STORMWATER AUSTRALIA

Stormwater Quality Improvement Device Evaluation Protocol

Field Monitoring

www.stormwater.asn.au

Acknowledgements

This document has been developed by Stormwater Australia and is broadly based on the Auckland Proprietary Device Evaluation Protocol (PDEP).

In preparing this document Stormwater Australia acknowledges the input made by the Auckland PDEP contributors in advancing this important cause.

Stormwater Australia also acknowledges the time and technical contributions made by the members of the Stormwater Quality Improvement Device Advisory Committee (SQIDAC):

- Andrew Allan, Stormwater Australia
- Stephen D. Baker, Rocla/Fletcher Building Products
- Dr Darren Drapper, Drapper Environmental Consultants
- Alan Hoban, Stormwater Australia
- Andy Hornbuckle, SPEL Environmental
- Dr Charles Kelly, Humes/Holcim
- Andrew Macklin, Ecosol
- Chris Rochfort, Star Water
- Michael Wicks, Stormwater360

Stormwater Australia is grateful for feedback during consultation periods received from:

- Environment Protection Authority, South Australia - Mr Tony Circelli;
- The Victorian Minister for Environment, Climate Change and Water - Hon Lisa Neville MP;
- The Commonwealth Minister for the Environment - Hon Greg Hunt;
- WaterNSW - Mr Malcolm Hughes;
- NSW Office of Environment and Heritage - Dr Marlene van der Sterren;
- Stormwater Queensland Committee

WaterRA and Swinburne University have contributed to improving the protocol through an Independent Peer Review.

Stormwater Quality Improvement Device Evaluation Protocol (SQIDEP)

Citation: Stormwater Australia (2018) Stormwater Quality Improvement Device Evaluation Protocol. Version 1.2 December 2018.

Version	*Release Date*
1.02	*2017*
1.1	12 *Dec 2018*
1.2	21 *Dec 2018*
1.3	31 *Dec 2018*

Disclaimer

About This Protocol

The introduction of water quality objectives for receiving waterways, and subsequent policies for stormwater quality improvement in the late 1990s has driven the need for stormwater treatment devices, whether manufactured (proprietary) or custom-designed, vegetated assets. Beginning with research undertaken in Australia by the CRC for Catchment Hydrology, there have been many monitoring programs and laboratory tests on a variety of treatment measures since. However, due to the variable nature of stormwater events across Australia, and the unique characteristics of different measures, rarely have the methods been replicated.

Stormwater Australia initiated a Literature Review of performance testing approaches with CSIRO in 2010. In June 2014, following a forum at the National Stormwater Conference, Stormwater Australia convened a Stormwater Quality Improvement Device Advisory Committee (SQIDAC) to commence the development of a benchmark field testing protocol. The advisory committee was chaired by Stormwater Australia and open to anyone, initially consisting of industry representatives, due to their experience with international testing protocols, and product research & development.

Drawing on this experience, the SQIDAC selected the Auckland Regional Council Proprietary Device Evaluation Protocol (PDEP) to use as a solid foundation for the preparation of the Stormwater Quality Improvement Device Evaluation Protocol (SQIDEP). Following revisions to reflect Australian rainfall conditions and feedback from the Auckland process, a draft SQIDEP was released for public consultation in December 2014. Amendments resulting from the consultation process were included in a second draft released by the SQIDAC for further industry consultation in September 2015. Several workshops were held with industry in 2015 and 2016 to engage and receive further feedback.

Independent peer review of the SQIDEP provided further comments in November 2017. This first formal release represents the culmination of four years of development and consultation to develop a robust, scientifically-based, and industry-recognised standard.

The SQIDEP provides a uniform set of criteria to which stormwater treatment measures can be field-tested and reported. These criteria should guide and inform field monitoring programs that seek to demonstrate pollutant removals for treatment measures included in the Model for Urban Stormwater Improvement Conceptualisation (MUSIC) or any other pollutant modelling software that may be released. This version does not address laboratory testing, but it is anticipated for the next revision. The SQIDEP will be supported by an independent evaluation process overseen by Stormwater Australia to ensure claims are assessed and verified.

The SQIDEP should be recognised to be a living document. Environmental conditions are highly variable. International protocols have identified that providing benchmark criteria for performance assessment is necessary. However, some criteria may require refinement once results from monitoring programs are received. Stormwater Australia commits to regular review of the criteria based on sound science and evidence.

Contents

Tables

Figures

GLOSSARY

Acronym	Term	Explanation
ADT	Average Daily Trips	Traffic movement count.
AEP	Annual Exceedance Probability	Probability that a given rainfall total accumulated over a given duration will be exceeded in any one year.
	Aliquot	A portion of a larger whole, especially a sample taken for chemical analysis or other treatment. For the purposes of this protocol a discrete sub-sample collected from a qualifying storm event.
APHA	American Public Health Association	Reference organisation.
ARI	Annual Recurrence Interval	Frequency of storm event.
ARQ	Australian Runoff Quality	Document published by Engineers Australia providing guidance on procedures for the estimation of urban stormwater contaminants and associated design guidelines.
ASTM	American Society for Testing Materials	Reference organisation.
BoE	Body of Evidence	One evaluation route in the SQIDEP, incorporating existing data from international sites and/or laboratory testing.
	Calibration	Utilising monitoring data points to adjust certain parameters used for the sizing methodology to ensure its representativeness.
	Claimant	Designer, vendor or supplier of permanent Stormwater Quality Improvement Device.
CRE	Concentration Removal Efficiency	A mathematical ratio of the difference between an influent concentration and an effluent concentration. Expressed as a percentage. Quantifies the ability of a device to reduce the concentration of a contaminant in stormwater.
	Controlled Field Test	Tests on a full scale device installed in the field, using artificially-produced influent to mimic stormwater flows.
	Effluent (or Outflow)	Stormwater exiting a treatment device.
DQO	Data Quality Objectives	
Device	Stormwater Treatment Device	Any permanent, repeatable man made device, structure or system designed primarily for the improvement of stormwater quality.
DRP	Dissolved Reactive Phosphorus	Any form of P that reacts with reagents in a colorimetric test following filtration of the sample through a 0.45 μm filter paper.

Acronym	Term	Explanation
	Evaluation Panel	Independent panel set up to make final decision on whether to accept submitted claims regarding device performance.
EMC	Event Mean Concentration	Weighted average pollutant concentration that reflects varying runoff concentration over the duration of the hydrograph.
ER	Efficiency Ratio	A Performance metric describing removal efficiency by comparing the difference between the average inlet concentration and the average outlet concentration.
ESA	Equivalent Standard Axles	Traffic movement count.
	Influent (or Inflow)	Stormwater entering a treatment device.
IET/ADP	Inter-event Time (also known as Antecedent Dry Period)	Time between a storm event's end and the subsequent event's beginning as designated by minimum time interval with no greater than 1mm of rainfall.
IQR	Inter Quartile Range	A measure of statistical dispersion, being equal to the difference between the upper and lower quartiles.
	Laboratory Tests – Scale Model	Tests undertaken in the laboratory on a scaled down model of the device. NOTE this data is not accepted for this SQIDEP.
	Laboratory Tests – Full Scale	Tests undertaken in the laboratory on a full scale model of the device. Future revisions of SQIDEP will include a Protocol for Laboratory Testing.
LPT	Local Pilot Trial	One evaluation route in the SQIDEP, requiring field installation and monitoring of full scale device performance.
MRE	Mass Reduction Efficiency	A mathematical ratio of the difference between the influent pollutant load by mass (i.e. concentration multiplied by flow volume) and the effluent pollutant load. Expressed as a percentage. Allows the total mass of contaminant captured by a device to be quantified.
NATA	National Association of Testing Agencies	Industry peak body responsible for certifying analytical agencies to ensure technical competence in undertake specific testing and analytical methods.
	Performance Metrics	Quantify pollutant removal capacity and consistency of treated effluent water quality.
PSD	Particle Size Distribution	Description of particle sizes (ranges) in stormwater flows.
QAPP	Quality Assurance Project Plan	Plan to show how performance testing in the field is undertaken in a way that ensures appropriate methods and procedures are followed.
RAE	Relative Achievable Efficiency	A performance metric. Expressed as a percentage. Determines pollutant removal relative to an irreducible minimum concentration or a water quality standard.
SFR	Specific Flow Rate	The flow rate through the device divided by the cross sectional area of the device.

Acronym	Term	Explanation
SQIDAC	Stormwater Quality Improvement Device Advisory Committee	An advisory committee reporting to the Stormwater Australia board.
SQIDEP	Stormwater Quality Improvement Device Evaluation Protocol	The testing protocol described in this document.
SSC	Suspended Sediment Concentration	A method for measuring sediment in stormwater according to a standard laboratory method (i.e. ASTM D3977-97 Test Method B or equivalent). Note some laboratories may refer to a similar method as a “low concentration TSS”. SSC is different to Total Suspended Solids (TSS) and results should not be used/reported interchangeably.
Tc	Time of Concentration	A measure of the response of a catchment to a storm event. It is the longest time required for water to flow from the most hydrologically remote point in a catchment to the catchment outlet. It is a function of the topography, geology, and land use within a catchment.
TFR	Treatable Flow Rate	The maximum flow rate treated by a device before bypass commences. The design TFR may be informed by the field test results.
TKN	Total Kjeldahl Nitrogen	The sum of organic nitrogen, ammonia (NH3), and ammonium (NH4+) in a sample.
TN	Total Nitrogen	The sum total of organic and oxidised nitrogen species (NOx).
TP	Total Phosphorus	Sum of organic and inorganic forms of phosphorus in unfiltered water samples.
TSS	Total Suspended Solids	A method for measuring sediment in stormwater according to a standard laboratory method (e.g. APHA (2005) 2540 D). Should not be used/reported interchangeably with SSC. The terminology is distinctly different from Gross Pollutants and should not be misinterpreted as such.
USEPA	United States Environmental Protection Agency	Reference agency.
	Validation	Utilising known data points to confirm a result or prediction.
VPD	Vehicles Per Day	Traffic movement count.
WERF	Water Environment Research Foundation	(United States) Industry body and reference organisation.

