ECOLOGY
ENGINEERING
DESIGN

E2DESIGNLAB

Sydney Water

Green infrastructure stormwater retention performance report - Stage 2 calibration

Version 1.2 / June 2024

Melbourne
Tenancy 4
Level 9 Carlow House
289 Flinders Lane
Melbourne VIC 3000
PO Box 19
Darling South VIC 3145
P +61 (0) 3 9654 7274

Brisbane
8A Princhester Street
West End QLD 4101
PO Box 5945
West End QLD 4101
P +61 (0) 7 3255 1571

Kunshan, China
505 Baitang Road
Zhoushi Town
Jiangsu Province

info@e2designlab.com.au
www.e2designlab.com.au

Document Control Sheet	
Report Title	Green infrastructure stormwater retention performance report – Stage 2 calibration
Version	1.2
Author(s)	Dale Browne, Zubin Irani, Penny Galbraith, Juliette Monetti
Approved by	Dale Browne
Signed	Dale Browne
Date	27/06/2024
Distribution	Lyndall Pickering – Sydney Water

Revision	Date	Approved	Details of Revision
V1.2	27/06/2024	Dale Browne	Draft for issue

enviro DEVELOPMENT™ PROFESSIONAL

1 Executive summary

The protection of waterways requires management of both stormwater flow volumes and frequency as well as quality. In Australia, the stormwater industry has focused mostly on stormwater quality for the last two decades. There is growing recognition that stormwater flow patterns and volumes must also be managed This recognition has led to prior work on stormwater flow volume objectives, commonly referred to as mean annual runoff volume (MARV) objectives (Sydney Water & E2Designlab, 2019).

In response to the recognised need to manage stormwater flow volumes, stormwater flow volume objectives have been established for the Western Sydney area for the Wianamatta/South Creek catchment. These are described in the DCP (NSW DPIE, 2022) and the Technical Guidance for achieving Wianamatta-South Creek stormwater management targets (State of New South Wales and Department of Planning and Environment, 2022).

Biofilters, also known as bioretention or raingardens, are widely used for stormwater pollutant treatment in Australia and other parts of the world. Increasingly, the potential for these assets to reduce stormwater volumes and allow more natural flow patterns to be delivered to waterways as well as reduce flooding is being recognised. There is a need to better understand their likely performance and Stage 1 of this project was undertaken to review the current state of knowledge of likely biofilter stormwater retention performance (Browne et al., 2020).

The MUSIC stormwater model is a conceptual continuous simulation link-node model with a simplified representation of a stormwater system. It is widely used for biofilter design in Australia and sometimes in other parts of the world. To improve biofilter design, users need to have confidence in the capacity of the model to predict stormwater volume reductions and understand how to configure it appropriately.

Calibration results

This study presents an assessment of how well MUSIC can match field biofilter outflow patterns and predict stormwater volume reductions for selected field sites.

The model was calibrated for three field data sets collected at different biofilters around Melbourne of varying designs. The data include a range of operational conditions with different flow rates, duration of events and antecedent conditions.

The model was calibrated and able to simulate water flows quite well for one of the three sites (Monash car park) with an NSE of 0.84. Some discrepancies were observed following longer dry periods. This may have resulted from soil cracking and macropore flows not represented in the model but this is a very good result and provides confidence for application of the model for relatively standard biofilter configurations.

For the other two sites (Wicks Reserve and Hereford Road), it was possible to calibrate the model relatively well (NSE of 0.70 and 0.71 respectively) but only using by physically unrealistic parameters. This was due to the Wicks Reserve biofilter featuring a constrained outlet which is not represented in the model and the Hereford Road biofilter having an elevated outlet which can be approximated but also affects how the model functions. The results indicate that these were approximated quite well through the modification of other parameters to compensate for the structural differences between these configurations and what the model supports. There is however an issue that such outcomes are not readily transferrable for application to other sites.

Statistical model results

The statistical models were calibrated but provided poor results for standard deviations in validation which was a key metric for success in the related earlier work (Zhang et al., 2021). Given the initial outcomes and with reflection on relevant literature it was considered that it would likely be difficult to achieve a successful statistical model without significant further work and that focus on calibration of continuous simulation-based models may be preferable at this time to better represent the range of patterns, behaviours and conditions experienced by a biofilter.

It is considered that the use of high quality consistently monitored data will be important for the success of statistical models and that broad-based fitting to the existing statistical data is likely to be challenging given the broad heterogeneity of data.

While this study has chosen to focus more on the continuous model calibration approach, this type of model remains of interest, successful precedents exist (Zhang et al., 2021)and it could be pursued through future research although at this stage continuous models look to be more promising.

Use of MUSIC for biofilters with a standard or typical configuration

We conclude that *MUSIC is capable of predicting stormwater flow volume reductions for standard biofilter configurations with good confidence, provided that it is appropriately parameterised*. It is most effective for assets with standard or typical design configurations aligned with the model structure.

Greater attention to parameterisation of the following is recommended:

- Use of an appropriate and representative ‘exfiltration rate’, ideally based on geotechnical testing of sub-soil infiltration rates. Even these can significantly under or over-estimate realised outcomes and sensitivity analysis as well as post construction monitoring where possible is recommended.
- Impervious fractions and soil parameters to more accurately estimate catchment flows (ideally with reference to calibrated values for the same or a similar catchment)

- Saturated hydraulic conductivity should be modelled at a conservative value (e.g. 100 mm/hour) for assessing treatment performance but sensitivity should also be undertaken at the design saturated hydraulic conductivity (average or ends of the specified range) to understand potential variations in performance. This *may be particularly important where outflow rates into a waterway are of interest such as in Western Sydney*.
- The PET factor should be revised to best represent the expected conditions. It will usually only be sensitive for assets with a large treatment area relative to catchment (>~5%) or where the effective area over which PET occurs varies significantly from the filter area. The default value of 2.1 may be regarded as indicative of likely performance of small to medium sized biofilters and raingardens while lower values (down to about 1) are more suitable for large scale biofilters or vegetated sponges.

Use of MUSIC for biofilters with a non-standard or a-typical configuration

Current design practice generally assumes a free outlet on biofilters. However, there are clear potential benefits to using constrained outlets on biofilters. This can increase stormwater volume retention performance which is a key area of focus. It could also allow biofilters to provide control of flow rates to provide appropriate flow patterns for waterway health as well as effective and more reliable flood mitigation outcomes. We may therefore expect increasing industry interest in simulating different configurations.

We caution that there is greater uncertainty when the biofilter model in MUSIC is used to represent non-typical configurations such as a constrained or elevated outlet, and by extension would expect greater variances in performance and uncertainty when modelling assets that vary from typical biofilters, such as passively irrigated tree pits and green roofs. The biofilter model may still be used in these circumstances given the limited range of alternatives. However, sensitivity testing and recognition of greater uncertainty levels should be considered in design as well as potentially using alternate models as a cross-check.

There is a clear opportunity for new or improved biofilter models to be developed that can support a broader range of assets that can be broadly classified as biofilters and structural configurations including a constrained outlet. We note that SWMM already has a constrained outlet (Lisenbee et al., 2021) and may represent a logical starting point for development of an improved biofilter model.

There are model input parameters with significant uncertainty. Current industry practice is to use the model deterministically using a single input value for these. There is a need for greater awareness of the range of uncertainty in these values and the implications for results.

It is recommended that a guide (or addendum to an existing guideline) is prepared to provide guidance to industry on sensitivity testing required. This should ideally involve input from both researchers and industry to ensure it is both practical and robust.

There is currently little or no monitoring, assessment or field testing to confirm that assets deliver on their design performance. Historically, this has been largely due to cost, primarily for stormwater quality sampling. However, where many large biofilter assets are being constructed, these may involve substantial investments of both public and private funding. The absence of testing and validation mean that these *significant investments may at risk* or not be as well targeted as they could be. *Even a* ***modest investment in testing could yield significant increases in environmental benefits realised and/or cost savings***.

It is recommended that a program is established to fund and progress field monitoring of a proportion of constructed biofilters (and other WSUD assets). The program should include model development, calibration and improvement to support assessment of performance and improved design guidance for biofilters. This program would ideally be expanded to include a broader range of biofilter asset types including green roofs, wicking beds and vegetated sponges.

CONTENTS

2 Introduction

The protection of waterways requires management of both stormwater flow volumes and frequency as well as quality. In Australia, the stormwater industry has focused mostly on stormwater quality for the last two decades. Guidelines for stormwater quality were established in the 1990's (Victoria Stormwater Committee, 1999). These have provided the basis or inspiration for most other guidelines and continue to be the primary reference for planning applications (WBM, 2015).

There is growing recognition that stormwater flow patterns and volumes must also be managed and stormwater guidance for stormwater flow volume reductions has been introduced.

The importance of restoring flow regimes to address stream health is well documented and increasingly recognised (DeBusk et al., 2010; Walsh et al., 2012). It has been suggested that source control efforts for harvesting, evapotranspiration and infiltration can help to achieve more natural levels and patterns of surface runoff (Burns et al., 2012; DeBusk et al., 2010).

New guidelines have recently been introduced within Victoria and New South Wales that add stormwater volume management as another layer requiring consideration (EPA, 2021) (EPA Victoria, 2021; NSW Department of Planning and Environment, 2022).

The NSW guidelines (for the Wianamatta/South Creek catchment in Western Sydney) set out a stormwater volume target of 2 ML/ha/year for urban development. This is higher than natural conditions (~1-1.5 ML/ha/year) but much less than fully urbanised conditions (~5-6 ML/ha/year). These are further augmented with specified targets or 'gates' on the flow duration curve and an alternative option consisting of no flow volume objective but a more stringent set of flow duration curve 'gates', see Figure 1.

The overall aim of these objectives is to minimise changes in waterway hydrology due to urbanisation. It is desirable that changes to the overall waterway flow duration curve should be minimised. Many potential indicators have been identified in the literature and may potentially be used (Duncan et al., 2014). It is acknowledged that pursuing stormwater volume reductions alone does not guarantee success in achieving the desired waterway flow patterns, however, they are recognised as a necessary inclusion or pre-cursor to making any meaningful progress. Simply attenuating and modifying flow patterns *using detention approaches alone will achieve little when flow volumes are 300-500% greater than natural conditions*.

The flow volume objectives are additional to the existing historical stormwater management objectives for flood mitigation and stormwater quality. The following is an example of typical planning requirements applicable for the Wianamatta/South Creek catchment in Western Sydney but may be taken as broadly indicative of what is needed to protect waterways throughout Sydney:

- Flooding
    - Mitigate flows to pre-development conditions for the 20% to 1% annual exceedance probability (AEP) event
    - Provide piped infrastructure for flows up to the 20% AEP event
    - Ensure overland flows are safely conveyed to avoid unsafe conditions for pedestrians and vehicles
- Stormwater quality - Reduce mean annual loads of pollutants
    - Gross pollutants: 90%
    - Total suspended solids: 90%
    - Total phosphorus: 80%
    - Total nitrogen: 65%
- Stormwater flow volumes (at point of discharge to local waterway)
    - Mean annual runoff volume (MARV): <= 2 ML/ha/year
    - 90%ile flow: 1,000 to 5,000 L/ha/day
    - 50%ile flow: 5 to 100 L/ha/day
    - 10%ile flow: 0 L/ha/day

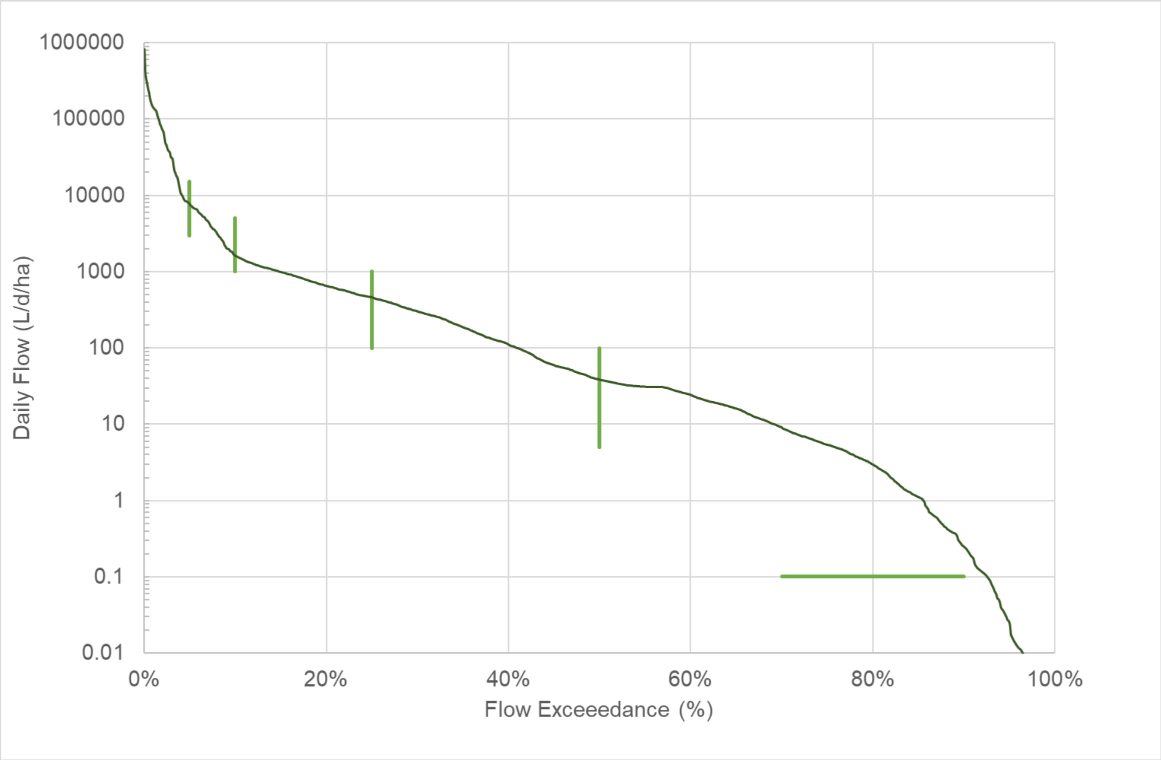


Figure 1 A flow duration curve demonstrating (mostly) compliance with the targets or 'gates'. These are largely achieved by stormwater harvesting but also significantly influenced by outflow rates from assets including bioretention at the lower end of the range

These objectives point to a potential major shift in focus and effort for stormwater management through the introduction of stormwater flow volumes and a greater focus on the necessity to achieve desirable waterway flow patterns to effectively protect downstream urban waterways.

These objectives are an important next step for the stormwater industry in its efforts to protect waterways from the impacts of urbanisation.

Sydney Water want to better understand the performance of different water sensitive urban design and stormwater green infrastructure interventions. This is particularly important given the context of the above guidelines that recognise the impact of urbanisation on stormwater flows and waterways and set corresponding stormwater flow volume reduction objectives for waterway protection. There is also potential for stormwater to be used to enhance liveability through increased urban greening.

The prior stage, Stage 1 of this study is documented in E2Designlab (Browne et al., 2020). The purpose of stage 1 was to identify, collate and review performance data for different types of green infrastructure assets with a focus on the following:

- Biofilters (also known as raingardens, bioretention)
- green roofs
- passively irrigated tree pits

The (Stage 1) study also sought to understand performance of other green infrastructure such as swales, simple grass buffer strips and permeable paving. However, there is relatively lesser performance data available for these and resources were therefore focussed on the first two asset types as well as some consideration of passively irrigated tree pits.

The basic mechanics of these WSUD asset are generally understood, but there has been limited attempts to improve our understanding of performance and reasons for differences in performance, leading to gaps in understanding.

