the most effective, is there a reason for this?

[Blank space for answer]
These last two questions were directed at contactors and council employees.