1 Introduction

The purpose of this protocol is to describe best practice procedures to evaluate the performance of stormwater quality improvement devices under field conditions. The SQIDEP and evaluation process will enable stormwater industry will be able to have increased confidence in the pollutant reduction claims associated with treatment measures tested in accordance with this protocol.

As stormwater quality objectives vary across Australia and are updated from time to time, this protocol does not seek to validate performance against a specific set of water quality targets. Rather it provides a robust set of criteria to enable consistent, repeatable assessment of water quality performance that can be compared against the targets in various jurisdictions.

This protocol describes the requirements for a Quality Assurance Project Plan and the reporting and evaluation of results.

This protocol is not intended to address:

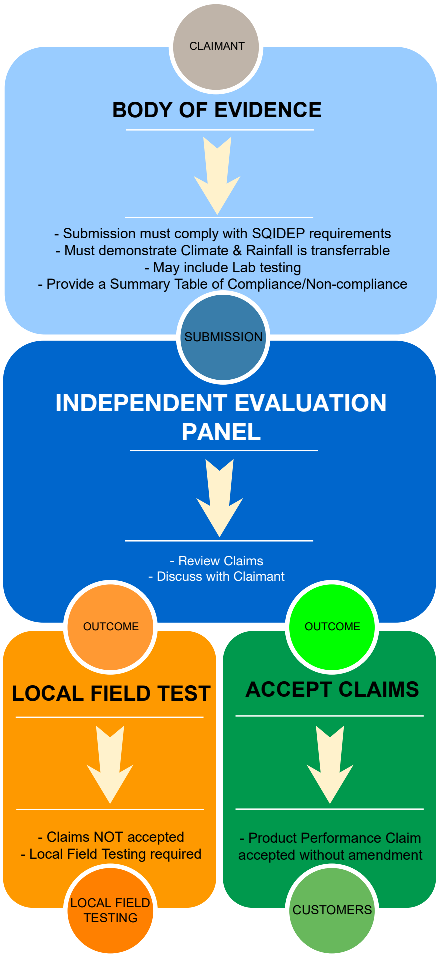
- Hydraulic performance characteristics such as head loss
- Laboratory testing practices (future revisions of SQIDEP are expected to include a laboratory testing protocol).

2 Process Overview

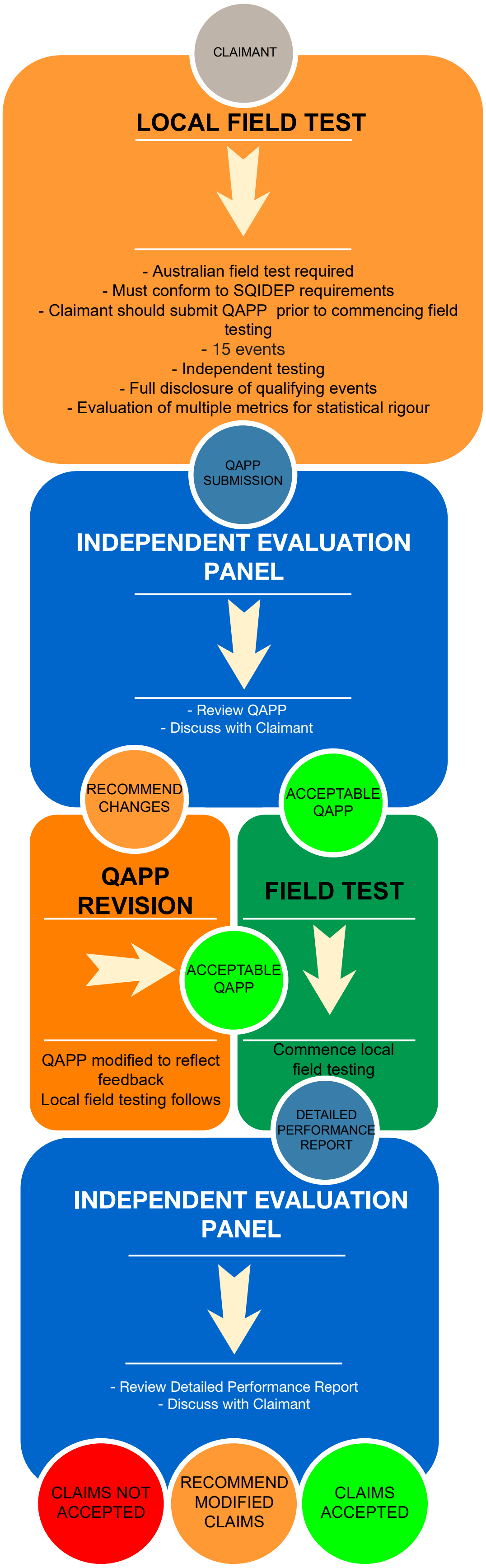
The following diagram shows the process for performance results to be considered under either a Body of Evidence or Field Evaluation pathway.

This protocol deals primarily with aspects of the Field Evaluation pathway.

The BoE pathway may be chosen if sufficient international field testing has been completed and can demonstrate compliance with the SQIDEP criteria and Australian climatic & rainfall conditions. Applicants seeking the BoE pathway are recommended to confirm with Stormwater Australia/the Independent Evaluation Panel prior to making a full submission.



Body of Evidence Pathway



Field Evaluation Pathway

Figure 1 — Assessment Pathways

3 Prior to Commencing Field Testing

Prior to undertaking a detailed field evaluation, it is important that there is sound evidence available to support the design of the trial. This may include laboratory testing (especially using a validation framework which included a range of full scale and challenges tests) or a solid theoretical base outlining key treatment mechanisms and credible scale/ pilot laboratory tests.

It is up to the claimant to identify a suitable trial site and convince the site operator of the veracity of the claims in these negotiations. Under this protocol the claimant can be afforded a reduced performance claim (as a precaution against ambit claims to enter and become established in the market) and two years to collect data and undertake analysis to prove the claim (unless an extension of time is justified). Claimants should consider that the design of the field tested device (eg. sizing relationships, vault volume, etc) may be reflected in minimum conditions and/or criteria, should the claim be accepted.

If the claimant can demonstrate the theoretical, laboratory or field performance of the device based on a desktop assessment, then a Quality Assurance Project Plan should be developed to support more rigorous field trials.

4 Quality Assurance Project Plan

A Quality Assurance Project Plan (QAPP) must be prepared prior to field testing. Its purpose is to describe how performance testing will be conducted; ensuring appropriate methods and procedures are followed and documented so that data obtained during testing is valid for verification of the device's performance.

The claimant is responsible for preparing the QAPP. The plan may be revised as necessary throughout the course of field testing with adjustments, notes and explanation provided.

The QAPP should contain background information on the device being tested, project organisation, sampling design and methods, laboratory methods, field and laboratory quality control, data management procedures, data review, and reporting.

The QAPP is developed for planning the monitoring programme and ensuring that the proposed methodologies are executed in line with the contents of the QAPP, which are aligned with the protocols outlined in the SQIDEP.

The QAPP must be agreed to by the claimant who should also commit adequate resources to implement the recommended testing. The QAPP should be developed by a person with knowledge of the SQIDEP and a good understanding of field sampling and analytical chemistry methods. Where appropriate, it shall be developed in consultation with the analytical laboratories selected, especially if specialist analysis is required.

The QAPP shall be based on the claimant's Performance Claim, and shall contain the details of:

a. Data Quality Objectives.
b. Organisational roles and responsibilities.
c. Description of test site.
d. Measuring rainfall.
e. Storm events sampled.
f. Flow monitoring.
g. Sampling location.
h. Sampling equipment.
i. Sampling methodology.
j. Sampling Quality Assurance and Quality Control.
k. Laboratory analysis.
l. Laboratory Quality Assurance and Quality Control.
m. Data management.
n. Reporting.

For each of these items (a-n) the SQIDEP document provides further commentary, which should be considered. Where the commentary contains numerical requirements (e.g. numbers of samples) if a discrepancy exists, Table 3 takes precedence

The QAPP shall describe the procedures that will be used to ensure data quality and integrity. The QAPP shall detail how the following will be achieved in accordance with recognised publications which are equivalent to, or complement, accepted methods.

While the primary focus of the QAPP is to ensure collection of relevant, quality data for evaluating performance claims, it is the responsibility of the parties involved to ensure that all activities are undertaken in a manner consistent with workplace health and safety considerations.

4.1 Data Quality Objectives

The Data Quality Objectives (DQO) are to obtain accurate and relevant data to assess the claimant's Performance Claim. Data quality will be assessed against the criteria of representativeness, completeness, and comparability.

Where a device provides quantity control, on-site monitoring must consider both contaminant concentration and mass or load transported. Data collected must be representative of typical storms, for each event that forms part of the device evaluation.

Verification may include conditions or design criteria related to the tested configuration or arrangement, e.g. vault volume per cartridge, catchment area per device, flow rate per catchment area.

Representativeness is largely achieved through the collection of flow-weighted event mean concentration samples, except for those contaminants that are obtained by grab samples.

The events sampled must also represent rainfall, and thus runoff, patterns for the catchment across an extended period of time typically (> 12 months) and be subject to the qualifying number of characteristic storms being achieved. Representativeness shall be assessed and reported.

Completeness of data will require that enough storm events are sampled to allow accurate evaluation (e.g. minimum of 15 events).

The data collected must be comparable to performance at other locations. Comparability requires that the contaminants analysed (e.g. total suspended solids versus suspended solids concentration) and the sampling and analytical methods used can be compared.

4.2 Organisational Roles and Responsibilities

There are many different parties involved in measuring device performance. These generally include the regulatory body, claimant, general contractors (including installation and maintenance contractors), testing organisation, analytical laboratory, site owner, and the evaluation panel. All have roles and responsibilities in the successful completion of a project (USEPA, 2002).

Organisational roles and responsibilities shall be clearly identified in the QAPP. The claimant, sampling organisation (including both equipment and sampling), analytical laboratory, and reporting organisation shall be clearly identified, along with limits of their roles. Ideally, key personnel, their titles, and contact information will be included. An organisational chart should clarify personnel and their roles (especially in confirming independence requirements).

4.3 Description of Test Sites

Ideally, a test site shall be selected so that the results are indicative of performance in other locations. The claimant shall propose a suitable site and demonstrate its appropriateness for performance testing.