There has been some excellent individual research, but no overall platform to synthesise or publicise learnings to improve learning and inform improved design and planning. There is no database of performance and little ability to share WSUD performance information.

Biofilters are recognised as an attractive and most likely effective water sensitive urban design (WSUD) response (also called low impact design (LID), sustainable urban drainage system (SUDS), nature-based solution (NBS) or green infrastructure).

It is hypothesised that biofilters designed to support infiltration that are sized adequately relative to catchment may achieve significant surface stormwater volume reductions through infiltration and evapotranspiration and this has been supported by the outcomes of the Stage 1 study.

The MUSIC stormwater model is a conceptual continuous simulation link-node model with a simplified representation of a stormwater system. It is widely used for design of stormwater management assets including biofilters in Australia. To improve biofilter design, users need to have confidence in the capacity of the model to predict stormwater volume reductions and understand how to configure it appropriately. It is noted that the SWMM model is more widely used for design elsewhere in the world although in Australia it is predominantly used for research.

The available evidence suggests that biofilters are likely to be effective for stormwater retention performance. However, actual expected outcomes and the accuracy of industry models to predict these are uncertain. Indeed, the MUSIC software model was initially developed to predict stormwater quality outcomes. Prediction of stormwater quantity is obviously also necessary for relative estimates of pollutant loads, although requires less modelling accuracy than that needed to accurately partition flows into outflows, overflows, infiltration and evapotranspiration. The science has developed significantly in the decade since the MUSIC biofilter model was last significantly upgraded. Some practitioners have questioned the accuracy of standard models in combination with existing guidelines (Hoban & Gambirazio, 2018, 2021) while others have expressed confidence in the models (Tanner et al., 2020a). More broadly, it is recognised further work is needed to calibrate and validate the models to build confidence (Lisenbee et al., 2021). In reviewing the literature on bioretention, Spraakman only identified 3 occurrences of MUSIC in the literature for assessment of hydrology, relative to 44 for SWMM. There is a clear need to review field monitoring data and re-assess the adequacy of relevant industry models to successfully predict stormwater retention performance.

There is a clear need to confirm stormwater retention performance for biofilters and MUSIC's prediction of this given limited calibration of the model to date as well as the limited success of some of these studies (Fowdar et al., 2022; Hennekam, 2021; Imteaz et al., 2013; Lintern et al., 2012; Parker, 2010). Randelovic (Randelovic et al., 2016) has also demonstrated successful calibration of a similar model (MPiRe) to individual storm events for the Monash biofilter. In contrast to the limited references for MUSIC, the SWMM model has been extensively studied and calibrated (Platz et al., 2020a)(Peng & Stovin, 2017).

Some key research questions have been identified as follows:

1. How much stormwater volume reductions can be achieved using biofilters
2. How well do the models reflect observations?
3. What are the key learnings from the research and how can performance data best inform water and land use planning?
4. What are the information and modelling gaps hindering us?

Stage 1 of this work focussed on addressing question 1 by identifying what stormwater volume reductions are being achieved by monitoring biofilters and other stormwater assets in the field.

Stage 2 has focussed on addressing question 2 using statistical analysis and interpretation of the data and by calibrating an industry-standard model with monitored field data.

Stage 2 has then sought to address questions 3 and 4 to draw out key learnings from the research and first two stages, identify gaps and to communicate these with industry.

2.1 Scope of investigation

This project encompassed the following tasks:

- Revision and update of stormwater performance data from Task 1
- Development of a multi-variable model to assess stormwater volume reductions
- Calibration of MUSIC to observed field data
- Interpretation and documentation of outcomes
- Communication of results and learnings for planning, modelling and implementation

2.2 Objectives

- The objectives of the overall study were to:
- Review field data of raingarden assets to understand performance, what the most influential factors are in design to promote volume reductions and the ability of the models to predict performance (Stage 1, prior report)
- To determine whether a multivariate model could be used to predict stormwater volume reductions seen in field data (This report)
    - Measured through the NSE values of calibration and validation datasets
    - Compared against % reductions in mean flow volumes (annual or period of record)
- To test whether selected field data can be calibrated within MUSIC and understand MUSICs ability to reflect realistic stormwater flow volume reductions and hydrographs.
    - Measured through the NSE values of calibration datasets
    - Compared against % reductions in mean flow volumes (annual or period of record) (This report)
- Provide recommendation for design, modelling and future research (this report)

3 Data review and analysis

3.1 Overview of data from literature

A review of available literature was undertaken in stage 1, focusing on identifying reported percentage stormwater volume reductions for field monitored biofilters. Furthermore, detailed hydrographs were also sought where such data was available with the intention of collating data to support model calibration.

This is documented in stage 1 and the reader is referred to the Stage 1 report for further detail while key outcomes are briefly summarised below.

Several meta-studies explore the effectiveness of bioretention systems for stormwater management including volume retention (Hoban & Gambirazio, 2018), (Hoban & Gambirazio, 2021), (Poresky et al., 2012), (Davis et al., 2012), (Spraakman et al., 2020). The Australian meta-study (Hoban & Gambirazio, 2021) found bioretention assets achieved a weighted average reduction volume of 51%. Studying US bioretention assets, Poresky and Liu found slightly higher results as shown in Table 1.

In this study, with an expanded international data set, we found broadly similar results with a mean of 53% and weighted mean of 58%..

Table 1 Stormwater volume reduction performance summary for biofilters

Paper ID	N	Min	Mean	Weighted mean	Median	Max
This study	26 assets 1596 events	8%	53%	58%	59%	87%
Hoban and Gambirazio (2021)	15 assets 513 events	-	-	51%	-	-
Poresky et. al. (2012)	20 assets 2074 events	11%	64%	-	65%	100%
Liu et al. (2014)	23 assets	0%	67%	-	75%	100%

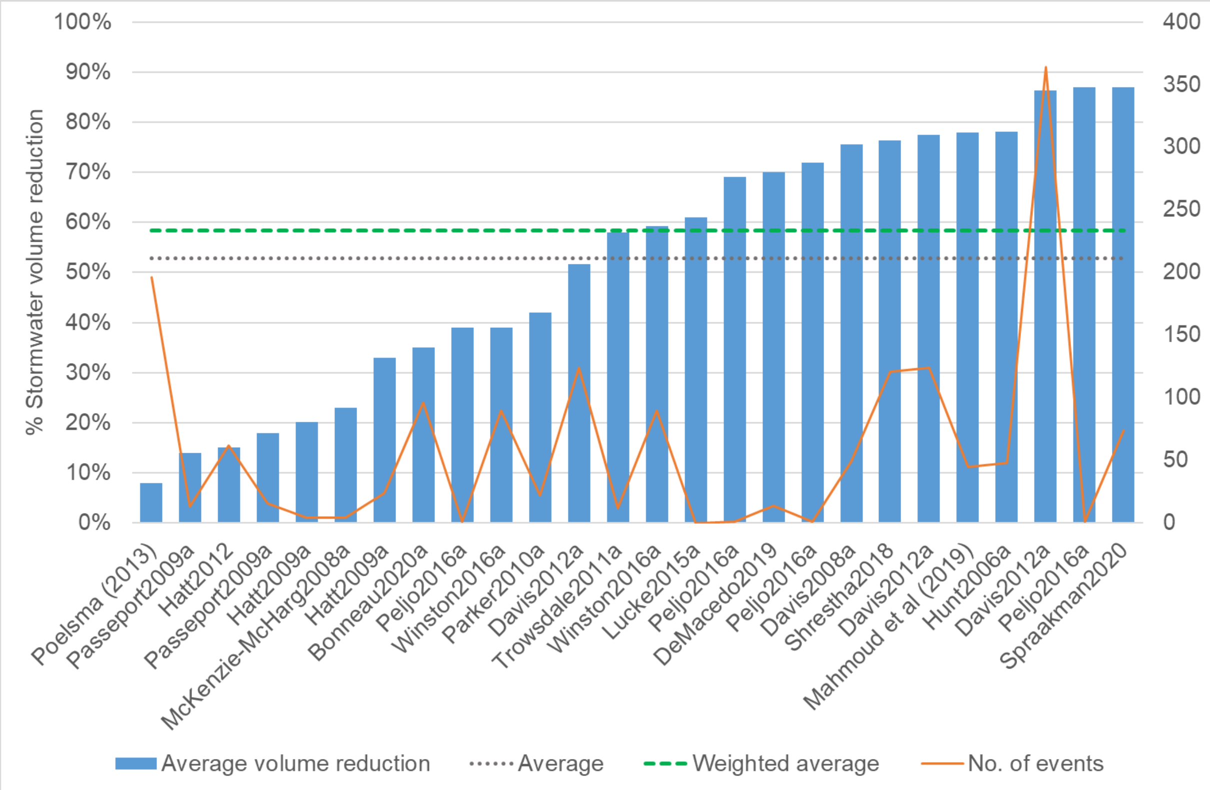


Figure 2 Stormwater volume reductions in bioretention

As may be expected, there is a broad range of results across the spectrum of possible outcomes as indicated by the minimum, maximum and distribution. It is however promising that the median and mean outcomes suggest substantial stormwater volume reductions in the range of 50-60% and only a handful of studies are at the upper and lower ends of the range.

The key findings from the data and review of the relevant papers can be briefly summarised as follows:

- Biofilters are generally effective for reducing stormwater volumes.
- Performance varies widely (as expected) and depends on climate, soils, design, size and other factors
- Biofilters are typically small relative to catchment and as a result infiltration (rather than evapotranspiration) is usually the dominant pathway for (surface) stormwater volume reductions to be achieved. *Where possible, biofilters should be unlined to enable infiltration and support waterway baseflows*.
- Satisfactory outcomes *may still be achieved in slow draining soils* or lined assets subject to context, design and potentially larger sizing than required for stormwater quality purposes
- Current modelling approaches with MUSIC in combination with current guidelines may tend to underestimate performance as suggested by Hoban and Gambirazio (2021) although in reviewing this, others have expressed a contrasting view (Tanner et al., 2020b). We note

that contemporary guidelines are moving away from suggesting zero infiltration rates to using rates based on geotechnical testing (E2Designlab for Melbourne Water, in prep) although this is not universal with the recent Western Sydney guidelines requiring design and modelling for zero infiltration due to potential sodic soil conditions.

- Generally, it is recognised that there are opportunities to improve guidance provided to practitioners on appropriate model parameterisation. Further calibration work is needed to support better guidance on modelling hydrologic performance. There may also be opportunities to improve process modelling.

3.2 Detailed data sets for modelling and calibration

The literature review was followed up by contacting researchers to request whether access to the original data including inflow and outflow hydrographs could be made available for further study and calibration. A number of researchers generously shared their data and these were compiled for six bioretention assets while a further two that were publicly available were obtained.

From researchers:

- Monash cark park biofilter
- Wicks Reserve biofilter
- Hereford Road biofilter
- Walker Street biofilters (2 cells)
- Wakerley biofilter (3 cells) (event data only)
- Kortright biofilter

Public (US EPA):

- Graham H.S. parking lot biofilter
- Villanova Traffic Island biofilter

While not documented further here, data for three additional biofilters have recently been sourced by Mohamed Sabbagh at Monash University (Pers. Comm, Mohamed Sabbagh, 2024):

- Ursuline College biofilter
- Holden North biofilter
- Holden South biofilter

From these, three high quality data sets were selected for the flow volume reduction analysis and calibration. These are continuous or event-based data sets recorded over 1- or 6-minute time steps. Gauges were set up at the inlet/s and outlet/s to measures all incoming and outgoing flows in the systems. The following data sets were analysed in this study:

- **Monash car park biofilter** (Hatt et al., 2009): continuous data sets for a three-cell raingarden located at the carpark at Monash University Clayton campus in Melbourne for a period of 5 months.
- **Wicks Reserve biofilter** (Bonneau et al., 2020): event-based data set for a large biofilter in the eastern suburbs of Melbourne for a period of 3 years. Due to missing recordings of rainfall data, rainfall events and periods between recorded rainfall events were infilled with modelled data (but do not form part of the calibration comparison)
- **Hereford Road biofilter** (Poelsma et al., 2013): continuous data for a raingarden located in the eastern suburbs of Melbourne for a period of 9 months.

Each of these sites had a good length of reasonably complete continuous flow data. The configuration of each asset varies. The Monash car park biofilter is a relatively standard biofilter with lining and a submerged zone (also called saturated anoxic zone) below the outlet invert. Wicks Reserve also has a submerged zone but is unlined to encourage infiltration and the outlet is also constrained. Hereford Road is similar but the underdrainage configuration is relatively unique, returning to the overflow pit and draining out at a level 100 mm below the surface to prevent long ponding periods.

Walker Street (Clifton Hill) was not chosen after considering the configuration and flow data. The site has a sediment pond which then directs flows to two biofilter assets, one downslope of the other. The outflows from the second biofilter are higher than the inflows. This suggests that infiltrated flows from the upstream biofilter may re-emerge at the second biofilter. These complex conditions would make calibration for this site challenging.

For Wakerley, we were only able to access limited event data although longer continuous data reportedly exists. This biofilter is also very large which can make effective monitoring of flows more challenging. While it is likely to be challenging and was not prioritised for further assessment, his asset could potentially be explored in future work if the full datasets can be accessed.

Data for the Kortright biofilter in Canada was received after work had commenced on Stage 2 so was not included in our analysis and calibration but has been the subject of analysis and calibration within parallel work by Mohammed Sabbagh.

Two further sites for biofilters were also available publicly from US EPA. These were of lower direct relevance for Australian conditions but of interest to pursue in future studies, particularly for potential comparison of calibration outcomes between MUSIC and SWMM.

4 Assets and data for calibration

This chapter contains a summary of the assets and data used for calibration.

4.1 Field measurements used for model testing

4.1.1 Monash Car Park Biofilter

Details of the Monash Car Park Biofilter are summarised below for reference. These were sourced from (Hatt et al., 2009) and also (Zhang et al., 2014) and the reader is referred to these papers for further details.

The Monash Car Park Biofilter is located at the Clayton campus of Monash University in Melbourne and was constructed in 2006 to treat runoff from the top level of a 4,500 m^2 multilevel car park with treated flows then used for irrigation of a sports oval. The catchment can be considered 100% impervious.

Monash University's Clayton campus receives around 680mm of precipitation per year, with an average maximum summer temperature of 26 degrees and average minimum winter temperature of 6 degrees.

The biofilter asset has a surface area of 45m^2 with a depth profile of 500mm filter media, 100mm transitional sand and 100m fine gravel. The biofilter is lined, being encased in a concrete box.

The biofilter consists of three separate cells, each with different filtration layers and plant covers to test performance of different soil-based filter media. Only two cells were used in this study. The original configuration was in place for the continuous flow data collected and used in this study. It was later modified with one of the cells replaced with a new configuration matching the FAWB specifications (FAWB, 2009) current at the time and used for the 'challenge tests' consisting of synthetic storm events. These were used in later papers by Zhang and Randelovic (Zhang et al., 2014) (Randelovic et al., 2016).

According to Hatt et al, (2009) the original configuration was as follows:

- All cells have a 500 mm filter media layer and a 200 mm drainage layer
- Cell 1 has filter media consisting of sandy loam
- Cell 2 has filter media consisting of sandy loam 80%, 10% vermiculite, 10% perlite
- Cell 3: sandy loam mixed with 10% compost and 10% light hardwood mulch (by volume).

Revised configuration (Zhang et al., 2014)

- Cell 1 is a biofilter without a submerged zone, and a filter media consisting of loamy sand as recommended by design guidelines at the time (FAWB, 2009); (sand 84.2%, silt 3.0%, clay 12.8%) and Carex apressa as the main plant species.
- Cell 2 contains a submerged zone, with filter media consisting of sand (sand 96.0%, silt 0.8%, clay 3.2%)
- Cell 3: sandy loam mixed with 10% compost and 10% light hardwood mulch (by volume).

Each cell was planted with the same mix of native rushes and sedges including Carex appressa, Carex tereticaulis, Lomandra longifolia, Isolepsis nodosa, Caleocephalus lacteus and Juncus spp (Hatt et al., 2009).

For the purposes of this study, the continuous data and configuration documented by Hatt (Hatt et al., 2009) was used.

6.1.4 Monitoring setup

The system is fully equipped for monitoring both flow and water quality. Three V-notch weirs installed in the covered inflow chamber are used to monitor inflow into the biofilters. The outflow from each cell is monitored by three small separate V-notch weirs. Overflows were not monitored.

6.1.5 Data available

A review of the data was undertaken to ensure that the data selected for the analysis was cleaned of as many obvious irregularities or noise in the recordings as possible. This predominantly included examples such as events where the outflows were recorded as being greater than the inflows.

A period of continuous monitoring data was available in 2007. This extended notionally from January to November 2007 with flow data collected for a total of 28 storm events. Inflow and outflow data was recorded, measured at a 1 minute time step.

It was found that the main monitored period was from 20th January to 31st May 2007. There were a further two events in November following a long gap in the monitoring record. However, the recorded flows for these were very different and outflows greatly exceeding inflows were reported for these events. Since this strongly suggested potential issues with these later events, these were excluded. Focus was therefore placed on calibration for the 5 month period from 20th January to 31st May 2007.

It was also identified that the high-flow bypass was engaged during 11 of the 28 monitored storm events. The bypass weir was not rated, therefore overflow volumes for these were unknown. It was assumed that overflows bypassed the monitoring and for the calibration measured flows and modelled drain outflows (which excludes overflows) were compared. Inflow and outflow data is recorded, measured at a 1 minute time step.

It was reported for the 17 runoff events without bypass that on average 33% of the inflow volume was retained with a range of 15-83% (Hatt et al., 2009).

The system has experienced issues with leakages and when the asset was reset, cracks in the concrete base were found. Analysis of results found that this leakage likely impacted the results of the monitoring study (Tanner et al., 2020a). This should also be considered when using the data and pursuing calibration.

The Monash biofilter is well suited for analysis of a system with a relatively low level of infiltration (being lined but leaky). A good level of continuous data is available.

4.1.2 Wicks Reserve

6.2.1 Site and climate

Details of the Wicks Reserve biofilter are summarised below for reference. These were sourced from Bonneau et al, (Bonneau et al., 2020) and the reader is referred to this paper for further details.

A biofilter was constructed at Wicks Reserve. It is located in the Dobson's Creek catchment in Boronia in the eastern suburbs of Melbourne Australia. The creek is a sensitive receiving waterway and the objective was to reduce direct surface stormwater discharges using a combination of subsidised household tanks and the biofilter. The 33 hectare catchment is peri-urban and was estimated to have 15% impervious surfaces.

Average annual rainfall in the catchment is 730mm, evenly distributed throughout the year with a slight winter-spring bias. Average annual areal potential evapotranspiration is 1050mm and higher in summer, displaying strong seasonality.

6.2.3 Asset

Runoff from impervious areas of the urban catchment reaches the catchment via two stormwater pipes which combine in a junction pit. Flows of up to 200 L/s are diverted to a gross pollutant trap (GPT) while higher flows bypass the stormwater management assets and are conveyed to the local creek. Bypass flows are monitored with a narrow crested rectangular weir. Flows from the GPT then flow through a sediment pond of approximately 175 m^2 and depth of 0.8 m. The GPT intercepts large particles such as litter, gravel and some coarse sediment. The sediment pond provides further treatment of coarse sediment (target particle size >125 μm).

Outflows from the sediment pond spill evenly across the basin filter media. The biofilter is 1,800 m^2 or approximately 4% of the upstream impervious catchment. This is larger than typical stormwater quality assets constructed in the region (commonly 1-2% of impervious catchment or less). The larger size was an intentional design choice to increase infiltration and replenish depleted stream baseflow. The sediment biofilter is located in heavy clay soil (hydraulic conductivities was measured between $5e^{-8}$ m/s and $5e^{-7}$ m/s [0.005 m/day to 0.05 m/day]. The biofilter is densely vegetated with grasses and rushes.

The biofilter is on average 0.8m deep in total. The top 0.35m filter media layer consists of loamy sand and the bottom 0.3 m is scoria to provide temporary water storage. Three 0.05m transition layers (medium-fine sand, coarse sand and gravel respectively) separate these layers. There is a slotted underdrain at the base of the basin which discharges through an elevated orifice in a discharge pit. The orifice is elevated by 0.5 m from the bottom of the basin, meaning that the bottom 0.5 m including the scoria storage and transition layers act as a submerged zone.

Only the southern side of the basin is lined to prevent intrusion of upslope groundwater into the basin while the remainder is unlined to allow infiltration to the surrounding soil. The asset contains an extended detention depth of 0.5m before discharging into an overflow pit.

Bonneau et al., (2020)reports that in winter, the surface of the bioretention basin can be covered by water for extended periods of time, up to a couple of weeks. In summer, the surface of the basin is covered with water for a few hours after a rainfall event or a few days at most in the case of a large event.

Figure 3 The Wicks Reserve Biofilter (a) GPT installation, (b) Biofilter construction, (c) establishment, (d) established, Sources (a) and (b) Dale Browne, (c) and (d) Clearwater case study (Clearwater, 2012)

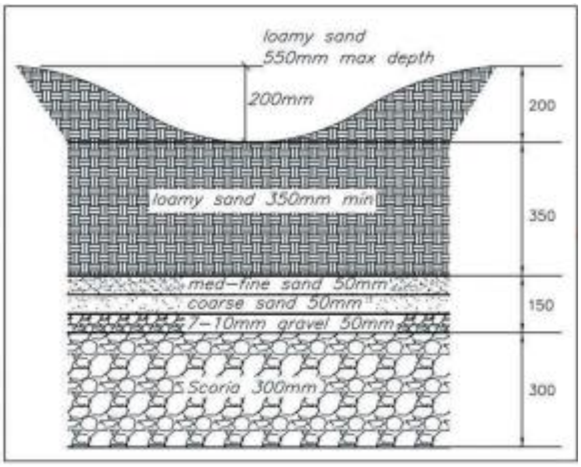


Figure 4 The biofilter cross-section (Bonneau et al., 2020)

6.2.4 Monitoring setup

Flow monitoring consists of flow gauges for upstream, downstream and bypass flows. There are also three water level probes in the biofiltration system and a rain gauge on site.

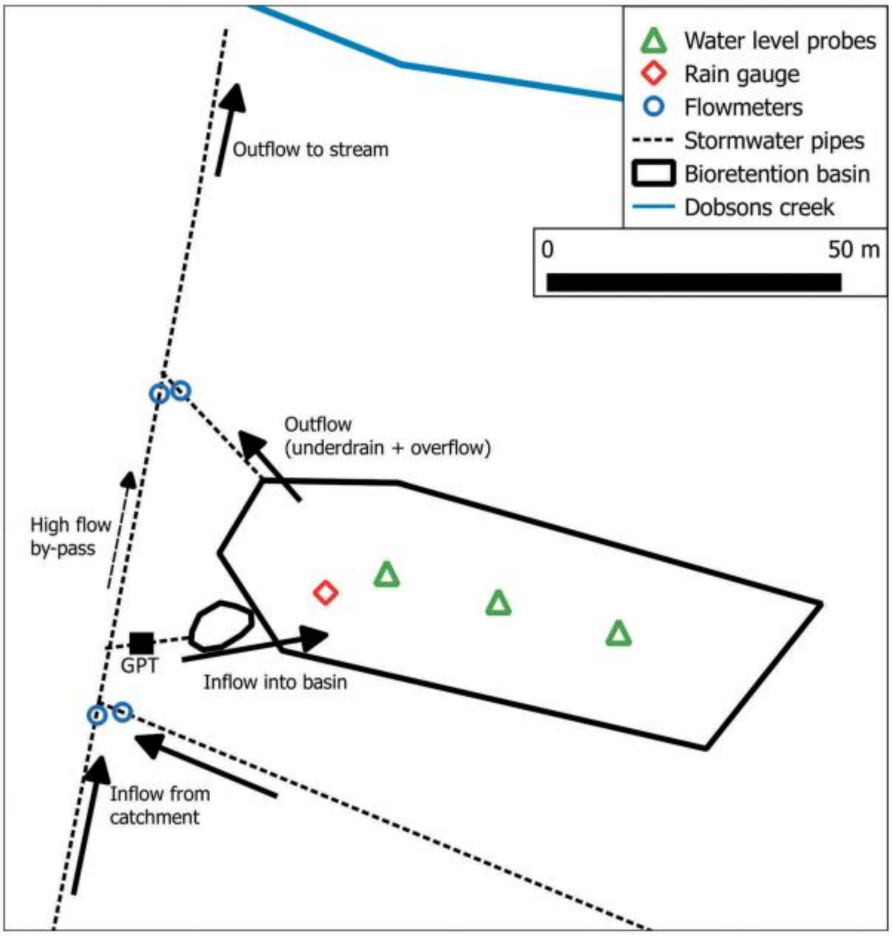


Figure 5 Plan view of the biofilter basin and monitoring systems. Inflows from the catchment are monitored in each of the two inflow pipes, outflows and overflows from the basin are monitored at the outflow pipe and bypass flows are monitored on the high flow bypass.

6.2.5 Data available

Flow data was collected between September 2013 and September 2016 while water level data was collected between March 2013 and December 2019. 96 natural events were monitored over the period of three years. Inflows (two gauges), outflows and bypass flows are available at a 6 minute timestep. Water level data for the bioretention system were available at a 1 minute timesteps for a continuous 6-year period. Sediment accumulation impacted flow data collection for certain periods. These missing periods were infilled using a linear regression from the 96 usable rainfall events captured.

Wicks Reserve is a relatively large and complex system but has an excellent and large data set available making it an attractive option for calibration and validation.

4.1.3 Hereford Road

Details of the Hereford Road biofilter are summarised below for reference. These were sourced from (Poelsma et al., 2013) and the reader is referred to this paper for further details.

The biofilter treats runoff from a 9,800 m^2 impervious catchment comprising of roads (6,170 m^2), roofs (3,050 m^2) and some other paved areas (580 m^2). The system has a surface area of 100 m^2, or approximately 1% of the impervious catchment area.

The biofilter is vegetated with indigenous sedges and shrubs. The extended detention depth is 300mm. Below this there is 400mm of filter media (loamy sand), a 200mm transition layer (sand and fine gravel) and 400mm of coarse aggregate (scoria). Treated outflows are allowed to discharge via riser pipes extending from the bottom of the biofilter. These allow water to drain to 100mm below the surface to prevent long surface ponding periods (Figure 2). This is an a-typical design with a relatively high effective outlet and deep submerged zone.

The system is not lined to allow exfiltration of water into surrounding soils. Point infiltration rates of the surrounding soil were approximately 3.8 mm/hr in the upper 300 mm, increasing to around 15 mm/hr in the lower depths.

The closest BOM station of Montrose records an annual rainfall of 1,027 mm annually. Maximum average temperatures of 26 degrees in February and a minimum average temperature of 13 degrees in July. Average precipitation is higher in winter, but there is significant rainfall all year round.

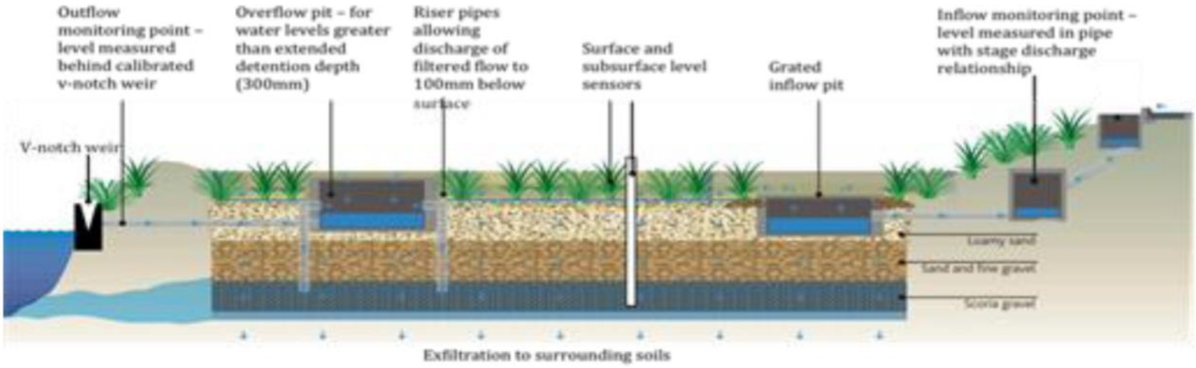


Figure 6-5 Hereford Road biofilter, after (Poelsma et al., 2013)

Data available
Available data from the monitoring period includes rainfall, PET, inflow and outflow rates. There are 9 months of data available with flows collected at a 1 minute time interval.

Two large rainfall events occurred for which flows could not be accurately measured. These were removed from the data set.

This site is well suited for calibration and validation given the length of continuous data available. However, the a-typical design configuration is likely to present some challenges.

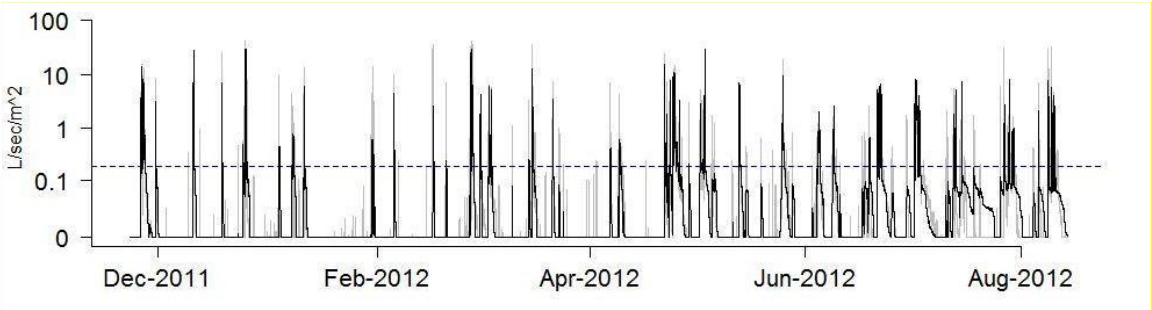


Figure 6-6 Hereford Road long-term hydrograph at outlet (Poelsma, Fletcher and Burns, 2013a)

4.2 Performance results

The anticipated performance of the assets based on the literature papers was reviewed to understand the likely stormwater retention performance as well as relative proportions of infiltration and evapotranspiration.

For Monash (Hatt et al., 2009), 28 events were observed, however overflows were not rated and therefore only the 17 smaller events not causing overflows were used to quantify the water balance. The results indicated that on average downstream outflows were reduced by 33% of the inflow volume although the breakdown of this into infiltration and evapotranspiration was not quantified.

For Wicks Reserve (Bonneau et al., 2020), inflows were modelled and infilled for intervening gaps in the data to provide a continuous data set. Outflows and overflows were both monitored and infiltration was estimated based on changes in water level. The results indicated that most of the inflows discharged as outflows (55%) with significant retention through infiltration at 31% and evapotranspiration being a relatively minor contributor at 4%.