This site shall be representative of the installation and land use appropriate to the device and intended market segments. The test site land use shall be detailed and described according to land use category (e.g. commercial, industrial, high density residential).

There should be limits on maximum pollutant concentrations included within the set of qualifying storms presented for analysis, because very high pollutant concentrations are likely to lead to an overestimation of treatment performance.

Where sites are likely to have atypical land uses or water quality characteristics, the claimant may elect to take samples to characterise the trial site to avoid committing to testing at a site where pollutant concentrations are likely above the maximums allowed for qualifying events, or below limits of detection. Baseline monitoring is optional. As an indication, stormwater should contain the average typical concentrations of contaminants as provided in recognised industry publications and studies (eg. Australian Runoff Quality, Lucke et al. 2018). Typical concentrations for commonly regulated pollutants are provided in Table 1.

It is acknowledged, however, that pilot sites may be limited to new developments through the regulators' development assessment process. Therefore, testing should continue for a period sufficient to demonstrate a range of typical catchment pollutant concentrations, and when this data is collected it can be used to augment concentrations referenced in other literature or to develop a site-specific dataset (i.e. in the case that there is a paucity of published data for the particular application).

The recommended mean influent concentrations are given as guide and may lead to disqualification of a pollutant parameter for an individual storm if agreed upon between the technical expert reviewer and independent evaluation panel.

Storms with influent pollutant concentrations below the recommended minimum in Table 1 may be excluded at the discretion of the Claimant.

A full description of the test site shall be provided, and shall include the following:

a. Catchment area, land use, percentage impervious cover;

b. Aerial photos and site photos;

c. Geology, hydrogeology, soil types, surface hydraulics;

d. Potential pollutant sources;

e. Baseline stormwater quality (if located on an atypical land use);

f. Site map, showing catchment area, drainage system layout, treatment device, and sampling points, preferably GIS compatible;

g. Treatable flow rate (TFR);

h. Expected catchment flows;

i. Make, model and capacity of treatment device;

j. Closest receiving water body;

k. Identification of bypass flow rates and/or flow splitter design;

l. Pre-treatment system, if any;

m. Site suitability – e.g. safety, access for flow measurements, power, phone; and,

n. Any known adverse site conditions.

Table 1 — Typical Untreated Stormwater Contaminant Concentrations

	Adopted minimum	Recommended Mean Influent Concentration & (Standard Deviation)[1]	Adopted maximum average for all qualifying storms: (Mean + 1SD)[2]	Maximum for any individual event: Mean + 2SD
TSS	Limit of detection	151 (+220)	371	591
TP	Limit of detection	0.34 (+0.37)	0.71	1.1
TN	Limit of detection	1.82 (+1.27)	3.09	4.4

Notes:

1: Recommended mean influent concentrations and standard deviations from Goonetilleke, A, Thomas, E, Ginn, S, and Gilbert D (2005), Understanding the Role of Land Use in Urban Stormwater Quality Management.

2: Applies to all storms. Individual storms may exceed these values.

4.4 Measuring Rainfall

Rainfall shall be measured by a rain gauge (pluviometer) that is capable of sampling at intervals of 5 minutes or less, and in increments no greater than 0.25mm. An electronic rain gauge connected to a data-logger is recommended. A non-recording rain gauge installed at the test site will allow the recording gauges totals to be calibrated and increase confidence in data.

The location of the rain gauge in relation to the test site shall be shown on a map. The rain gauge shall be installed and maintained according to manufacturer's instructions, and as a minimum be checked, cleared of debris regularly and calibrated at least two times during the testing period (if a non-recording gauge is used this can be emptied and 'reset' to achieve this). It is also recommended that rain gauges are checked prior to each anticipated storm event targeted for sampling to ensure they have not clogged and rainfall data is recorded during the monitoring event. Rain gauges shall be protected from excessive wind velocities that could skew accuracy of measurement.

Guidance on installation of rain gauges can be found at the Bureau of Meteorology website:

http://www.bom.gov.au/climate/how/observations/rain-measure.shtml

4.5 Qualifying Storm Events

A sufficient number of qualifying storm events is required for a statistically robust data set to support assessment. In addition to achieving a sufficient number of events, the data set should ensure a range of flow conditions are demonstrated.

Pollutants with relatively consistent concentrations and removal rates may require a smaller number of events to be analysed to achieve statistical-confidence for performance. Where pollutant concentrations and removal rates are more variable additional samples may be required to account for this variability.

In all cases a minimum of 15 qualifying events is required, but an upper number of tests needs to be determined based on an assessment of the data using credible statistical methods (such as ANOVA/ t-test techniques) to achieve at least 90% statistical significance between paired samples of influent and effluent (Toifl et al. 2017).

If the level of statistical significance is not able to be demonstrated more events must be sampled until the 90% statistical significance is achieved. Where this statistical validation may require excessive events (e.g. >30) to prove (as determined by the recognised statistical power techniques such as the equation described by Burton and Pitt, 2001) an altered claim can be considered commensurate with evidence.

Australian rainfall patterns have changed significantly in recent time and are likely to continue to do so (Westra et al, 2013). Furthermore, Australian rainfall is more variable than the US and New Zealand where SQIDEP reference protocols were sourced.

Setting specific criteria regarding minimum or maximum storm durations and antecedent dry periods would result in the exclusion of many storm events. It would likely prolong the monitoring program with potentially limited improvement in the strength of the dataset. In fact, omission of events based on arbitrary hydrologic criteria may diminish the representativeness of the overall dataset. Furthermore, stormwater quality is highly variable between and within storm events.

First flush phenomena are variable, site specific and more strongly applicable to TSS (Lee et al. 2002, Liu, 2011, Modugno et al. 2015), than to TP or TN (Acharya et al 2010). Lee et al (2002) also found no correlation was observed between the first flush phenomenon and the antecedent dry weather period. The recommended sampling methodology is to collect composite samples comprising aliquots taken throughout the storm event. This results in an aggregated (averaged) concentration that smooths out the impacts of intra-storm pollutant variability. As such, the over-riding criteria should be whether the average influent concentration appropriately reflects the expected stormwater quality from the catchment land-use.

If the stormwater quality is within the expected range (between the limit of detection and the maxima from Table 1), a sample should not be excluded from the dataset simply based on antecedent dry weather period. Each monitoring programme will need to identify the period delineating the end of one event and beginning of the next. This is typically at least 24 hours or the time taken to reset monitoring equipment.

Similarly, given the variability of rainfall events in Australia, if there were a requirement to capture 100% of the hydrograph this would mean a large number of events would be excluded from the dataset and be unlikely to increase the rigour of the predictive capacity of the testing programme.

It is more important that the overall dataset contains a diversity of storms. Submitted reports will need to provide hydrographs of the sampled events demonstrating the programme has representatively captured the event.

At least two (2) peak inflows from the sampled events should exceed 75% of the design TFR of the device, and 1 at or greater than its design TFR.

Sampling events should be sufficiently distributed throughout the monitoring period to capture seasonal influences on storm conditions and device performance.

There is no stipulated minimum storm event duration, however for the majority of qualifying events (80%) at least 8 aliquots are required if discreet aliquots are being collected (continuously variable sampling is also acceptable).

The intent of each monitoring programme is to capture as much of the event as is practicable.

The independent evaluation panel must be satisfied that the qualifying storms being assessed includes a good range of storm events including longer and shorter duration storms of varying magnitude and that at least 50% of qualifying storms should include the first 70% storm hydrograph coverage. Where storm events are longer than 8 hours in duration, sampling over the first 8 hours is regarded as sufficient.

Where claimants wish to conduct the field trials on multiple sites and pool the results, the onus is on the claimant to demonstrate to the Independent Evaluation Panel that there are no aspects of the test sites or device configurations that would invalidate such an approach (In many cases this is likely to increase the rigour of results by expanding the range of conditions under which the device is tested; better reflecting real-world conditions).

The QAPP should allow results to be presented and interpreted along with qualifying storm characteristics as summarised in Table 7.

4.6 Flow Monitoring

Flow monitoring equipment must be able to continuously monitor flow at regular intervals to match rainfall information (5 minutes or less is recommended) throughout the duration of a storm event and over the expected range of flows. Depth measurement or area/velocity devices are most common – selection will depend upon the test site and method of conveying stormwater (WSDE, 2011).

Flow measurement at the inlet and outlet is recommended. Monitoring of bypass flows is optional, however at a minimum the monitoring information should be sufficient to identify periods during which the device is operating in bypass mode.

Proper presentation and interpretation of treated and bypassing flows is considered critical to ultimately supporting design outcomes. Where the monitored treatment measure is effectively a sealed "black box", flow monitoring from either the inlet or outlet may be sufficient. If the treatment measure relies on hydrological changes (e.g. biofilters, wetlands, infiltration) inlet and outlet flow measurement is required.

Flow monitoring equipment shall be described (make and model), and the description shall include flow splitter and bypass flow set points, and flow conditions (gravity or pressure) (WSDE, 2011). Equipment must be installed and calibrated according to manufacturer's instructions and as described in the QAPP.

As a minimum, the equipment's internal desiccators, sensors and connections shall be inspected regularly (USEPA, 2002). It is also recommended that in-pipe sensors and intakes be checked prior to each anticipated storm event targeted for sampling to ensure they are not blocked, damaged, or covered by sediments or gross pollutants.

When determining suitable flow monitoring configurations, the relevant Australian Standards should be consulted, and references provided.

4.6.1 Accounting for internal bypass flows

Some devices by nature of their internal geometry and treatment process may allow a portion of flows entering to be internally bypassed. In these instances, flow monitoring should be sufficient to determine the treatment effectiveness across the devices (i.e. inlet and outlet conditions).

4.7 Sample Location

The inlet sample shall be taken as close as possible to the device, at a minimum this should be at a point where total site runoff is sampled. The QAPP should identify whether effluent characterisation accounts for total storm flow, including bypass if it occurs.

For typical installations, gross pollutants (>1000 µm) should be excluded (WSDE, 2011) from any captured water samples, unless this is being claimed for the device.

Outlet flow should be sampled either prior to or after mixing with bypass flow and Claims identify the inclusion/exclusion of bypass flows (Figures 2 and 4).

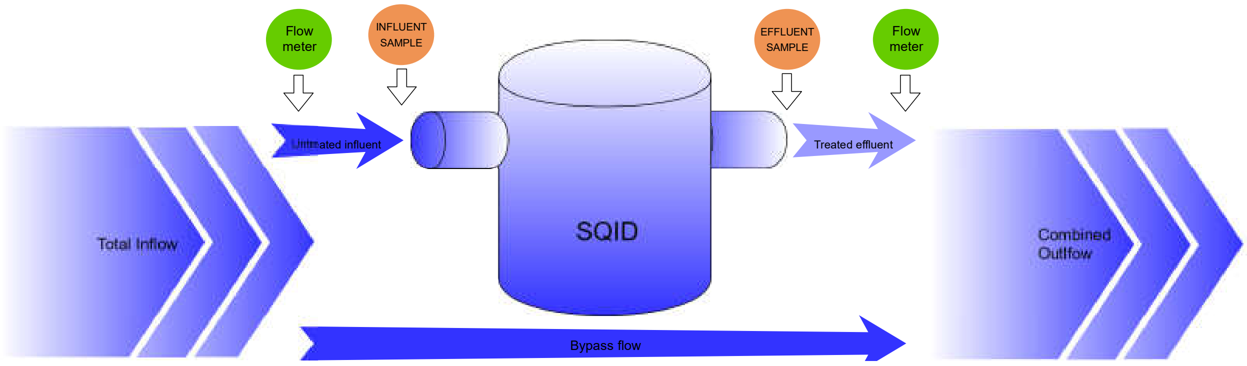


Figure 2 — Flow Sensor and Sample Intake Locations (bypass flows not accounted for in analysis)

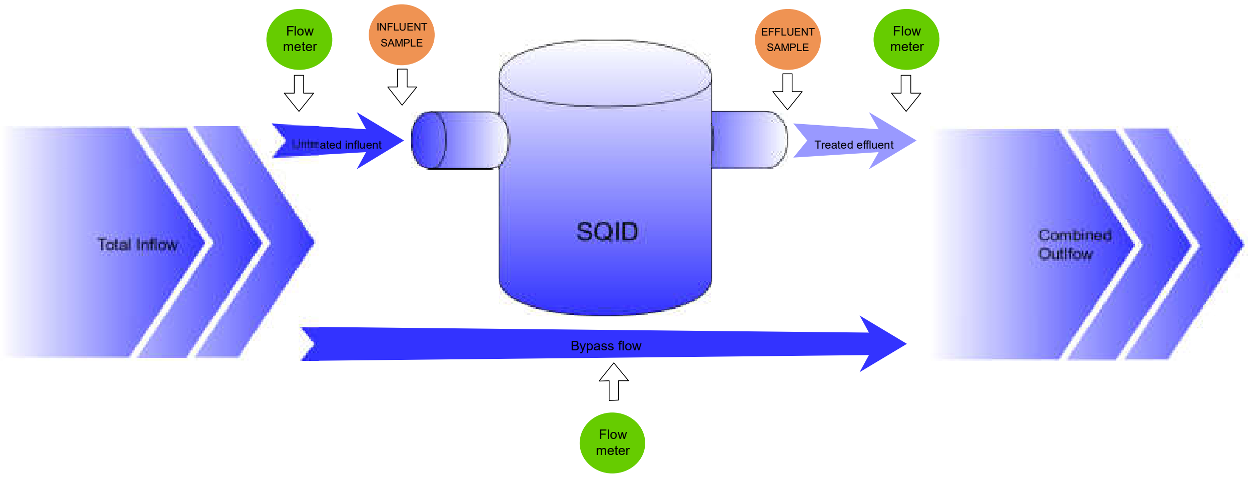
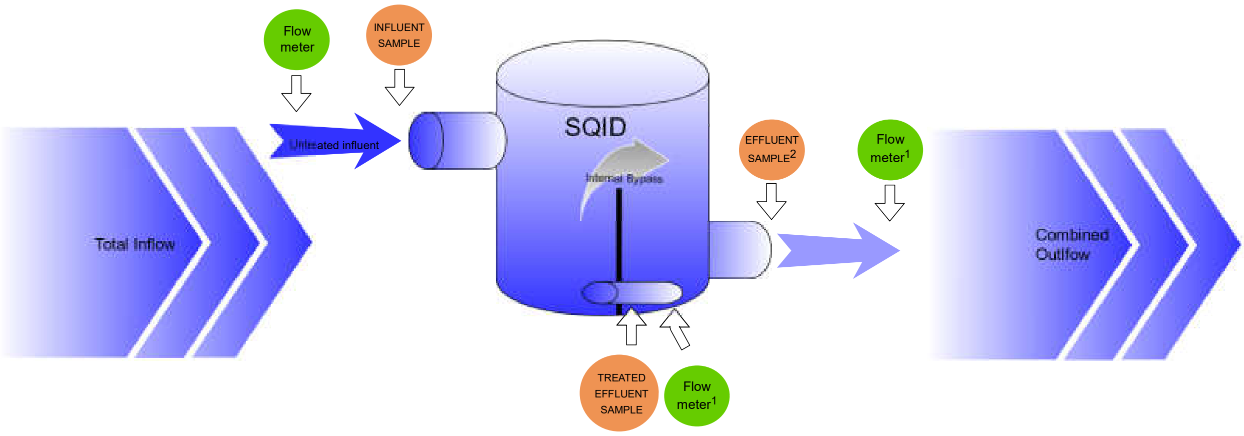


Figure 3 — Flow Sensor and Sample Intake Locations (bypass flows accounted for in analysis)



[1] Flow meters may be at either or both locations

[2] Effluent samples may be collected once recombined, if Claim is to include bypass flows.

Figure 4 — Sample location accounting for any internally bypassing flows

If a claim is being made for performance including bypass, the contribution of bypass (if/when it occurs) shall be incorporated into the calculation of device efficiency (USEPA 2002) or design tools as appropriate as described below.

If the outlet flow is sampled prior to mixing with bypass flow (Figure 3) it should be noted when the bypass condition occurs (but it is not necessary to measure bypass flows).

The performance claim must be made in relation to the device up to TFR, and no removal can be claimed for the bypass flows.

In this circumstance the performance claim claimed must be qualified as such; sizing and design advice must recognise this fact.

If the outlet and bypass flow is to be sampled together (Figure 4), samples should be collected after sufficient mixing has occurred and prior to comingling with any other runoff. In this event bypass flows must be measured, and the concentration of the bypass flows assumed to be the same as the device influent.

Figures 2 to 4 shows sampling and flow monitoring configurations for devices that allow various bypass configurations. These may identify changes in treatment veracity over the device's deign life (e.g. filters becoming clogged reducing hydraulic throughput).

If internal bypass occurs and samples are collected immediately after the treatment element (i.e. separate to internal bypass flows) additional flow monitoring will be required to allow the treatment effect for all flows entering the device to be calculated, where this is intended to be claimed.

There is potential for some stormwater constituents to stratify during conveyance. To avoid sampling stratified flow, all sampling points shall be located where mixing of the flow is maximised (USEPA, 2002; WSDE, 2011).

Sampling locations should be consistently located upstream and downstream of the tested device to allow representative consideration of stratification. It is recommended the location is agreed through the QAPP to accommodate operational realities of field testing.

Hydrocarbons, or other light, non-aqueous phase liquid which are likely to remain in a floating, free state at time of arrival at the testing site, shall be sampled in accordance with recognised guidelines. Where emulsified hydrocarbons are expected, justification should be provided, and samples should be collected from a zone of representative mixing with appropriate collection and preservation techniques used.

4.8 Accounting for scour

For devices installed online, scouring might occur during large events. Any scouring effects shall be assessed and reported. The assessment may be in the form of hydrodynamic modelling, or other approach with appropriate justification.

Otherwise if the device is an offline device; and/or there is sufficient evidence that scouring is not present, it shall be provided as part of the Performance Report.

In situations where there is the potential for scour to occur, preventative strategies can be recommended as part of installation or maintenance methodologies.

4.9 Monitoring Equipment

Evaluation of device performance requires measurement of stormwater inflow into the device, outflow, stormwater quality, and rainfall.

Equipment is required to measure rainfall, inflow and outflow volumes, and some method of determining the bypass volumes must be incorporated (measurement or calculation).

Equipment is also required to sample stormwater for laboratory analysis. For all equipment, the make and model of equipment, and procedures and schedule for calibration, inspection and cleaning shall be provided (USEPA 2002).

Consideration should be given to access for monitoring Equipment , equipment security and protection, and power (access to grid, or unobscured sky for solar) and phone supply/modem (if the site is to be remotely telemetered, either land line access or cellular reception).

The potential for power failure and subsequent loss of samples should also be considered (WSDE, 2011).

4.9.1 Automated Samplers

Automated samplers are to be used for all water sampling, except where grab samples are required (i.e. to ensure timely sample preparation, preservation or monitor unstable parameters).

The sampler shall be installed and calibrated according to the manufacturer's instructions and maintained between each sampling event. Information provided shall describe how the sampler will be programmed, how sampling will be triggered, and how the sampler purges and rinses between samples (USEPA, 2002; WSDE, 2011).

The bottle changing procedure shall also be described. The suction tube material, length and vertical lift should be described, and the location of the tube inlet relative to flow conditions should also be described. Teflon shall be used for sampling organic constituents (WSDE, 2011).

4.10 Sampling Methodology

As a minimum, flow-weighted composite samples should be collected utilising an automated sampler, whenever possible. However, some contaminants may require grab sampling under some sampling protocols.

4.10.1 Automated Sampling

Where the constituent being measured does not require grab sampling, automated sampling should be undertaken. Samples can be taken by automatic flow-weighted compositing, or discrete samples that can be composited later.

Where samples are manually composited, it is recommended this is undertaken at the analytical laboratory to minimise risks of contamination.

4.10.2 Grab Sampling

Grab sampling is required for constituents that transform rapidly, require special preservation, adhere to bottles, or where compositing can mask the presence of some contaminants through dilution.

Grab sampling is recommended for pH, temperature, total petroleum hydrocarbons (TPH), oil and grease, mercury (Hg), hexavalent chromium ($Cr6+$), bacteria, cyanides, total phenols, and residual chlorine.For all other constituents and pollutants,sampling should use automated samplers.