For Hereford Road (Poelsma et al., 2013), infiltration and evapotranspiration were estimated using sub-surface level data during dry periods with infiltration taken to be level changes occurring overnight from 10 pm to 5 am when evapotranspiration would be minimal with additional level change during daytime attributed to evapotranspiration. This indicated that infiltration was clearly dominant for the asset and associated climate and soil conditions in north-eastern Melbourne. This was observed despite relatively heavy clay soils and low infiltration rates in the area.

Table 2 Stormwater volume reduction performance summary for biofilters

Biofilter	**N events**	**% outflow**	**% overflow**	**% infiltration**	**% evapotranspiration**
Monash Car park	17*	67%	-	33%	
Wicks Reserve (Bonneau et al., 2020)	96	55%	5%	31%	4%
Hereford Road	-	70%**		8.9%	1.4%

*Data for 11 of 28 events excluded as overflow not rated

*Percentage of rainfall rather than inflow

4.3 Prior calibration work for MUSIC and sites of interest

Calibration of MUSIC to field data has been undertaken within a number of studies and these are identified along with key learnings below. SWMM has been more extensively studied and selected key papers for this are also discussed below.

4.3.1 MUSIC

Lintern et al (Lintern et al., 2012) assessed the performance of a stormwater biofilter model, MUSIC Version 5 (eWater, 2024), using storm event data from the Monash Car Park biofilter, situated in Clayton in the south-eastern suburbs of Melbourne. The results indicated the model underestimated retention performance with modelled outflow volumes being 12-32% less than observed. It was able to replicate some of the observed peak flows but was not effective for representing low flows. Two objective functions were used, the coefficient of efficiency (Nash & Sutcliffe, 1970) which favours high flows and a more unbiased objective function that favours low flows.

The results indicated significant differences in the calibrated hydraulic conductivity, depending on which objective function was used with the results for the unbiased function being closer to the measured values. On the other hand, total outflow volumes (of primary interest for this study) were modelled more closely using the coefficient of efficiency.

The study found that the model was able to predict some hydrographs accurately but that there were a considerable number of events, particularly smaller events where flow rates and durations were not accurately predicted, resulting in underestimations of total outflow volumes of 12 to 35%. It was noted that further work may take into account temporal and spatial variations in hydraulic conductivity as well as the specified parameters for soil moisture at field capacity and the water stress point. It was noted that these latter parameters were determined for biofilter columns in a laboratory setting planted with Carex appressa and that they may need to be calibrated for a broader range of soil types since they theoretically depend on soil texture.

Imteaz et al., (2013) reviewed the performance of two biofilters, one that was new and one that was approximately 6 years old within the City of Manningham in Melbourne's north-east. Storm events were simulated using a water truck discharging into the upstream drainage while outflows were monitored. Three tests were performed with different flow rates for each bioretention. The systems were soaked prior to the tests to create saturated soil conditions.

A hydraulic conductivity of 100 mm/hour was assumed for both assets. The assets are relatively large with filter areas of 332 and 506 m^2 and surface areas of 1,506 and 800 m^2 respectively for the Habitat Park and Yarraleen biofilters.

The assets were simulated using MUSIC and inflow events of similar size to those observed were selected for comparison. The modelled outflows for these events were then compared with the measured outflows.

Despite the limitations of the study, the modelled and measured flow volume reductions were relatively close, being in the order of 86-91% which suggests that most of the flow volumes were retained within the biofilters or infiltrated.

Fowdar et al. (Fowdar et al., 2022) investigated the performance of a biofilter, wetland and swale in Punggol, Singapore. The assets were modelled using MUSIC and the model was calibrated and validated to the monitored data with the aid of the PEST optimisation software. Nash Sutcliffe Efficiencies (NSE) for event outflow rates ranged from -3.13 to 0.86 were achieved with similar performance for validation (-1.49 to 0.71). Estimation of peak flows tended to be under-estimated to varying degrees. This was assumed to be due to heterogenous behaviour of the biofilter including preferential flows and partial engagement of the surface area while uncertainties in inflows may have contributed. For event volumes, the NSE was -0.24 for calibration and 0.34 for validation. Cumulative modelled volumes were within 21% of measured values and considered acceptable. It was noted that while the model shows promise in predicting shape and timing of peak flows reasonably well, it may under-perform for assets with features significantly different to guidelines such as being over-sized.

4.3.2 MPiRE

MPiRE is a research bioretention model developed for the purposes of assessing micro-pollutant removal. The model is based on a bucket approach for water simulation, similar to MUSIC while while more sophisticated pollutant removal algorithms are proposed.

Randelovic et al (2016) calibrated the model to a series of challenge tests conducted at the Monash Car Park biofilter (discussed above). It was found that the model was able to simulate flows reasonably well although there were some discrepancies for long dry periods with good to very good Nash-Sutcliffe efficiencies from 0.62-0.90 achieved for outflows. Soil moisture was also considered and it was found that MUSIC performed very well for a standard biofilter but less well for the cell with a high organic content.

4.3.3 SWMM

The biofilter module in SWMM has had numerous calibrations undertaken. A couple of these are briefly outlined below.

Lisenbee et al evaluated the performance of SWMM and DRAINMOD-Urban for representing a biofilter at Ursuline College near Cleveland, Ohio, USA. The catchment was 3,600 m^2 consisting mostly of car park draining to a 182 m^2 bioretention.

They found that while DRAINMOD-Urban better matched the shape of drainage hydrographs, SWMM provided closer estimates to measured event volumes with an NSE even when uncalibrated of 0.7. The authors explored different calibration approaches including calibrating to hydrographs and to event volumes. They found the initial NSE (for event volumes) could be further improved by calibrating for event volumes to obtain a result of 0.86 or become worse when calibrating to hydrographs to 0.59. These results suggest a potential trade-off between fitting to event volumes and hydrographs with SWMM. Review of the outflow hydrographs indicate that drainage rates are truncated and this was explained by the percolation equation used in SWMM with outflows limited by saturated infiltration rates. Monitored data suggest the bioretention never achieves full saturation even with ponding. This suggests that there may be more complex processes occurring (such as trapped air pockets) that are not fully represented in the models. The overall outcomes indicate SWMM and DRAINMOD-Urban are capable of achieving good results with SWMM appearing to be better for event volumes and DRAINMOD-Urban for matching hydrographs.

Platz et al undertook calibration of a range of WSUD assets including bioretention for SWMM (Platz et al., 2020a). Two assets were considered, the Graham Bioretention Cell and the Villanova BioInfiltration Traffic Island with the first modelled using the bioretention LID and the second using the raingarden LID. A small number of selected events were calibrated. SWMM was found to satisfactorily represent the outflow hydrograph with NSEs in the range 0.55 to .97 with an average of 0.76.

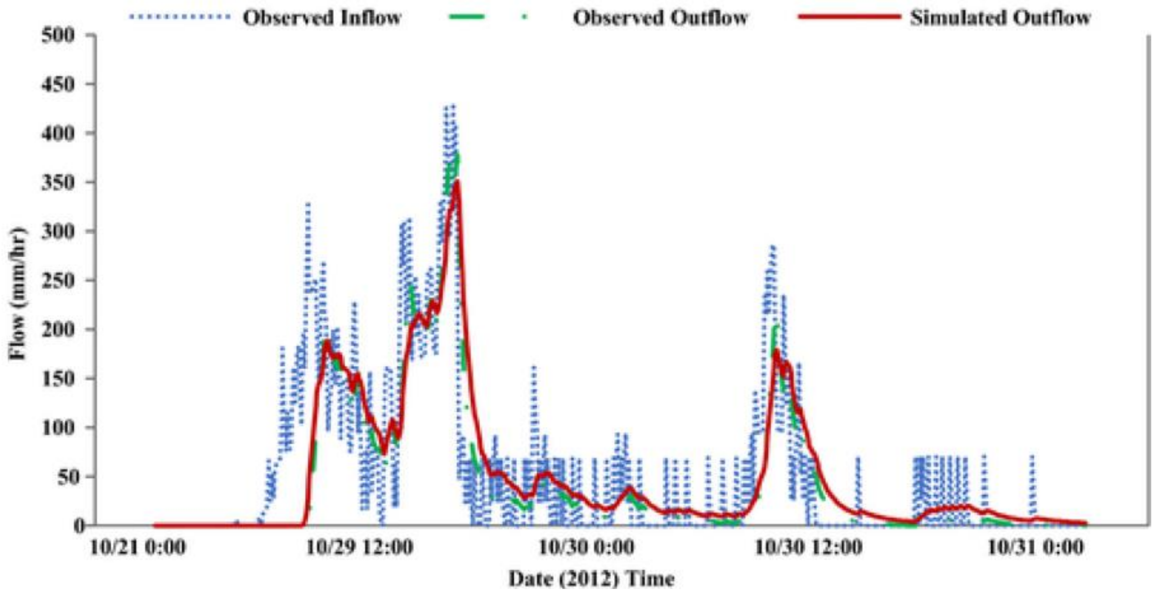


Figure 6 A representative hydrograph for the Villanova BioInfiltration Traffic Island indicating a good fit

4.3.4 Summary

Various authors have pursued calibration of MUSIC and other bioretention models. A summary of the assets, calibration configurations and outcomes are shown in Table 3. The outcomes were mixed, suggesting some good success with calibration but also some poorer results. As observed by Fowdar (Fowdar et al., 2022)and Lintern (2012), the model seems to perform best for a relatively standard biofilter configuration with typical sizing, soils and configuration and some of the results obtained and fits to hydrographs are quite good. However, the model has greater difficulty in representing a more diverse range of asset configurations. This can potentially be attributed to the simplified representation of the asset as having uniform homogenous soil conditions, uncertainties in inputs and other factors. This means that the model can be used with reasonably confidence for assets with typical or guideline sizing, soils and configuration but lesser confidence where these characteristics diverge from these.

Some learnings for calibration are as follows:

- The model should be either calibrated to or compared with the objective of interest (in our case mean annual outflow volumes)
- Both manual and automated (e.g. PEST) calibration approaches seem to offer adequate results
- Hydraulic conductivity appears to be the most commonly calibrated and sensitive parameter but other parameters such as infiltration rate, PET factor and those for dimensions such as surface area may also need to be tested and calibrated to obtain the best fit

Table 3 Summary of model calibration outcomes

Author	Asset/Cell	Year of data	Model	Number of events	Calibration	Parameters calibrated	Statistics used	NSE / φ (Outflows)	Measured % volume reduction	Modelled % volume reduction	Measured k (mm/hour)	Calibrated k (mm/hour)
Lintern	Monash car park biofilter	2007	MUSIC		Manual	k	NSE, unbiased function	0.62 / 2.16	-	-	123	230 / 100
								0.84 / 0.33	-	-	144	200 / 170
								0.65 / 0.37	-	-	77	60 / 60
Randelovic	Monash car park biofilter Cell 1	2011-2012	mPIRE		Manual		NSE (Outflows and soil moisture)	0.62-0.88	-	-	123 144	105 158
Randelovic	Monash car park biofilter Cell 2	2011-2012	mPIRE		Manual		NSE (Outflows and soil moisture)	0.88-0.90	-	-		105 158
Imteaz	Habitat Park	~2014	MUSIC	3	None		Outflow rate	NA	86%	90%	NA	100 adopted
	Yarraleen	~2014	MUSIC	3	None		Outflow rate	NA	89%	91%	NA	100 adopted
Fowdar	Punggol biofilter	Jan-Sep 2018 (Calibration) Oct 2018-	MUSIC	7	PEST	Surface area	NSE	-3.13 to 0.86	-	-		

Author	Asset/Cell	Year of data	Model	Number of events	Calibration	Parameters calibrated	Statistics used	NSE / φ (Outflows)	Measured % volume reduction	Modelled % volume reduction	Measured k (mm/hour)	Calibrated k (mm/hour)
		Mar 2019 (Validation)				PET scaling factor						
Lisenbee	Ursuline College		SWMM			K_{sat}, soil suction, conductivity slope, void ratio, seepage rate	NSE, event volume NSE					
Platz	Graham bioretention cell		SWMM		PEST		NSE, RSR, PB	0.67				
Platz	Villanova BioInfiltration Traffic Island		SWMM		PEST		NSE, RSR, PB	0.86				

5 Statistical analysis and model

5.1 Introduction

A statistical analysis was completed on a combination and mix of all three data sets mentioned in Section 3. The outputs from the models are the Nash-Sutcliffe Efficiency (NSE) and total percentage reduction in stormwater flow volume over the period of record or event.

The impact of stormwater biofilter design and operational variables on nutrient removal – a statistical modelling approach (Zhang et al, 2021) was used as the foundation for the statistical analysis used in this study and an overview of their method is provided here for reference, see the paper for further details. In the original study, multivariate statistical models were used to assess the impact of raingardens on nutrient removal. Three different multivariate models were used and are shown below in equations 1, 2 and 3. The models featured a combination of design variables which defined the system characteristics (plant species, filter media type, filter media depth, submerged zone depth) and operational variables which more define the rainfall events (antecedent dry weather period, inflow flux, inflow concentration, average infiltration rate).

In the Zhang study the following models were tested:

Model 1 (Variability around the mean model) assumes that outflow concentration is modelled with a mean outflow concentration determined by specific system design, and a noise caused by real-time operational conditions at a point of time.

Model 1: Variability around the mean model

$$Outflow_{i,j} = Outflow_j + \Delta outflow_{i,j}$$

$$Outflow_j = int + a * FM_j + b * FMD_j + \mathrm{c} * Veg_j + d * SZD_j$$

$$\Delta outflow_{i,j} = a_j * ADWP_{i,j} + b_j * IR_{i,j} + c_j * Flux_{i,j} + d * C\_in_{i,j}$$

Model 2 (Operational conditions model) is a multi-level model which assumes outflow concentration at time i for design j is a linear function of operational variables with the regression coefficients determined by the design variables.

Model 2: Operation conditions model

$$C\_Out_{i,j} = int + a_j * ADWP_{i,j} + b_j * IR_{i,j} + c_j * Flux_{i,j} + d_j * C\_in_{i,j}$$

$$a_j = int_a + x_a * FM_j + y_a * Veg_j + z_a * FMD_j + w_a * SZD_j$$

$$b_j = int_b + x_b * FM_j + y_b * Veg_j + z_b * FMD_j + w_b * SZD_j$$

$$c_j = int_c + x_c * FM_j + y_c * Veg_j + z_c * FMD_j + w_c * SZD_j$$

$$d_j = int_d + x_d * FM_j + y_d * Veg_j + z_d * FMD_j + w_d * SZD_j$$

Model 3 (Design characteristics model) is another multi-level model with the same structure as Model 2, but with it being a linear function of design variables with regression coefficients determined by the operational variables.

Model 3: Design characteristics model

$$C\_Out_{i,j} = int + a_i * FM_{i,j} + b_i * Veg_{i,j} + c_i * Flux_{i,j} + c_i * C\_in_{i,j}$$

$$a_i = int_a + x_a * ADWP_i + y_a * IR_i + z_a * Flux_i + w_a * C\_in_i$$

$$b_i = int_b + x_b * ADWP_i + y_b * IR_i + z_b * Flux_i + w_b * C\_in_i$$

$$c_i = int_c + x_c * ADWP_i + y_c * IR_i + z_c * Flux_i + w_c * C\_in_i$$

$$d_i = int_d + x_d * ADWP_i + y_d * IR_i + z_d * Flux_i + w_d * C\_in_i$$

For each model, all possible combinations of design factors and operational conditions were modelled to investigate how each variable impacted the treatment performance and how each combination predicted outflow concentrations of TP and TN for raingardens.

5-fold calibration and validation approach was used to fit and rigorously test the 1,350 models. This consisted of randomly selecting 80% of data from full datasets for model calibration and using the remaining 20% for validation. The process was repeated 5 times to ensure robustness of testing. The Nash-Sutcliffe Efficiency (NSE) for C_out was used as a performance indicator for the calibration and validation.