Grab stormwater samples are discrete samples (not composited), normally collected within the first 30 minutes after the onset of runoff, but no later than within the first 60 minutes. If grab sampling is required, the approach and justification shall be clearly documented. The QAPP shall describe how the criteria for a qualifying storm event will be met. The availability and preparedness of sampling staff shall also be demonstrated.

4.10.3 Flow-Proportional Sampling

This method of sampling collects an aliquot each time the auto-sampler is prompted, based on a flow volume interval. This interval is typically programmed into the datalogger, and should be described in the QAPP. As many aliquots as possible should contribute to the composite sample (Fassman, 2010; Ma et al, 2009) and should provide statistical confidence in representativeness.

This protocol requires that at least 80% of the submitted events have at least 8 aliquots collected from the event to form the composite sample. These aliquots should be collected from both the rising and falling limbs of the hydrograph. Refer Table 7.

4.10.4 Time-Proportional Sampling

This is a method of sampling that is best suited to auto sampling techniques.

Prediction of the type of storm event prior to its occurrence is necessary to enable effective time-proportional sampling. It is difficult to get the time intervals correct with Australia's multi-peaked events and therefore, this method should be used with caution.

A statistically representative number of discrete samples or sample aliquots shall contribute to each composite sample, with the emphasis on the hydrograph's rising limb (Fassman, 2010), at both the inlet and outlet of the device.

The sampler should be programmed to take the maximum number of aliquots possible (USEPA, 2002). All samples collected from qualifying events should be analysed and reported.

While time-proportional sampling is acceptable, it is challenging to implement well. The compositing methodology must be clearly documented and justified.

The flow data will need to be readily available following a storm event to properly composite time-proportional samples and ensure holding times are met.

4.11 Sampling Quality Assurance and Quality Control

Operation and maintenance schedules for sampling equipment (e.g. automated), flow monitoring and rainfall equipment shall be provided. Sample blanks for field and analytical testing will be supplied in accordance with the QAPP and recommendations in the EPA guidelines.

Chain of custody documents identifying sample, collection agency, collection time, preservation used, and laboratory receipt of sample and sample condition shall be provided.

4.12 Laboratory Analysis

All analysis shall be undertaken at laboratory or analytical facilities with current NATA accreditation for requested analysis (including limit of reporting).

The method chosen for analysis shall be detailed in the QAPP, including any justifications as considered necessary (e.g. depending on expected catchment conditions, analysis methods maybe chosen based on limit of reporting). Analyses should be in accordance with National or International standards (eg. APHA, 2017).

4.12.1 Laboratory Quality Assurance and Quality Control

Proper quality assurance and control procedures are critical within any laboratory engaged to undertake samples. Generally, the use of NATA accredited facilities will ensure that a high standard of quality management is adhered to.

Beyond the accreditation for specific tests, the laboratory should also be able to provide a suitable chain of custody documentation to identify sample receipt and condition, the samples should be properly labelled and stored pending testing, and holding times for samples should be observed.

In addition to field-based quality assurance samples, the laboratory should have its own procedures to demonstrate confidence in sample preparation methods and analysis, including the use of duplicates, spikes, surrogates and blanks. When examined together, quality assured laboratory and field data give the highest confidence in the measured results.

4.12.2 Laboratory Data Management

All documentation pertinent to undertaking field testing, sample collection and analysis, and reporting of results should be retained in full and presented in a logical and easy to follow format for evaluation.

It is desirable to receive testing results in electronic format to facilitate analysis and assessment, but copies of the accompanying certificates of analysis which include test results and laboratory quality assurance results should also be retained. Also include Chain of Custody documentation and any relevant field notes identify sample collection time, location and prevailing conditions.

Where electronic copies are provided these should be delivered to the independent party who is working with the claimant to deliver field testing. Full copies of original results (electronic and hardcopy) should be retained and submitted to the independent evaluation panel.

4.13 Reporting

Reporting must be prepared according to the approved QAPP by an organisation independent of the claimant.

The reporting organisation must understand the hydraulics and treatment mechanisms of the SQID, with knowledge and experience of proper sampling and flow measurement practices, and have the ability to properly interpret and report, without prejudice, the flow and water quality data.

A Statutory Declaration disclosing the nature of any commercial relationship between the claimant and the report author (or its affiliates) and must be supplied.

This work will typically be undertaken on a fee for service basis, and the payment of professional fees does not necessarily constitute a breach of independence.

5 Performance Reporting

The performance of a device needs to be reported consistently for efficient and accurate evaluation.

This section discusses the requirements of the SQIDEP in terms of the framework of the Performance Report and the performance metrics.

5.1 Non-Detects

Non-detects are values reported to be at or below a reporting limit and/or detection limit.

When analysing data for outlet concentrations, they need to be considered, as removing them may result in biased and non-representative estimated summary statistics of the monitored site (Drapper et al. 2018, Helsel & Cohn, 1988, Helsel, 2009).

Specifically, some treatment measures may consistently produce outlet concentrations below detection limits, therefore exclusion of these events will prevent a successful conclusion of the monitoring programme.

However, inclusion of non-detect data on influent concentrations holds little scientific value. Therefore, events reporting non-detects on the inlet may be excluded from data analysis.

Effluent sample results below the limit of detection (LOD) shall be set at 0.5 x LOD and must be accompanied by a sensitivity analysis showing impact on performance metrics of adopting both LOD and 0).

If there are a large number of non-detects, the applicant can propose the use of an alternate statistical method to analyse them, such as (1) regression on order statistics (ROS) method; (2) maximum likelihood estimation (MLE) method (Kayhanian, 2011); or another scientifically-justified method.

5.2 Framework for Reporting

A Detailed Performance report (DPR) is required after the local pilot trial (LPT) is completed.

Devices evaluated using the body of evidence (BoE) route are also required to summarise existing data and report using the following framework.

The requirements for reporting are as follows:

a. Device information (extracted and summarised from QAPP);

b. Sizing methodology and its description, including any non-validated or non-referenced assumptions;

c. Catchment characteristics (Area, approximate grade, landuse type) – photographs shall be provided;

d. Roles and Responsibilities of all parties involved;

e. Sampling and analytical methodologies (extracted and summarised from QAPP);

f. Discussion of all/any maintenance activities performed on the treatment measure including, nature, interval, modifications, repairs, replacements, observations;

g. Data reporting (for all qualifying events);

h. Discussion of any factors affecting the performance, including scaling effects and particle size distribution of both the influent and effluent. Other factors shall be included if deemed appropriate;

i. Box and Whiskers Plot for the Influent and Effluent Concentrations

j. Statistical Significance Testing;

k. Data quality (below);

l. Performance metrics (below), results and discussion; and

m. Statutory Declarations from associated organisations

5.3 Data Quality

The data collected shall be assessed and reported for the following factors:

a. Representativeness, completeness and applicability of rainfall/ runoff; and

b. Values relative to the detection limits of the analytical methods applied.

5.4 Performance Metrics

The pollutant removal capacity of a device needs to be consistent, and provided that suitable information is collected at the time of field trials, multiple metrics can be determined and should point to a consistent interpretation for the highest levels of confidence in evaluating results.

The SQIDEP allows a number of performance metrics to be presented as follows:

a. Five (5) types of percent removal efficiencies;

b. Event Mean Concentration (EMC) and (if applicable) Mass Discharge Variability; and

c. Statistical significance of differences (if any) between inlet and outlet EMCs.

The details of each performance metric are outlined below.

Equation 1

$$EMC = \frac{\sum_{i=1}^{n} V_i C_i}{\sum_{i=1}^{n} V_i}$$

Where:

Vi Volume of flow during period i

Ci Concentration associated with period i

n Total number of aliquots collected during event

5.4.1 Performance reliability and the statistical analysis of data

A qualifying number of sampling events is required to verify the statistical representativeness of the removal efficiency.

All performance metrics are supported by analysis of data collected when following this protocol, and all should provide a supporting case for the final accepted removal efficiency.

Performance reliability can be measured statistically by several methods.

It is assumed that the pollutant concentrations are likely to be log-normally distributed, however this assumption should be verified through statistical techniques and appropriate techniques employed to prepare the dataset for analysis.

Statistical parameters for evaluating the performance of the device seek to understand the difference between paired influent and treated effluent samples (i.e. treatment effect).

a. Ensure that the 90% Confidence Interval of the arithmetic average is provided (CRE and/or MRE calculated as recommended).

   A Confidence Interval of greater than 90% is required for a claim to be considered valid.

   Devices which can demonstrate reduced levels of variance in the treated effluent are likely to perform more predictably.

   Standard statistical techniques can be used to estimate the variability in a dataset.

   One such procedure is provided below.

b. Measure the spread of the effluent data by analysing the distance of the lowest and upper most point from the 1st and 3rd quartile values (effluent EMCs) against the inter-quartile range (IQR). Within 1.5 times IQR is desired.

c. Calculate the arithmetic mean above and below the standard deviation (CRE and/ or MRE). Within one standard deviation is desirable.

d. Calculate the difference between the arithmetic average and the median (CRE and/or MRE). Within 10% is desired.

5.4.2 Performance Metrics

The methods considered under the SQIDEP to compute and analyse removal rates and efficiencies are:

- Average and Median Concentration Removal Efficiency (CRE);
- Mass Reduction Efficiency (MRE);
- Relative Achievable Efficiency (RAE);
- Summation of Loads (SoL);
- Efficiency Ratio (ER); and
- Flow Based Variability (FBV) Curve;
- Event Mean Concentration and Mass Discharge Variability

Analysis should clearly indicate how treatment and bypass flows (either external or internal to the device) have been accounted for in the presentation of results. Reports may choose to report some, or all, of the above metrics, however, as a minimum AvCRE and ER shall be provided.

5.4.3 Average and Median Concentration Removal Efficiency

Pollutant Concentration Removal Efficiency (CRE) is computed to determine the reduction in pollutant concentration through a device.

Calculations depend on the sampling equipment configuration, as per Figure 2 - Figure 4.

The formula for computing CRE is as follows:

Equation 2

$$CRE(\%) = \frac{EMC_{in} - EMC_{out}}{EMC_{in}} \times 100$$

Where:

EMC_{in} is the event mean concentration measured in the inflow for each event; and EMC_{out} is the event mean concentration measured in the corresponding total outflow for each event.