Optimisation functions within the SciPy package were used in Python 3.7 to calculate the NSE value for each model. 'BFGS' optimisation method was used to calculate NSE values and initial values for regression coefficients were set to zero.

5.2 Methodology

The multi-variate models from the Zhang 2021 were adopted as a basis and then adapted to suit the flow retention objectives. Whereas the previous statistical models were based around an outflow concentration and determining which combination of operational and design variables influenced it, the models were now based around the outflow volume.

We hypothesised that it is likely that targeted analysis of a small number of well known assets may yield greater insights and understanding than attempting to work with the large and complex database available from the literature in entirety given the findings by others (Poresky et al., 2012) that found it difficult to identify clear relationships and results from a large and quite heterogenous data set. For this reason and to ensure that both design and operational variables could be assessed using data at an event level, we focussed efforts on assessment of three of the field studies (Monash car park, Wicks Reserve and Hereford Road).

The functions adopted for the study are presented below.

Model 1: Variability around the mean model

$$C\_Out_{i,j} = Outflow_j + \Delta outflow_{i,j}$$

$$Outflow_j = int + a * EDD_j + b * FMD_j + \mathrm{c} * SZD_j$$

$$\Delta outflow_{i,j} = a_j * ADWP_{i,j} + b_j * IR_{i,j} + c_j * Flux_{i,j}$$

Model 2: Operation conditions model

$$C\_Out_{i,j} = int + a_j * ADWP_{i,j} + b_j * IR_{i,j} + c_j * Flux_{i,j}$$

$$a_j = int_a + x_a * FMD_j + y_a * EDD_j + z_a * SZD_j$$

$$b_j = int_b + x_b * FMD_j + y_b * EDD_j + z_b * SZD_j$$

$$c_j = int_c + x_c * FMD_j + y_c * EDD_j + z_c * SZD_j$$

$$d_j = int_d + x_d * FMD_j + y_d * EDD_j + z_d * SZD_j$$

Model 3: Design characteristics model

$$C\_Out_{i,j} = int + a_i * FMD_{i,j} + b_i * EDD_{i,j} + c_i * SZD_{i,j}$$

$$a_i = int_a + x_a * ADWP_i + y_a * IR_i + z_a * Flux_i$$

$$b_i = int_b + x_b * ADWP_i + y_b * IR_i + z_b * Flux_i$$

$$c_i = int_c + x_c * ADWP_i + y_c * IR_i + z_c * Flux_i$$

$$d_i = int_d + x_d * ADWP_i + y_d * IR_i + z_d * Flux_i$$

Input data for the models was provided for three sites where research studies were undertaken and have been documented in the Section 3.2: Hereford Road, Wicks Reserve, and Monash Biofilter. Refer to Section 4.1 for further information on the field measurements used for models.

The model input data consisted of discrete rainfall events and continuous datasets recorded at the asset sites. The statistical models required discrete rainfall events, so the continuous data sets were split and organised into rainfall events. This raised a number of challenges around the consistency of collation of the various datasets and how they differed. It is unclear how the Wicks Reserve event data was collated and split into discrete events. Hence, the splitting of the other continuous data sets into rainfall events will differ as there are a few questions to address. These include:

- How to define start and end of rainfall events?
- When to split one rainfall event into two or more?
- How do our assumptions compare to the provided discrete datasets?

The statistical models above were tested for different configurations of design and operational variables. This included adopting three or four variable models, and different combinations of variables. The variables tested were:

- Design: FMD, EDD, SZD, media depth, area ratio, catchment area, asset area
- Operational: ADWP, IR, flux, evapotranspiration

Different combinations were iterated to see which resulted in the NSE values with the lowest standard deviation. As per the research paper, the input data was randomly split into 80% training data and 20% test data. This process was repeated five times and NSE value used as the performance indicator.

Table 4 Results summary of the five-fold training and test for different models

Model	No	NSE (Train)		NSE (Test)		Design variables	Operation variables
Model 1 (Variability around the mean model)	Bo1	0.519	±0.028	-3.190	±1.199	FMD	ADWP
	Bo2	0.692	±0.044	0.530	±0.223	FMD, EDD	ADWP, IR
	Bo3	0.944	±0.015	-1.418	±1.656	FMD, EDD, SZD	ADWP, IR, flux
Model 2 (Operational conditions model)	Bo1	-2.429	±1.869	0.370	±0.039	FMD	ADWP
	Bo2	0.498	±0.344	0.798	±0.031	FMD, EDD	ADWP, IR
	Bo3	-0.317	±0.955	0.980	±0.009	FMD, EDD, SZD	ADWP, IR, flux
Model 3 (Design characteristic model)	Bo1	0.004	±0.003	-0.017	±0.020	FMD	ADWP
	Bo2	0.648	±0.032	0.593	±0.148	FMD, EDD	ADWP, IR
	Bo3	0.942	±0.017	0.900	±0.053	FMD, EDD, SZD	ADWP, IR, flux

The outputs of the statistical models resulted in varying NSE values with high variance. The training model typically resulted in high NSE values whereas the test models resulted in low to medium NSE values. It was also not uncommon for models to result in negative NSE values. Similarly, the training model had lower standard deviation values whereas the test model had standard deviation multiple orders of magnitude higher. This indicates that while the initial calibration may have been reasonably successful, that it was not readily transferrable to other events or data. Some of the reasons for this may be the level of noise in the data

There is potential that with further investigations and refinement, that more promising results can be obtained. However, given the limited success and recognition that other studies seeking statistical correlations have also recognised challenges (Poresky et al., 2012), it was decided that it may be preferable to focus greater efforts on the MUSIC calibrations.

It is hypothethised that varying methods, sources and the structurally different configurations may have contributed to a somewhat heterogenous data set with a high ratio of noise to signal that can make it difficult to draw out relationships with confidence. Such conclusions are not dis-similar to the prior efforts such as (Poresky et al., 2012) using a larger data set. The latter found a few tenous relationships but that it was difficult to clearly identify or articulate the key driving factors for performance based on the data. While a range of likely influential factors were chosen for analysis, it is also possible that other factors not considered may influence results, for example differences in vegetation species, form and maturity.

This is not to say this approach cannot work, but that either a dataset with a greater level of consistency may be needed (such as a series of assets with mostly common features, context or monitoring programs) or deeper analysis and more powerful statistical approaches could possibly be used to gain some insight.

Some key learnings that we can draw from this work are:

- The level of noise in the existing literature data makes it difficult to draw firm conclusions on driving factors for performance. This is likely to be exacerbated rather than improved by using a larger data set
- A sub-set of higher quality and consistent data may be most useful to gain further insights
- Targetted pair-wise studies may be a preferable way of identifying the effects of various factors on performance (e.g. assets with one change in design configuration or identical assets in different locations)
- The diversity of field assets make it difficult to assess statistically. Model calibration approaches may possibly be a preferable approach although ideally models should have structural flexibility to accommodate a range of potential configurations.

6 Methods for calibration to MUSIC

6.1 Calibration approach

Two calibration approaches were used, manual calibration and calibration using an optimisation tool, PEST. It is noted that manual calibration was undertaken within the scope of this project while the automated calibration was undertaken wholly as a research extension of the project to support a complementary journal paper.

The manual calibration is valuable for gaining an understanding of parameter sensitivity and can be an effective way of achieving a good calibration in its own right. It can also be useful for identifying a suitable or good starting point for the automated calibration which may then have a better chance of finding an optimum. The automated calibration may then only need to make minor further refinements or alternately may find a different optimum point. The outcome can be further validated by commencing automated calibration further from the manual optimum to see what outcomes are obtained and whether these validate the manual outcomes and indicate the automated process is working (that is it can find a solution that is similar or better than the manual calibrations).

6.2 Objectives and metrics

Measured and model predicted outflow rates and mean annual runoff volumes of the period of analysis were used for the calibration.

The performance measures used in this study were the Nash–Sutcliffe efficiency (NSE) used for optimisation to obtain a best fit to the overall outflow hydrograph and the mean annual runoff volume (MARV) reduction (%) which is the main objective of interest. The percent bias (PBIAS) was also calculated for consideration.

Nash-Sutcliffe Efficiency (NSE)

The Nash-Sutcliffe Efficiency (NSE) is a measure of how well the modelled time series matches the observed time series. "The NSE (Equation 1) is a normalized statistic that expresses the relative magnitude of the residual variance (noise) compared to the variance of the measured data (information)" (Nash & Sutcliffe, 1970). NSE values range between $-\infty$ and 1.0, where NSE = 1 is the optimal value. It generally favours good fitting of the peaks (Beven J., 2001).

$$1-\frac{\sum_{i=1}^{n}\left(O_i-P_i\right)^2}{\sum_{i=1}^{n}\left(O_i-\bar{O}\right)^2}$$

Equation 1 Nash Sutcliffe efficiency

The NSE was used in this study as the objective function for undertaking calibration of the MUSIC model. The NSE was calculated using the finest time-scale available of 6 minute across the full time series. This presents a demanding test as it requires the model to both represent the overall volumes as well as the timing and magnitude of flows throughout events.

MARV reduction (%)

The focus for MARV objectives such as those set in the states of Victoria and NSW are to reduce direct stormwater surface runoff flows entering waterways and adversely changing the hydrology through increased frequency and duration of event flows. Reducing overall volumes tends to reduce both the frequency and duration of these event flows and restore more natural hydrologic patterns. While other approaches such as peak flow attenuation has possible value for flood mitigation, it is generally recognised that reducing the increase in flow volume due to urbanisation is essential for restoring more natural flow patterns. The outflow volume and mean annual runoff volume reduction is important as the model must provide a reasonable prediction of the outflows and corresponding volume retention.

The mean annual runoff volume (MARV) reduction is calculated as the average difference between the inflows to the asset and the outflows from the asset over a year. For a biofilter, 'losses' that are realised as downstream flows may be through infiltration into the surrounding soils or evapotranspiration.

MARV reduction = (Inflow – Outflow) / Inflow

Equation 2 MARV reduction

Outflows should ideally consist of total outflows including both outflows and overflows. However, in practice, bypassed flows and overflows are either not monitored at all or not accurately measured during field monitored.

Note that we have reported this as percentage reduction in runoff volume relative to stormwater inflow volume over the period of record for each dataset since the periods of record vary and this allows easier comparison. This shall be referred to as % MARV in the following for consistency with the adopted terminology in guidelines.

Further consideration may then be given to the breakdown of modelled fluxes and how this compares with the observed and calculated fluxes.

Percent bias (PBIAS)

"The PBIAS measures the average tendency of the simulated data to be larger or smaller than their observed counterparts" (Gupta et al., 1999). "The optimal value of PBIAS is 0.0, with low-magnitude values indicating accurate model simulation. Positive values indicate model underestimation bias, whereas negative values indicate overestimation bias" (Gupta et al., 1999). PBIAS is calculated with Equation 3 where PBIAS is the deviation of the evaluated data, expressed as a percentage. The PBIAS provides an indication of whether modelled outflow data accurately represents peaks and troughs in observed outflow data and can help indicate poor model performance (Gupta et al., 1999).

$$\left[\frac{\sum_{i=1}^{n}(O_i-P_i)}{\sum_{i=1}^{n}O_i} * 100\right]$$

Equation 3 PBIAS

6.2.1 Manual calibration

Initially, calibration was undertaken manually to understand model behaviour and general sensitivity to different parameters. A consideration for pursuing manual calibration was that automated parameter optimisation is neither directly accessible via the MUSIC interface nor simple to carry out, and also that users generally have to change parameters manually in MUSIC to achieve a desired result. This therefore better reflects what is likely to be possible for a broader suite of users that may choose to pursue calibration for other assets.

Parameters describing the dimensions of the asset were generally adopted as reported although in some cases these were also varied.

A distinction that must be made clear for the PET Factor is that MUSIC takes the areal potential evapotranspiration (APET), otherwise referred to as Morton's Wet Environment evapotranspiration as its basis evapotranspiration.

The filter media saturated hydraulic conductivity, underlying soil exfiltration rate and the 'PET Factor - which can be used to adjust the level of evapotranspiration and works similar to a crop factor, are anticipated to be key variables affecting hydrologic biofilter performance. Unlike fixed design parameters, these may vary during asset operation and are difficult to accurately measure, making measurements highly uncertain. For this reason, it was assumed that these factors should be used as calibration parameters since they are likely to be key factors in influencing the goodness of fit between observed and modelled hydrographs and MARV reductions.

This is a reasonable representation of evapotranspiration occurring across a catchment or treatment area surrounded by a broad area where similar levels of evapotranspiration are occurring and thus suppressing potential evapotranspiration. It is widely used in catchment models. However, in MUSIC it is also used in the treatment modules.

If used ‘as is’ without any factoring (that is a PET factor of 1.0), there is a risk the areal potential evapotranspiration may underestimate evapotranspiration from small assets such as raingardens which are not surrounded by similar or irrigated landscapes (for which the point potential evapotranspiration may have been a better starting point). The default PET factor in MUSIC is set to 2.1 which essentially factors up the areal potential evapotranspiration to a more realistic level based on a calibration to biofilter columns at Monash. Such an approximation is sufficient for typical biofilters where PET is usually not more than 5% of the overall water balance. However, a greater level of consideration and accuracy may be needed for larger raingardens or roof gardens, should the ‘biofilter’ node in MUSIC be used to represent these.

It is further noted that the point potential evapotranspiration may be in the order of double (or 2.1) times the areal potential evapotranspiration but is unlikely to be much more than this. The above therefore *should not be misinterpreted* to suggest that PET factors of greater than 2.1 are either feasible or likely for a typical biofilter. Furthermore, these factors represent the ‘maximum potential’ evapotranspiration and the actual realised evapotranspiration will always be a proportion of this depending on soil moisture and plant stress.

Some design parameters such as the depth of the overlying surface pond (Extended detention depth) were also varied, recognising that the constructed outcome may not match the design intent or that this may provide a better calibration.

Parameters calibrated manually included:

- Filter media saturated hydraulic conductivity, K_{sat} (mm/hour)
- Exfiltration rate into underlying soil (mm/hour)
- PET Factor (-)
- Extended detention depth (EDD) (m)

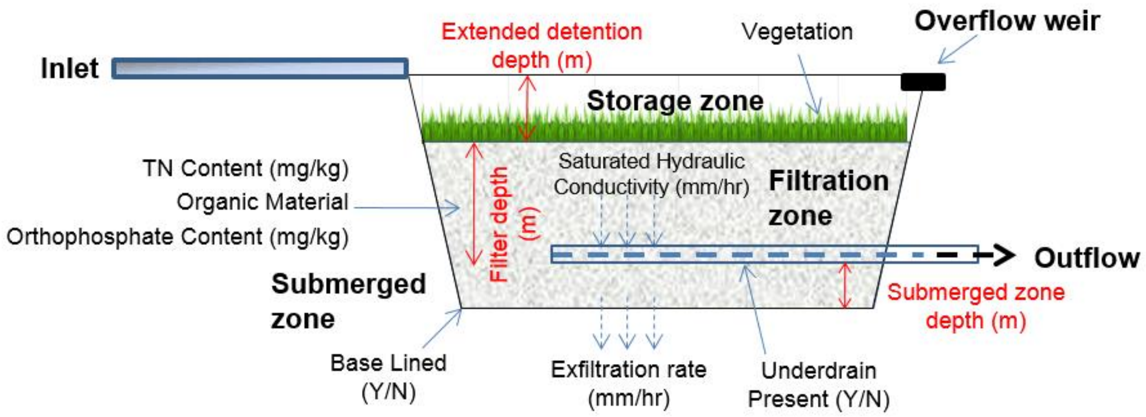


Figure 7 Schematic representation of biofilter in MUSIC indicating parameters of interest (eWater, 2024)

Models were calibrated as best as possible based on optimising the operational variables including filter media saturated hydraulic conductivity, exfiltration rate into underlying soil and PET factor.

ideally optimisation would be achieved just by modifying these uncertain variables. However, extended detention depth was also added as a calibration parameter, recognising the constructed outcome may not match the design intent or may be difficult to measure accurately (notably for Wicks Reserve which had an undulating surface). Furthermore, MUSIC does not account for plant volumes and even small changes in storage volumes may significantly affect hydrologic performance. Furthermore, for assets such as Wicks Reserve and Hereford Road, the saturated hydraulic conductivity and extended detention depth should be regarded more as 'black box' calibration parameters rather than real physical numbers due to the difficulties in representing the physical configuration of these assets in MUSIC (with constrained and greatly elevated outlets respectively).

The remaining parameters were set based on design information, defaults or best judgement.

Parameter ranges for the optimisation parameters were set to cover a physically realistic range initially. This was then broadened where outcomes indicated this may be necessary for optimal calibration.

Several versions of the model with varying parameter sets were configured, run and outputs extracted. This was undertaken on one or several models at a time with the parameter sets to be tested adaptively chosen depending on outcomes.

Relatively good outcomes were achieved using this approach with promising NSE's and matches between the observed and modelling reductions in stormwater flow volumes. In some cases, good results for both were achieved. However, for Wicks Reserve in particular, it was clear that NSE and mean annual volume outcomes were being traded off against each other for parameter sets within a physically realistic solution range.

For Wicks Reserve, it was apparent that the outflows were constrained by the outlet while this was not represented at all by the model since it assumes a free underdrain (in contrast, SWMM can model a constrained underdrain but historically did not allow representation of a submerged zone (Lisenbee et al., 2021), although this is now possible). It appears that representing a constrained outlet may be important for simulating the Wicks Reserve biofilter.

To compensate for the lack of this functionality in MUSIC, *the saturated hydraulic conductivity of the filter media was revised substantially downwards* to approximate the outlet drain constraint and the extended detention depth increased to approximate storage within the filter media. This was successful in significantly improving the numerical calibration performance, albeit at the cost of *using a model representation that did not very closely reflect the real conditions*.

6.2.2 PEST autocalibration

To supplement the initial efforts, we used the PEST software (Doherty, 2015, 2024) which can be found at (https://pesthomepage.org/ to calibrate the bioretention module in MUSIC. PEST is a suite of tools for model independent parameter estimation and uncertainty analysis. It is intended to automate calibration of numerical models based on sampling and simulation. PEST solves the non-linear least squares problem using a Levenberg-Marquardt algorithm to minimise differences between model outputs and observed data. PEST has previously been used with MUSIC (Dotto et al., 2009, 2011; Fowdar et al., 2022) (Dotto et al., 2009, 2011; Fowdar et al., 2022), SWMM (Platz et al., 2020a) as well as for calibrating models for other individual WSUD assets like infiltration trenches (Browne et al., 2013; Browne & Deletic, 2011).

Autocalibration tools like PEST can be used to calibrate MUSIC (run in command line mode) based on optimising its performance for a given objective function (e.g. NSE) for a defined set of parameters using a global optimisation. PEST was trained to read MUSIC's input and output files through configuration of template and instruction files respectively. The relevant parameters to be varied through calibration were the same as above:

PEST was configured to factor the saturated hydraulic conductivity and exfiltration rate in the log domain given that these can potentially widely across one or more orders of magnitude.

The observations were read in and weighted with a weighting of 1 for periods of active observations with flows greater than zero. A weighting of 0 was used for periods outside these as well as for periods of modelled inflows for Wicks Reserve. The modelling of inflows was essential to ensure appropriate flows and soil moisture levels in the biofilter but it was considered safer to exclude them from calibration to mitigate the risks of influence by any possible systematic effects from any errors in the inflow estimation.

A PEST control file was configured to manage the calibration process. It was configured to run in the 'vanilla' parameter estimation mode with mostly default settings. PEST solves the inverse problem using the Gauss-Marquardt-Levenberg method. It first calculates the Jacobian matrix then uses this to calculate an improved set of parameters. The method can accommodate nonlinear behaviour of the model's outputs with respect to its parameters which is essential for MUSIC.

PEST was set to optimise the Nash Sutcliffe Efficiency (NSE) calculated for modelled outflows vs observed outflows at a 6 minute timestep for the whole time series.

It is noted that observations were available for some sites at a finer time-scale but this is the finest timestep that MUSIC can report. The use of the fine 6 minute time scale resulted in a relatively large number of observations for calibration (~250,000 for Wicks Reserve).

The input parameter ranges and starting values for the automated calibrations are summarised below in Table 5.

Table 5 Input parameters for automated calibration

Parameter	Range	Monash	Wicks Reserve	Hereford Road
Filter media saturated hydraulic conductivity (mm/hour)	Lower	10	1	10
	Start	190.08	10	100
	Upper	300	400	300
Underlying soil exfiltration rate (mm/hour)	Lower	0.01	0.01	0.01
	Start	0.03	0.2	0.4
	Upper	3.6	100	10
PET Factor	Lower	-	1.0	1.0
	Start	1.0	1.0	2.1
	Upper	-	3.0	2.1
EDD	Fixed	0.25	0.75	0.3

7 Calibration results

The calibration results for each of the assets are summarised in the following sections followed by a discussion of the overall outcomes.

7.1.1 Monash car park biofilter

Calibration was undertaken comparing modelled drain outflows and observed outflows. Initial comparisons between measured outflows and total modelled outflows yielded poor results due to overflows occurring. The results comparing outflows (for events without overflows which were not monitored) are presented below in Table 6. Note that the % MARV reduction here are not indicative of the actual reductions realised by the asset due to these reflecting the difference between inflows and drain outflows. The calculated losses for the final calibrated model are shown in Table 7.

It can be seen that using a conservative model guideline value for saturated hydraulic conductivity (100 mm/hour), zero exfiltration and default or design values for PET factor and extended detention depth respectively resulted in a moderate NSE at best of 0.57.

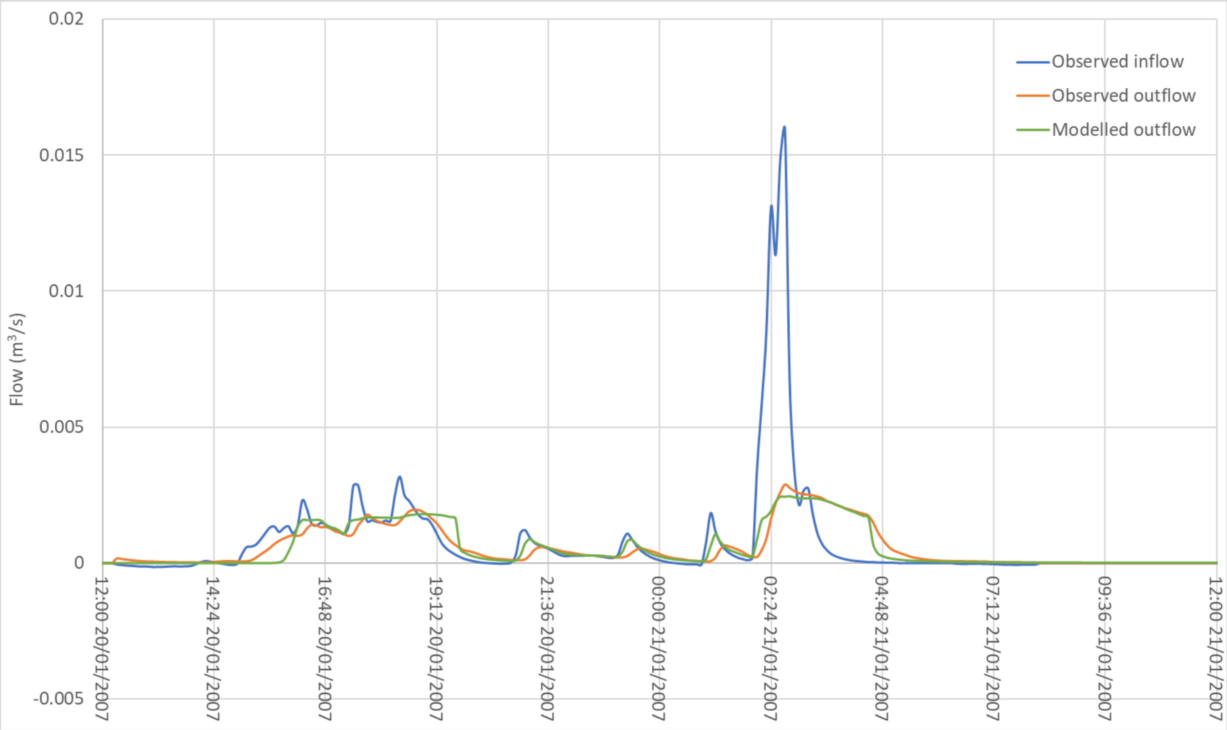
This was improved significantly improved through the use of a more representative saturated hydraulic conductivity with the best results for manual calibration being an NSE of 0.77 and MARV difference of +0.1%.

Calibration using the PEST optimisation software was undertaken to see whether the outcomes could be improved. This further improved the outcomes with PEST achieving an NSE of 0.84 while still maintaining a very good percentage difference in mean annual runoff volume of -0.3%.

The results from this calibration are by far the best and this is perhaps a reflection that the asset design is fairly typical and consistent with the configuration expected in MUSIC.

Table 6 Selected scenario results for the Monash Carpark Biofilter continuous data

Scenario	**Parameter Value**				**Model performance**		
	EDD (m)	**K_{sat} (mm/hr)**	**Exfiltration (mm/hr)**	**PET factor**	**% MARV reduction (Difference from observed)**	**NSE**	**PBIAS**
Observed	-	-	-	-	24.1%	-	-
Base Case	0.25	100	0	2.1	38.9% (+14.7%)	0.57	19.4%
1	0.25	150	0	2.1	31.2% (+7.0%)	0.76	9.3%
2	0.25	150	20	2.1	52.1% (+27.9%)	0.71	36.8%
3	0.25	300	0	2.1	19.1% (-5.1%)	0.71	-6.7%
4	0.35	150	0	2.1	24.3% (+0.1%)	0.77	0.15%
PEST	0.25	190.08	0.03	1	23.8% (-0.3%)	0.84	-0.40%



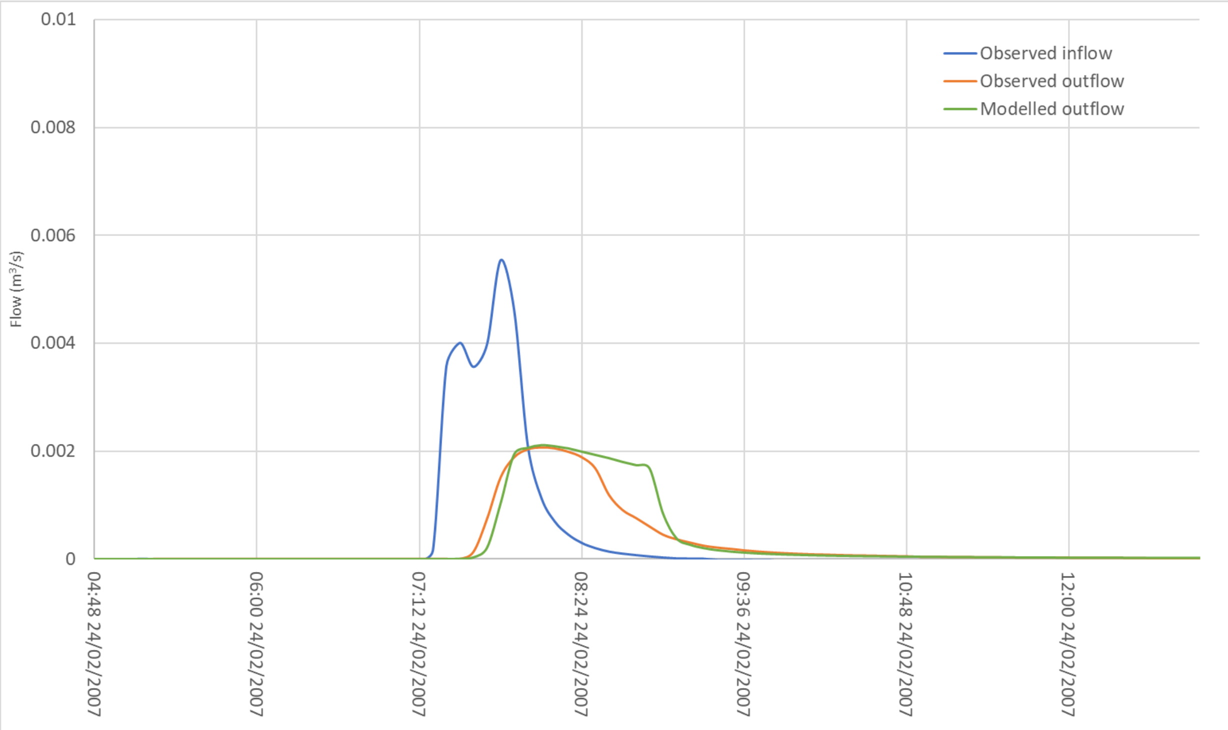


Figure 8 Example events from calibration

The differences between inflows and drain outflows appears substantial given that the asset was intended to be lined. However, we also need to factor in flows that bypassed or were overflows to more accurately calculate the 'losses' realised through infiltration and evapotranspiration.

To get an estimate of actual losses, the model results from the final model calibrated using PEST over the full period including events for which overflows occurred (that were excluded from the calibration) were examined, see Table 7. This indicates that the observed losses over the period of record were in the order of 32 kL/year or 3.4% of the total inflows. This is much less than the 23.8% modelled when comparing inflows to drain outflows (as in the calculations above for the calibration).

It was recognised during post-monitoring assessment of the asset (Tanner et al., 2020a) that the asset had been leaking due to cracking of the concrete and root intrusion. The optimisation outcomes reflect this, indicating a low infiltration rate can be used to approximate the leaking flows to good effect. Based on the breakdown shown in Table 7, they account for a modest proportion of the overall volume reductions.

Table 7 Summary of detailed flows for PEST calibrated MUSIC model

Scenario	Result (kL/year)
Inflow	946.1
Outflow	742.4
Weir flow	184.6
Infiltration	3.5
Evapotranspiration	28.4
Total losses (infiltration and evapotranspiration)	31.9
Actual losses (percentage)	**3.4%**

7.1.2 Wicks Reserve

The Wicks Reserve biofilter was calibrated for total outflows for the period from 25/09/2013 to 20/08/2016. Only event periods for which monitored outflow data were available (96 events) were calibrated, which were within the timeframe 25/09/2013 to 14/08/2016.

A selection of the parameters tested and resulting modelling performance are summarised in Table 8.

A baseline was established using the design extended detention depth of 0.45, the measured hydraulic conductivity of 230 mm/hour and exfiltration of 1 mm/hour. This produced fair performance with an NSE of .046 and % MARV reduction of 31.5%, indicating a significant over-estimate of flow volume reductions.

Visual inspection of the model results suggests reasonable representation of the event flows and timing of peaks, though there is a tendency for the model to represent a shorter and sharper peak followed by a rapid recession in the model.