For Sampling Configuration shown in Figure 2, EMC_{out} is the event mean concentration measured in the treated effluent.

Note, under this interpretation, if bypass is not measured, no credit can be reliably claimed for bypass and design guidance should allow bypass to be excluded from treatment.

If bypass occurs and is measured:

For Sampling Configuration shown in Figure 3, EMC_{out} should be calculated using Equation 2.

For the Sampling Configuration shown in Figure 4, EMC_{out} will automatically include any bypass flow if/when it occurs and Equation 1 can be used.

An alternative, when bypass occurs and it is measured, EMCout for the event is calculated as:

Equation 3

$$EMC_{out} = EMC_{treated\ out}\left(\frac{V_{treated\ out}}{V_{total\ outflow}}\right) + EMC_{in}\left(\frac{V_{bypass}}{V_{total\ outflow}}\right)$$

Where:

$EMC_{treatedout}$ is the event mean concentration measured in the treated effluent for the event

$V_{treatedout}$ is the measured flow volume treated by the device (not bypassing)

V_{bypass} is the flow volume of the event that bypasses the treatment device (measured or calculated)

$V_{totaloutflow}$ is the flow volume downstream of the junction of the bypass and treated effluent, as per Figure 14-2a, b (measured and/or calculated).

To calculate and report the Average CRE:

1. Calculate CRE for each event, according to Equation 1.
2. Calculate the arithmetic average of the CRE over all events.
3. Calculate the 90% confidence interval for the arithmetic average of CRE.

To calculate and report the Median CRE:

1. Calculate the median CRE over all events.
2. Compute the difference between the arithmetic average CRE and the median CRE.
3. Calculate the arithmetic mean above and below as the standard deviation for CRE.

Variation <10% between the median and average CRE indicate that the overall statistic is not influenced by an extreme event/s.

If median and average values are greater than 10% different, the data set should be inspected for the presence of an extreme value(s) which may need further investigation or explanation.

5.4.4 Average and Median Mass Removal Efficiency

Pollutant Mass Reduction Efficiency (MRE) is reported to determine the total mass of pollutant captured by the device. MRE calculations are relevant for devices which may provide runoff quantity management. The formula for computing MRE is as follows:

Equation 4 MRE

$$MRE(\%) = \frac{(V_{in} \times EMC_{in}) - (V_{out} \times EMC_{out})}{(V_{in} \times EMC_{in})} \times 100$$

Where:

V_{in} is the flow volume of each event, measured at the inlet;

V_{out} is the total outflow volume of each event, measured downstream of the junction of the bypass and treated effluent, as per Figure 14.2

EMC_{in} is the event mean concentration measured in the inflow for each event; and EMC_{out} is the event mean concentration measured in the total outflow for each event, as described by Equation 1 or Equation 2.

To calculate and report the Average MRE:

1. Calculate MRE for each event, according to Equation 3.
2. Calculate the arithmetic average of the MRE over all events.
3. Calculate the 90% confidence interval for the arithmetic average of MRE.

To calculate and report the Median MRE:

1. Calculate the median MRE over all events.
2. Compute the difference between the arithmetic average MRE and the median MRE.
3. Calculate the arithmetic mean above and below as the standard deviation for MRE.

Variation <10% between the median and average MRE indicate that the overall statistic is not influenced by an extreme event/s.

Close agreement of median and average MRE indicate that the overall statistic is not influenced by an extreme event/s. If median and average values are greater than 10% different, the data set should be inspected for the presence of an extreme value(s) which may need further investigation or explanation.

5.4.5 Average and Median Relative Achievable Efficiency

The relative achievable efficiency (RAE) is computed to mitigate the influence of influent EMC on the percent removal calculations. The RAE is a function of benchmark or 'irreducible' concentration.

This was derived from the 'best' median effluent concentration across all stormwater treatment devices reported in the International Stormwater BMP Database (Geosyntec and Wright Water Engineers, 2008). The RAE is calculated as follows:

Equation 5 RAE

$$RAE(\%) = \frac{EMC_{in} - EMC_{out}}{(EMC_{in} - C^{*})} \times 100$$

Where:

C* is an irreducible concentration used as a benchmark and taken from Table 2 (Fassman, 2010);

V_{out} is the flow volume of each event, measured at the outlet

EMC_{in} is the event mean concentration measured in the inflow for each event; and EMC_{out} is the event mean concentration measured in the total outflow for each event, as described by Equation 1 or Equation 2.

To calculate and report the Average & Median RAE:

1. Calculate RAE for each event, according to Equation 5.
2. Calculate the arithmetic average of the RAE over all events.
3. Calculate the 90% confidence interval for the arithmetic average of RAE.
4. Calculate the median RAE over all events.
5. Compute the difference between the arithmetic average RAE and the median RAE.

Variation <10% between the median and average RAE indicate that the overall statistic is not influenced by an extreme event/s.

5.4.6 Summation of Loads

The summation of loads method allows performance to be measured by calculating the ratio of all outlet loads to inlet loads.

Equation 6

$$SOL = 1 - \frac{Sum\ of\ inlet\ loads}{Sum\ of\ outlet\ loads}$$

Equation 7

$$SOL = 1 - \sum_{i=1}^{n} \frac{C_{inlet}V_{inlet}}{C_{outlet}V_{outlet}}$$

Where:

i duration of sample period

n number of aliquots

C_{inlet} C_{outlet} inlet and outlet concentrations respectively

V_{inlet} V_{outlet} volumetric flow rate of inlet and outlet respectively

To report SOL accurately, it is necessary to have a high (>50%) coverage of all events in the monitoring period to correlate concentration and flow volume. Reports will need to present this information.

Table 2 — Recommended C* Values Based on all Parameters in the 2008 International BMP Database Summary (Geosyntec and Wright Water Engineers, 2008, eWater, 2010)

Parameter	C*	Based on Treatment Device(s)	Number ofStudies
Total Suspended Solids	6 mg/L	Media filter	33
Total Dissolved Solids	N/A1		
Total Phosphorus	0.06 mg/L as P	Media filter	28
Dissolved Phosphorus	0.05 mg/L as P	Retention pond Constructed wetland	12 4
Total Nitrogen	1 mg/L as N	Biofilter	12
Total Kjeldahl Nitrogen	0.77 mg/L as N	Biofilter	22
Total Nitrate	0.20 mg/L as N	Wetland basin	5
	0.25 mg/L as N	Retention pond	12
		Wetland channel	3
Total Nitrate + Nitrite	0.05 mg/L as N	Retention pond	22
Total Lead	1.20 µg/L as Pb	Wetland basin	5
	2.20 µg/L as Pb	Biofilter	50
Dissolved Lead	1.00 µg/L as Pb	Media Filter	17
		Biofilter	38
		Wetland basin	2
Total Zinc	19.00 µg/L as Zn	Retention pond	34
Dissolved Zinc	10.00 µg/L as Zn	Retention pond	9
	19.20 µg/L as Zn	Biofilter	41
Total Copper	3.0 µg/L as Cu	Wetland basin	4
	5.0 µg/L as Cu	Media filter	27
		Retention pond	27
Dissolved Copper	4.37 µg/L as Cu	Retention pond	9
	5.90 µg/L as Cu	Biofilter	41

5.4.7 Efficiency Ratio

The efficiency ratio (ER) is defined in terms of the difference between the average Event Mean Concentration of influent and effluent pollutants calculated over all of the analysed events.

Equation 8

$$ER = \frac{average\ inlet\ EMC - average\ outlet\ EMC}{average\ inlet\ EMC}$$

5.4.8 Flow Based Variability (FBV) Curve

Some testing protocols call for each device to establish performance efficiencies for a range of storms (Geosyntec and Wright Water Engineers, 2009, Auckland Regional Council, 2010).

Over a sufficiently long enough period of time the ability for a device to remove pollution will lead to average effects which could be reasonably expected across different installations.

However, design processes will benefit from the ability for performance to be described at different flow rates, particularly where continuous simulation design techniques are employed, or where a device is being recommended for a sensitive application where a greater resolution of performance is required.

Where data is available and lends itself to presentation in a FBV format, this is desirable.

If direct flow versus performance data is not available there may be a case to consider how controlled tests (e.g. laboratory) may be used to augment valid field data to generate curves, which can then be included in design guidance.

In developing a FBV curve, a line of best fit which describes the performance claim should be produced for the entire curve or for any discrete part. This line of best fit must have a corresponding correlation co-efficient of greater than 0.5.

The FBV shall be presented as a graph with the data points, and the R^2 displayed. The proposed FBV data points (ie. those that would coincide with the FBV curve) shall be presented in a separate table.

The method for selecting the appropriate FBV shall be described in the submitted report. The standard options for line-of-best-fit in Excel could be considered, as a guide.

Other forms of the curve could be used that adjust for device scalability such as the volumetric loading rate (Lps/m^3).