It became clear through visual inspection of the baseline model hydrographs, see Figure 9 and Figure 11 that there were persistent elevated outflows occurring from the biofilter following events in the monitored data. However, MUSIC does not represent these flows and they cannot easily be replicated through calibration using parameters within realistic physical ranges as determined through the initial calibration efforts. For the modelled flows, the outflow rate rapidly decreases following an event. This may be expected if we consider the drainage equation in MUSIC which is as shown in Equation 4 (Lintern et al., 2012). The flow rate is related to the soil moisture saturation level and decreases exponentially as soil moisture within the biofilter reduces.

$$Q = Aks^{\beta}\frac{h + D}{D}$$

Equation 4 Drainage from biofilter filter media in MUSIC

Where:

Q = outflow rate (m^3/s)

A = filter area (m^2)

k = saturated hydraulic conductivity

S = soil moisture saturation (-)

β = Dingman coefficient (-)

h = depth of ponding (m)

D = depth of filter media (m)

The observed extended recession periods in the observed data may be attributed to the size of the outflow pipe, which for such a large biofilter becomes the limiting constraint for outflow rates.

There is no functionality within MUSIC to represent a limited outflow pipe such as this although other models such as SWMM have this capability (Lisenbee et al., 2021).

In lieu of this, calibrations were then tested to approximate the ‘constrained outlet’ behaviour by revising parameters and relaxing the constraints on parameters from the notional physical limitations. In effect, the intention is to approximate the constrained outlet by limiting the saturated hydraulic conductivity of the filter media and increasing the extended detention depth to approximate storage within the filter media.

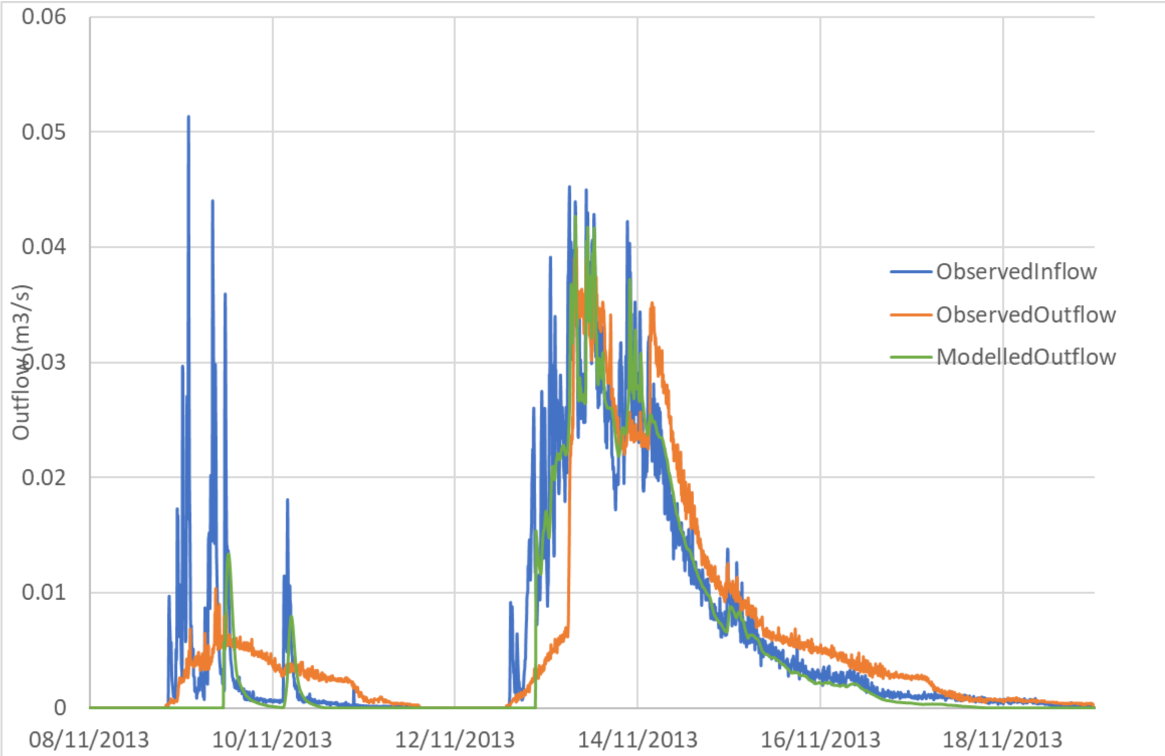


Figure 9 Typical hydrographs for Wicks Reserve – baseline (November 2013)

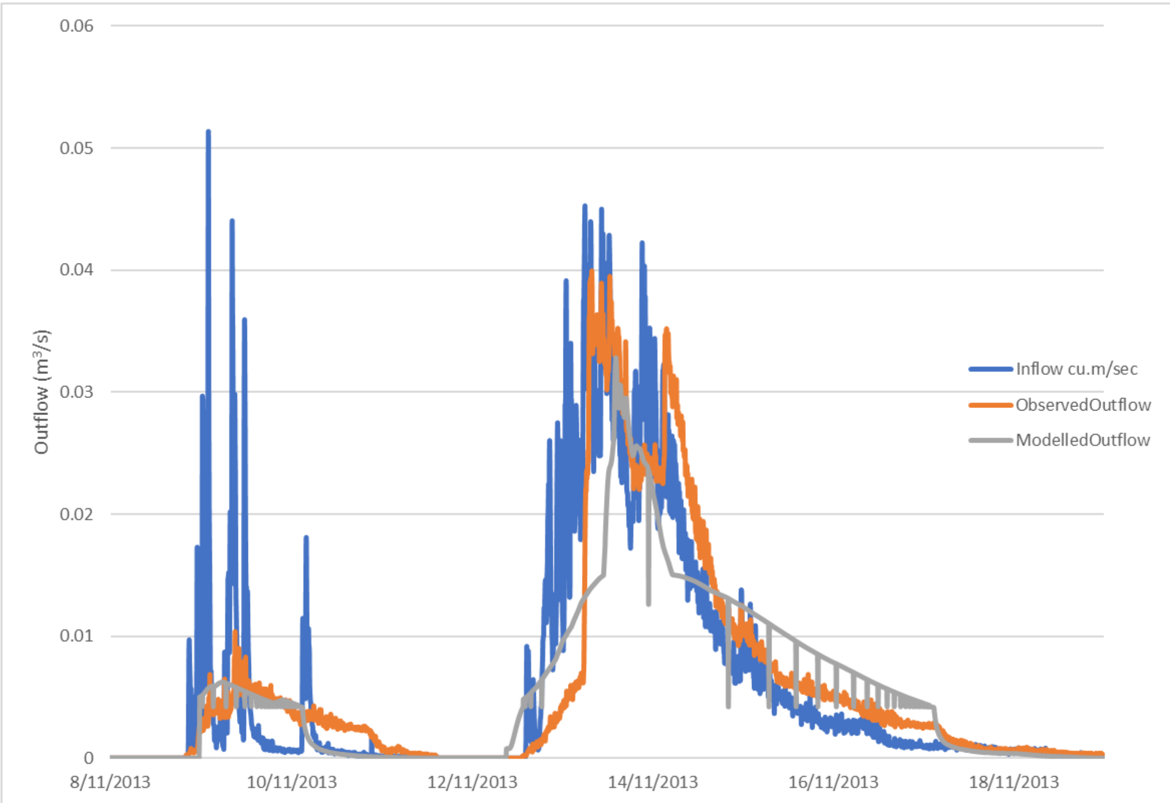


Figure 10 Typical hydrographs for Wicks Reserve following calibration with PEST (Nov 2013)

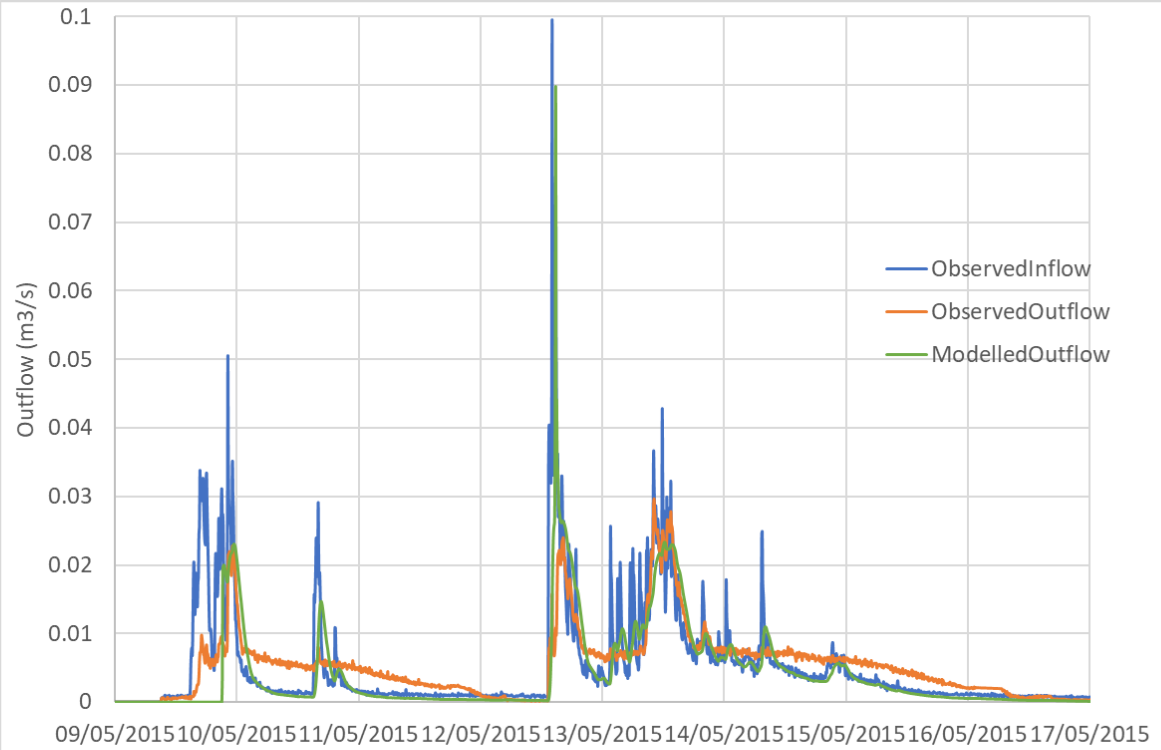


Figure 11 Typical hydrographs for Wicks Reserve – baseline (May 2015)

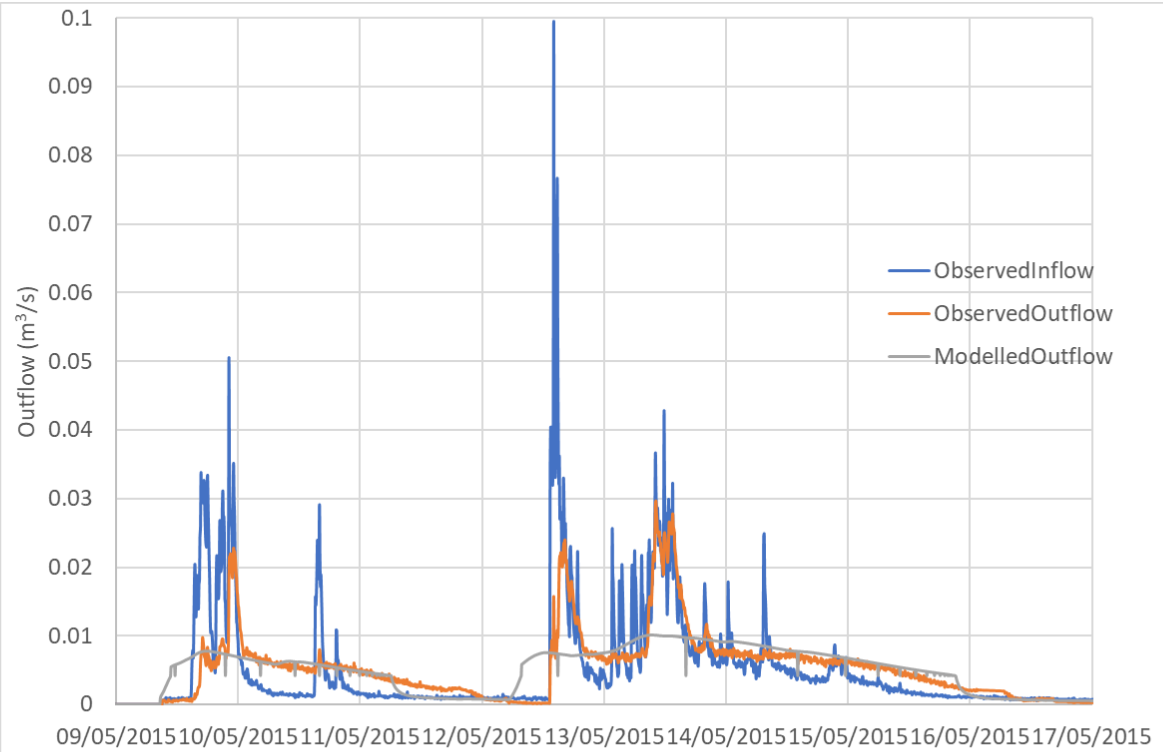


Figure 12 Typical hydrographs for Wicks Reserve following calibration with PEST (May 2015)

The calibration outcomes were improved significantly improved through the use of the greatly reduced saturated hydraulic conductivity and increased extended detention depth with the best results for manual calibration being a MARV difference of +19.8% and NSE of 0.67.

Calibration using the PEST optimisation software was undertaken to see whether the outcomes could be improved. This further improved the outcomes with PEST achieving a percentage difference in mean annual runoff volume of 1.5% and an NSE of 0.70. The PBIAS was, however, very high (71%) indicating potential accuracy issues with the model.

Overall, it is considered that it is possible to calibrate MUSIC to assets such as these with a constrained outlet. However, the use of highly modified parameters means that the transferability of parameters to make predictions for other scenarios or assets would likely be less than the outcomes for the Monash Biofilter with a more typical asset configuration.

Table 8 Selected scenario results for Wicks Reserve

Scenario	Parameter value				Model performance		
	EDD (m)	K_{sat} (mm/hr)	Exfiltration (mm/hr)	PET factor	% MARV reduction (Difference from observed)	NSE	PBIAS
Observed	-	-	-	-	12.2%	-	-
Base	0.45	230	2	2.1	47.1%	0.48	10.6
1	0.65	70	2	1	46.8% (+34.6%)	-0.10	9.2%
2	0.75	10	10	1	88.1% (+75.9%)	0.23	82.8%
3	0.75	10	2	2.1	49.1% (+36.9%)	0.63	13.3%
4	0.75	5	1	1	34.6% (+22.4%)	0.59	-16.5%
5	0.75	10	1	1	32.2% (+20%)	0.67	-17.9%
6	0.75	10	2	1	48.1% (+35.9%)	0.63	11.5%
7	0.85	10	1	1	32.2% (+20.0%)	0.67	-17.9%
8	0.75	20	1	1	32.0% (+19.8%)	0.49	-18.5%
9	0.75	10	0.2	1	14.1% (+1.9%)	0.67	-70.0%
PEST	0.75	8.644	0.193	1.024	13.7% (1.5%)	0.70	-70.7%

7.1.3 Hereford Road

A selection of the parameters tested and resulting modelling performance are summarised in Table 9.

MUSIC demonstrated relatively poor performance in predicting the percentage MARV reduction when a relatively high initial exfiltration rate was adopted of 10 mm/hour. This was based on approximately the average of the reported measured range of 2-17 mm/hour (Poelsma et al., 2013).

However, this was readily improved by adopting a low exfiltration rate that would be more consistent with typical guideline recommendations for a medium to heavy clay and at or below the lower end of the measured range. Reducing the exfiltration rate to values as low as 0.3 or 0.4 gave the best fit to the volume reductions while retaining a slightly better NSE than the base case.