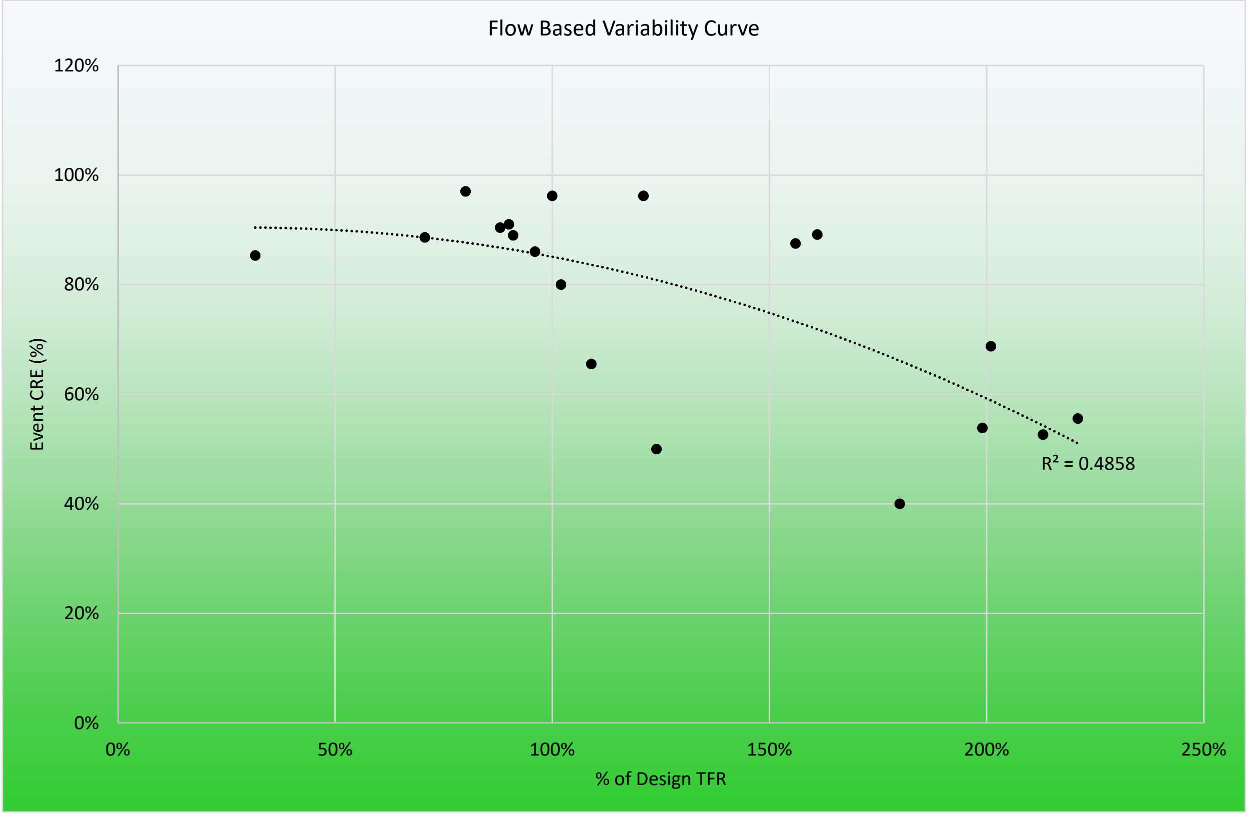


Figure 5 — Example of FBV curve

---

[1] The recommendations are based on a rationale by Dr. E Fassman; the Geosyntec and Wright Water Engineers (2008) data summary does not provide recommendations for C*.

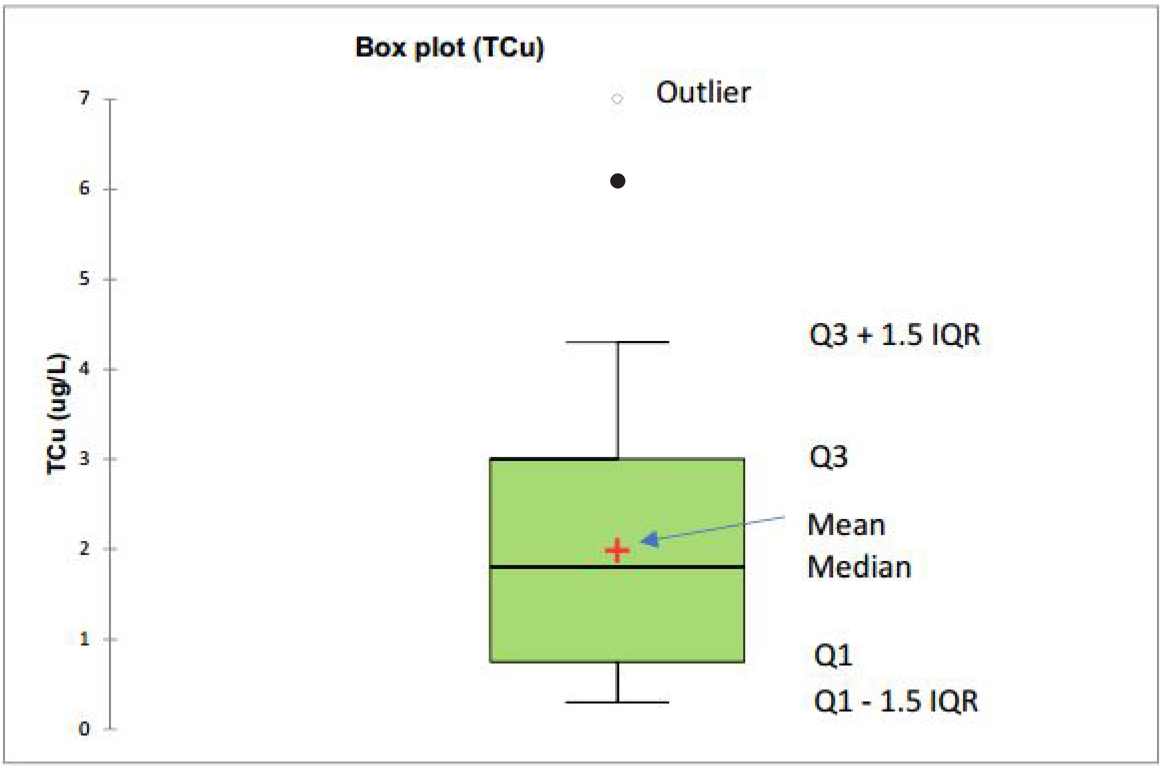


Figure 6 — Example Box and Whisker Plot with explanation of Terms, adapted from Geosyntec and Wright Water Engineers (2009)[1]

5.4.9 Event Mean Concentration and Mass Discharge Variability

The event mean concentration and Mass Discharge variability are required to verify the ability of the device to manage large variability in EMCs and mass discharges.

Box and whisker plots should be prepared for influent and effluent EMCs as well as mass loads (where presented).

The number of EMCs and mass loads contributing to each distribution should be clearly indicated.

The following explanation of a box and whisker plot is an excerpt from Geosyntec and Wright Water Engineers (2009):

> *"Box plots (or box and whisker plots) provide a schematic representation of the central tendency and spread of the data. A standard box plot consists of two boxes and two lines.*
>
> *The lower box expresses the range of data from the 25th percentile (1st quartile or Q1) to the median of the data (50th percentile, 2nd quartile, Q2). An upper box represents the spread of the data from the median to the 75th percentile (3rd quartile or Q3). The total height of the two boxes is known as the interquartile range (Q3 – Q1). A "step" is 1.5 times the interquartile range.*
>
> *Two lines are drawn from the lower and upper bounds of the boxes to the minimum and maximum data points (respectively) within one step of the limits of the box. Asterisks or other point symbols are sometimes used to represent outlying data points.*
>
> *Some statistical packages, including stand-alone software and third-party spreadsheet extensions, also include the confidence interval about the median as notches in the boxes about the center line or can be customized to include specific data percentiles (e.g., 5th, 10th, 90th, and 95th)."*

The above explanation is illustrated in Figure 6

* While the Y-axis label indicates total copper (TCu) expressed as µg/L, the procedure is equally applicable for TSS and many other water quality parameters.

5.5 Statistical Significance Testing

Statistical significance testing of differences between inflow and outflow EMCs and Mass Loads is required. This significance testing determines whether the difference is too large to have occurred by chance or too small such that it is insignificant.

The selection of the appropriate statistical significance test depends on the distribution and size of the data sets. For most water quality results, the distribution is usually log-normal, except for some constituents such as pH (Pitt & Maestre, 2005). Hence, statistical testing should be performed on log-transformed data, where appropriate.

The statistical significance testing on influent and effluent data sets should be tested with the following tests, as applicable:

1. Normality Test (Shapiro-Wilk, Anderson-Darling, Lillefors, or Jarque-Bera)
2. Sign Test
3. (Non-parametric)Wilcoxon-Mann-Whitney Rank-Sign Test (non-parametric)
4. Paired Student's t-Test (parametric)

Non-parametric tests (tests 2 & 3) are only needed if the data is not normally distributed (even after log transformation). If the data is normally distributed, only the paired t-test (test 4) applies.

The paired Student's t-Test assumes normal (or log-normal) datasets. Therefore, if normality cannot be confirmed, the conclusion of the t-Test is less reliable. The tests should be performed to validate the statistical difference between the influent and effluent data sets. The goal is to satisfy a 90% confidence interval.

5.6 Reporting Scour

The effects of scour shall be reported if the device is an online device. Alternatively, the claimant can provide evidence that the magnitude of scour is negligible in the device.

The effects of scour shall be hydro-dynamically modelled; otherwise an alternative option of demonstrating scour effects should be developed as part of the QAPP and reported accordingly.

Table 3 — Minimum data and qualifying event requirements for assessment

Sampling Events	Field Testing Criteria
Type of Event	Rainfall Events[2]
Minimum Number of Events	The greater of: a. 15 events, and b. Sufficient events to achieve 90% confidence interval, as determined by defensible statistical method (e.g. ANOVA, t-test) that examines influent and effluent pairs. This may vary between target pollutants (based on catchment variability). In this event, statistical analysis can be undertaken separately for each species of interest.
Minimum Rainfall Depth	Sufficient to collect minimum sample volume (based on laboratory analytical requirements).
Recommended Inter-event Time	Min 6 hours[3]
Device Size	Full Scale (where a 'family' of devices are being included as part of the claim sizing relationships must be provided for evaluation along with any basis of justification).
Runoff Characteristics	Target pollutant profile of influent and effluent
Runoff Volume or Peak Flow	At least 2 events should exceed 75% of the design water quality volume/ TFR and 1 event greater than 100% of the TFR.
Sampling Procedures and Techniques	
Automated Sampling	Composite samples on a flow- (preferred) or time-weighted basis
Minimum Number of Aliquots	80% of field test collections should have at least 8 per event[4]. Notwithstanding aliquots should be collected to provide hydrograph coverage of rising and falling limbs.

2: Must not Include Controlled Field Tests. See glossary for the definition of controlled field tests.

3: Interevent time or antecedent dry period (ADP) will be dependent on sampling practicalities and catchment pollutant generation. Shorter ADP events may be considered where influent concentrations are above detection limits. Including minimum qualifying concentrations and aliquot collection will impose a limitation on events that can be included in analysis, but if samples are collected, their analysis and/ or omission should be disclosed for completeness of data presentation.

4: Aliquot collection is determined by sample device collection rate and storm duration. For more intense, shorter duration storms a reduced number of aliquots may result. The protocol adopts a practical approach to ensure shorter duration events are able to be included in analysis and evaluation to achieve statistically robust outcomes.

Sampling Events	Field Testing Criteria
Hydrograph coverage	At least 50% of qualifying storms should include the first 70% storm hydrograph coverage (or, for storms longer than 8 hours, capture of the first 8 hours). Programmes should aim to capture full hydrographs for all events, but flexibility will be considered for large volume, long duration events. Dependent on catchment and rainfall patterns, multiple peaks should be accounted for (at least 1 occurrence).
Grab Sampling	Only for constituents that transform rapidly, require special preservation or adhere to bottles, or where compositing can mask the presence of some contaminants through dilution.
Sampling Location	As identified and agreed in the submitted QAPP.
Sampling Procedures and Techniques	
Chemical and Physical analytes	As identified and agreed in the submitted QAPP.
Minimum and maximum (influent) pollutant concentrations for qualifying events	Minimum concentrations: exclude if below limit of detection. Maximum: mean+2SD for any single event, and mean +1SD in the aggregate dataset. Refer **Table 1.**
Analytical Methods	NATA accredited sample handing and analytical methods. Refrigerated autosamplers may be required to adequately preserve samples.
Requirements	
Flow Measurement Location	Inlet, Outlet and Bypass, as applicable. Based on relevant accepted measurement protocols for flow type (e.g. open channel, in pipe)
Precipitation Measurement	Automatic rain gauge (pluviometer)
Recording Intervals	5 minutes or less
Rainfall Recording Increments	No greater than 0.25mm
Rain Gauge Calibration	Twice during monitoring period

Sampling Events	Field Testing Criteria
Data Analysis and Reporting	
Performance Indicators	Based on the Performance Claim stated in Detailed Performance Report. (Can include but not limited to TSS, Metals, TPH, TP & TN). The target pollutants and testing rationale must be described in the QAPP & Detailed Performance Report. Where a device is claiming total reductions of a particular pollutant, it is not necessary to include speciation. If speciation is not undertaken then reductions of sub-species cannot be claimed.
Data Analysis and Reporting	
Performance Indicators Calculation	Concentration Removal Efficiency (CRE) (See Section 6.4.3) (Arithmetic average and median. If difference is 10% or greater, inspect data set closely) Mass Removal Efficiency (MRE) (See Section 6.4.4) (Arithmetic average and median. If difference is 10% or greater, inspect data set closely) Relative Achievable Efficiency (RAE) (See Section 6.4.5) (Arithmetic average and median. If difference is 10% or greater, inspect data set closely Summation of loads (SoL) (See Section 6.4.6) (Arithmetic Average and median. If difference is greater than 10% inspect dataset closely) Efficiency Ratio (ER) (See Section 6.4.7) (Arithmetic Average and median. If difference is greater than 10% inspect dataset closely) Flow Based Variability (FBV) (See Section 6.4.8), including a plot of one of the above performance measures against the 25, 50, 75, 100 and 125 percent of the treatable flow rate. Provide details on the selected curve and the associated R^2 value.
Performance Variability Schematics	Box and Whisker Plots of inlet and outlet EMCs.
Statistical Significance Testing	Log-transformed inlet and outlet paired samples at 90% confidence level.
Sizing Methodology	A sizing methodology must be provided that allows an evaluation of performance of other devices in a 'family' to be reviewed. This should include relationships established under defensible theoretical/ modelled conditions or testing undertaken under either field or laboratory conditions.

6 References

1. Acharya, A., Piechota, T.C. and Acharya, K. 2010, Characterization of First Flush Phenomenon in an Urban Stormwater Runoff: a Case study of Flamingo Tropicana Watershed in Las Vegas Valley, Conference Proceedings – World Environmental and Water resources Congress 2010: Challenges of Change, ASCE.
2. APHA. 2017 Standard Methods for the Examination of Water and Wastewater, 23rd ed.; American Public Health Association, American Water Works Association & Water Environment Federation: Washington, DC, USA.
3. Auckland Regional Council, 2012, Proprietary Devices Evaluation Protocol (PDEP) for Stormwater Quality Treatment Devices; Version 3; Auckland Regional Council: Auckland, New Zealand.
4. Burton, A. & Pitt, R. 2015. Stormwater Effects Handbook: A Toolbox for Watershed Managers, Scientists and Engineers. Available online: http://unix.eng.ua.edu/~rpitt/Publications/BooksandReports/Stormwater%20Effects%20Handbook%20by%20%20Burton%20and%20Pitt%20book/hirezhandbook.pdf
5. Drapper, D., Hughes, L. and Rahman, J. 2018. Much ado about next to nothing; a stormwater quality perspective on handling results below detection limits, Conference Proceedings, Stormwater 2018 conference, Sydney Olympic Park, Sydney, October 9-12, 2018.
6. Fassman, E.A., 2010. Sampling Requirements and Reporting Statistics for the Proprietary Devices Evaluation Protocol Development. Prepared by UniServices for Auckland Regional Council. Auckland Regional Council Technical Report Number TR001/2010.
7. Geosyntec Consultants and Wright Water Engineers, 2008. Analysis of Treatment System Performance: International Stormwater BMP Database (1999-2008). Water Environment Research Foundation, American Society of Civil Engineers, U.S. Environmental Protection Agency Federal Highway Administration, and American Public Works Society. Available Online:http://www.bmpdatabase.org/MonitoringEval.htm#PerformanceEval
8. Geosyntec Consultants and Wright Water Engineers, 2009. Urban Stormwater BMP Performance Monitoring. Prepared under support from the U.S. Environmental Protection Agency, Water Environment Research Foundation, Federal Highway Administration and Environmental and Water Resources Institute of the American Society of Civil Engineers. October. Available Online: http://www.bmpdatabase.org/Docs/2009%20Stormwater%20BMP%20Monitoring%20Manual.pdf
9. Goonetilleke, A., Thomas, E., Ginn, S. and Gilbert, D. 2005, Understanding the role of land use in urban stormwater quality management, Journal of Env. Management, 74(1), 31-42, Elsevier BV.
10. Helsel, D.R. and Cohn, T.A. 1988. Estimation of descriptive statistics for multiply censored water quality data. Water Resource Res., Vol. 24, 1997-2004.
11. Helsel, D. 2009, Much Ado about Next to Nothing: Incorporating Nondetects in Science. Annals of Occupational and Environmental Hygiene, 257-262
12. Kayhanian, M., 2011, Influence of Non-Detects on TMDL Mass Loading Calculations. Stormwater Updates-March, Volume 6, No. 3
13. Kayhanian, M., Singh, A., & Meyer, S., 2002. Impact of Non-Detects in Water Quality Data on Estimation of Constituent Mass Loading. Water Science and Technology, 219-225.
14. Lee, H., Lau, S.L., Kayhanian, M. and Stenstrom, M.K. 2004, Seasonal first flush phenomenon of urban stormwater discharges, Water Res., 38(19), 4153-4163, Elsevier BV.
15. Liu, A. 2011. Influence of Rainfall and Catchment Characteristics on Urban Stormwater Quality. Ph.D. Thesis, Queensland University of Technology, Brisbane, Australia, 30 October 2011.

16. Lucke, T, Drapper, D. and Hornbuckle, A., 2018. Urban stormwater characterization and nitrogen composition from lot-scale catchments – new management implications, Science of the Total Environment, 619-620, 65-71, https://doi.org/10.1016/j.scitotenv.2017.11.105, Elsevier B.V.
17. Ma, J.-S., Kang, J.-H., Kayahanian, M., and Stenstrom, M., 2009. Sampling Issues in Urban Runoff Monitoring Programs: Composite versus Grab. Journal of Environmental Engineering 3(135): 118-127, ASCE.
18. Modugno, M., Andrea, G., Gorgoglione, A., Iacobellis, V., la Forgia, G., Piccinni, A.F., and Ranieri, E. 2015, Build-Up/Wash-Off Monitoring and Assessment for Sustainable Management of First Flush in an Urban Area, Sustainability, 7(5), 5050-5070, mdpi.
19. Pitt, R. and Maestre, A., 2005, Stormwater quality as described in the National Stormwater Quality Database (NSDQ), Proceedings of the 10th International Conference on Urban Drainage, Copenhagen, Denmark, 21-26 August. IWA.
20. Toifl, M., O'Halloran, R., Palombo, E. and Imteaz, M. 2017, Literature Review for Stormwater Quality Improvement Device Validation Protocol (SQIDEP), Final Report – WaterRE Project 3038-16, Water Research Australia Ltd, Swinburne University.
21. USEPA/United States Environmental Protection Agency Edison NJ. 2002, ETV Verification Protocol Stormwater Source Area Treatment Technologies, Draft 4.1. March 2002. Available Online: https://archive.epa.gov/nrmrl/archive-etv/web/pdf/04_vp_stormwater.pdf
22. Westra, S., Evans, J., Mehrotra, R., Sharma, A. (2013), A conditional disaggregation algorithm for generating fine time-scale rainfall data in a warmer climate, Journal of Hydrology, Volume 479, pp.86-99
23. WSDE/Washington State Department of Ecology, 2011. Technical Guidance Manual for Evaluating Emerging Stormwater Treatment Technologies: Technology Assessment Protocol – Ecology (TAPE). Publication Number 11-10-061 (Revision of Publication 02-100-037). August 2011.

5 Change Log

Version 1.2

- General text and formatting improvements
- Removal of minimum pollutant concentrations in qualifying storm events. Storms with low concentrations of pollutant provide useful data, reflect the variability of field conditions, and add to the statistical rigour of the testing program. Storms are disqualified based on water quality if the pollutant concentrations are well above expected levels, as such results are likely to lead to an over-estimation of treatment performance. This approach also mitigates against the risk of test sites being spiked with high pollutant loads.
- Removal of requirement to undertake baseline monitoring to characterise water quality on test sites. This is only likely to be of value where a site has atypical water quality.
- Adjustment to hydrograph coverage criteria. Australian storms are much more variable than in other countries, and adoption of US or NZ storm sampling criteria would result in an excessive number of storms being disqualified.
- Reduction in minimum inter-event period to 6hrs, allowing for successive storms to be sampled.

Version 1.3

- General text and formatting improvements.
- Correction of Equation 8.