The NSE compares the observed and modelled flows at the modelled timestep of 6 minutes and thus requires a more accurate representation of the actual hydrographs than simply matching the % MARV reduction. The base case model achieved a fair NSE value (NSE = 0.54), indicating the model may not have adequately represented dynamic hydrograph variations or peak flows.

PBIAS values varied with a high (inadequate) score for the base model run but were mostly very good or at least fair for the following calibrations.

The obvious challenge with the Hereford Road biofilter is that the outlet structure has an unusual configuration whereby water must rise to close to the surface level of the biofilter before flowing out. It was anticipated that representing this using 100 mm of filter media and 900 mm of submerged zone could result in unrealistic results. Therefore, as a compromise, this was represented initially using 300 mm of filter media and 600 mm of submerged zone. However, this was tested in later calibrations and it was found that a good fit could still be obtained with a filter media depth of 100 mm and a submerged zone of 900 mm.

After initial manual efforts, PEST was trialled to further calibrate the model. The first run tested varying saturated hydraulic conductivity and exfiltration to optimisation NSE. The outcome for this was a high saturated hydraulic conductivity of 300 mm/hour (the upper limit allowed) and an exfiltration rate of 1.67 mm/hour. However, this also resulted in a significant increase in mean annual flow volume reductions to 18.6% relative to 8.5% for the observed data.

A second PEST optimisation was conducted allowing the PET factor to also be varied. The outcome for this was a saturated hydraulic conductivity of 245 mm/hour and an exfiltration rate of 1.055 mm/hour while the PET factor only marginally reduced to 2.08. This improved the mean annual flow volume reduction from 18.6% to 14.8% (still nearly double observed) with just a small drop in NSE from 0.61 to 0.58. In effect, it is apparent that the % MARV reduction and NSE are being traded off, similar to what was observed for Wicks Reserve and also to observations by Platz (Platz et al., 2020b), calibrating SWMM to the Ursuline College biofilter.

Given the limited improvements using PEST, further manual calibrations were undertaken.

It was found that (substantially) reducing the saturated hydraulic conductivity could significantly improve the NSE outcomes while still preserving the target mean annual outflow volumes within a desirable range. This resulted in very low saturated hydraulic conductivity values being adopted, in the order of 10-15 mm/hour for the best results. It is unlikely that the loamy sand filter media would actually have a saturated hydraulic conductivity this low in a constructed biofilter. Therefore, as for Wicks, this suggests that the parameter is being used as a 'black-box' parameter to compensate for other structural deficiencies in the model and provide a good approximation of the observed data.

The adopted 'best' model was calibration 17, which adopted the structural configuration of the asset, had an NSE of 0.71 and percentage MARV reduction of 8.6% (c.f. 8.5% observed). This is quite a good calibration outcome and promising result, albeit with the caveat that the saturated hydraulic conductivity is likely not physically realistic.

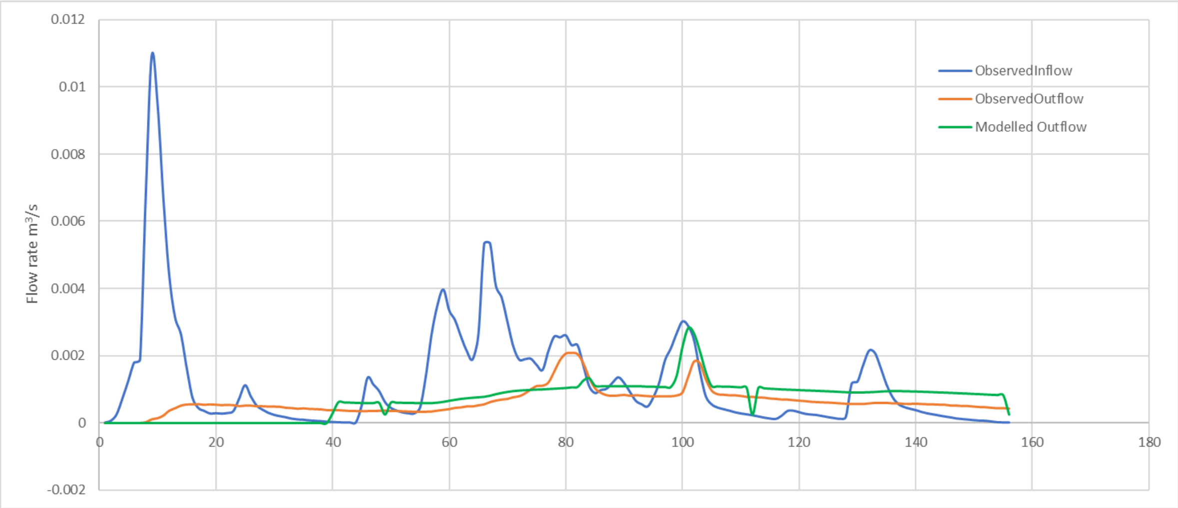


Figure 13 Typical hydrograph for Hereford Road following calibration with PEST (Jan 2012)

Table 9 Hereford Road calibration results

Scenario	Parameter Value					Model performance		
	EDD (m)	Filter depth (m)	K_{sat} (mm/hr)	Exfiltration (mm/hr)	PET factor	% MARV reduction (difference from observed data)	NSE	PBIAS (%)
Observed	-		-	-	-	8.5%	-	-
Base Case	0.3		100	10	2.1	40.3% (+31.8%)	0.54	34.8
1	0.3		100	2	2.1	20.6% (+12.1%)	0.57	13.3
2	0.3		100	1	2.1	14.5% (+6.0%)	0.57	6.6
3	0.3		50	0.3	2.1	7.1% (-1.3%)	0.56	-1.4
4	0.3		150	0.4	2.1	8.5% (0.0%)	0.56	0.07
5	0.3		100	0.3	2.1	7.1% (-1.4%)	0.56	-1.5%
6	0.3		100	0.4	2.1	8.5% (0.0%)	0.56	0.0%
PEST 2 Par*	0.3		300	1.67	2.1	18.6% (10.2%)	0.61	13.0
PEST 3 Par**	0.3		245.16	1.055	2.08029	14.8% (6.4%)	0.58	9.7
9	0.3	0.2	100	0.4	2.1	8.0%	0.55	-0.5%
10	0.3	0.25	100	0.4	2.1	8.2%	0.55	-0.3%
11	0.3	0.25	100	0.5	2.1	9.4%	0.55	1.0%
12	0.3	0.25	50	0.5	2.1	9.5%	0.62	1.1%
13	0.3	0.25	50	0.4	2.1	8.2%	0.62	-0.2%
14	0.3	0.25	75	0.4	2.1	8.2%	0.57	-0.3%
15	0.3	0.25	25	0.4	2.1	8.3%	0.70	-0.1%
16	0.3	0.25	15	0.4	2.1	8.4%	0.71	0.0%
17	0.3	0.3	15	0.4	2.1	8.6%	0.71	0.1%
18	0.3	0.3	10	0.4	2.1	8.7%	0.72	0.3%
19	0.3	0.3	10	0.35	2.1	8.0%	0.72	-0.5%

*Saturated hydraulic conductivity and exfiltration

**Saturated hydraulic conductivity, exfiltration and PET factor

7.1.4 Discussion

We have demonstrated that with calibration, MUSIC is capable to produce reasonably realistic outflow hydrographs that follow the observed patterns with good statistical (Nash Sutcliffe Efficiency) outcomes, assessed at a high resolution (6 minute timestep). Furthermore, it can usually provide reasonably good predictions of mean annual flows (or flow volumes over the period of observations at least). In some cases, both of these can be produced simultaneously, in others there may be trade-offs between predictions of the hydrograph and mean annual flows with parameter sets that tend to predict one well while predicting the other less well.

Our experience from the study indicates that successful calibration of MUSIC is challenging, particularly for long continuous time series and requires interpretation of the design, careful setup and configuration of the model to reflect actual context and conditions as well as augmenting or infilling of the monitoring data to fill in gaps or missing data. One challenge is simultaneously achieving good representation of both the overall outcomes (i.e. water balance and volume reductions) as well as temporal matching of flows within an event.

The calibration results indicate that the stormwater flow volume reductions are highly sensitive to the exfiltration rate, which is not surprising. The Nash-Sutcliffe efficiency, representing the hydrograph is also sensitive to the saturated hydraulic conductivity as may be expected. The PET factor was usually not that sensitive and was often set to a lower value, although it did not make a substantial difference for the assets calibrated (which are of typical sizes for stormwater quality treatment performance and are relatively small with respect to their catchment).

The results have clearly identified one limitation in representing actual asset designs in MUSIC is the lack of a constrained outlet pipe in MUSIC as some experimental systems have either high submerged zones or constrained outlet pipes. Interestingly, SWMM is reported to have this capability (Lisenbee et al., 2021). For conventional design of biofilters, this is potentially academic as it is less common in current practice, although it was featured in one of the field sites.

However, we expect a renewed focus on optimising stormwater volume reductions using bioretention following introduction of new guidelines in some parts of Australia. These are likely to put a renewed focus on optimising retention performance of biofilters and restricting the outlet is an obvious and cost-effective way to improve both retention performance as well as downstream flow patterns. For the latter, there is obvious potential application in Western Sydney where stormwater treatment assets must be designed to achieve stormwater flow retention objectives as well as meet a series of ‘gates’ along the flow duration curve. The outflows from an appropriately sized biofilter can often exceed these limitations and the ability to constrain this would improve their flexibility in meeting these guidelines. Taking these into account, we can expect to see a broader range of biofilter configurations and there is a need for our models and tools to adapt to support greater flexibility in design.

8 Key findings and conclusions

Green infrastructure stormwater retention performance

In Stage 1 of this study the following findings were made:

- Confirmed biofilters, tree pits and green roofs are generally effective for reducing stormwater volumes.
- As expected, performance varies widely according to climate, soils, design, and size relative to catchment
- Biofilters and tree pits are typically small relative to catchment and infiltration is usually the dominant pathway for volume reductions. Where possible, biofilters and tree pits should be unlined to enable infiltration and support waterway baseflows
- Satisfactory outcomes may still be achieved in slow draining soils or even lined assets subject to context, design and potentially adopting larger sizing than is required for stormwater quality purposes.
- Green roofs are typically large relative to catchment and evapotranspiration is the pathway for volume reductions. Green roofs can substantially reduce stormwater volumes from roofs. It is noted that Sydney Water is collaborating with WSU to measure the performance of "purple roofs" which also incorporate detention and may reduce or eliminate the need for detention storage tanks. Green roofs may potentially be approximated using the biofilter node in MUSIC as the best currently available option, although we caution there are differences in structural configuration and potential performance. The representation of 'purple' roofs incorporating detention will necessitate the capacity to constrain outflows in some way or introduce a delay or roughness coefficient into the submerged zone.
- Within or intra-event processes depend mostly on inflow patterns. Between or inter-event processes on local climate and soil conditions for corresponding evapotranspiration and infiltration. Calibration for intra-event processes can potentially draw on data from many sources and contexts while continuous data for a local or similar climate is ideal to support inter-event process simulation.
- Five Australian detailed data sets for bioretention assets were obtained through this study that have previously not been readily or publicly available. Three of these have continuously monitored data and two have event data. These provide a basis for field calibration of biofilter models.

Statistical analysis

Statistical analysis was undertaken in an effort to identify the most influential factors driving performance. This built on work to understand design and operational factors influencing water quality treatment (Zhang et al, 2021).

We investigated the influence of:

- Design variables: FMD, EDD, SZD, area ratio
- Operation variables: ADWP, infiltration average, flux in, evaporation average
- other design variables: raingarden area, storage volume
- other operational variables: 3 day rainfall, event duration.

The analysis incorporated a range of field data from the literature review and yielded a range of results. However, the outcomes for the confidence indicators were low, indicating high levels of uncertainty and that the results could not necessarily be relied upon. It was therefore difficult to draw any conclusions with confidence. The challenge was that when drawing on a database of existing studies, these have been undertaken by different research groups, in different contexts in terms of climate and soils, using different assets and configurations. This means that there are a large number of variables and factors that may influence performance in addition to measurement uncertainties and errors. This resulted in a heterogenous data set with a high ratio of noise to signal. Our conclusions are not dis-similar to prior efforts such as (Poresky et al., 2012) which found a few tenous relationships but that it was difficult to clearly identify or articulate the key driving factors for performance based on the data. While a range of likely influential factors were chosen for analysis, it is also possible that other factors not considered may influence results, for example differences in vegetation species, form and maturity.

This is not to say this approach cannot work, but that either a dataset with a greater level of consistency may be needed (such as a series of assets with some common features, context or monitoring programs) or deeper analysis and more powerful statistical approaches could possibly be used to gain some insight. It is likely that targeted analysis of a small number of well known assets may yield greater insights and understanding than attempting to work with the large and complex database available from the literature in entirety.

Some key learnings that we can draw from this work are:

- The level of noise in the existing literature data makes it difficult to draw firm conclusions on driving factors for performance
- A sub-set of higher quality data may be required to gain further insights
- Targeted pair-wise studies may be a preferable way of identifying the effects of various factors on performance (e.g. assets with one change in design configuration or identical assets in different locations)
- The diversity of field assets make it difficult to assess statistically and model calibration approaches may possibly be a preferable approach

MUSIC model calibration

MUSIC (v6.3) was calibrated for continuous or event monitored outflow data for three selected biofilter assets.

We have demonstrated that with calibration, MUSIC is capable to produce reasonably realistic outflow hydrographs that follow the observed patterns with good statistical outcomes based on Nash Sutcliffe Efficiency (NSE), assessed at a high resolution 6 minute timestep.

MUSIC can provide good predictions of stormwater flow volumes and percentage reductions over the period of record. In some cases both of these can be produced simultaneously, in others there may be trade-offs between predictions of the hydrograph and mean annual flows with parameter sets that predict one well predicting the other more poorly. Similar findings where a good fit to one objective results in a poorer fit to another have been found by others (Lintern et al., 2012).

Successful calibration of MUSIC is challenging, particularly for long continuous time series and requires interpretation of the design, careful setup and configuration of the model to reflect actual context and conditions as well as augmenting or infilling of the monitoring data to fill in gaps or missing data. One challenge is achieving good representation of both the overall outcomes (i.e. water balance and volume reductions) as well as temporal matching of flows within an event.

Outcomes and learnings for saturated hydraulic conductivity and exfiltration rates

The (filter media) saturated hydraulic conductivity and (underlying soil) infiltration/exfiltration rate can vary widely and be difficult to measure accurately. Where laboratory measurements were undertaken, the calibration process indicated that (an arguably more precise estimate) of the model values of these (to match outflows) may still differ from those measured. In the case of Wicks Reserve, the calibrated exfiltration rate was at the lower end of the range predicted from geotechnical assessment (0.03 c.f. 0.2-2 mm/hour) while for Hereford Road the adopted value of 0.4 mm/hour it was well below the measured range of 2-17 mm/hour.

Comparing outcomes between a model configured simply following guideline recommendations and defaults with a calibrated model, it is clear that there is a significant difference in outcomes and that just using default and guideline values may not provide reliably accurate estimates of performance. This is not surprising as guidelines tend to take a conservative approach, recommending use of parameters that may predict lower performance than the design intent.

The key differences observed were:

- Guidelines may recommend the use of a zero exfiltration rate or a conservative exfiltration rate to soils based on geotechnical testing. The results from the study suggested that an even lower exfiltration rate may be appropriate to represent observed flow conditions and that infiltration is therefore less than anticipated (but also higher than zero in most cases including slow draining soils and potentially even when lining is attempted).

- Guidelines may recommend use of a saturated hydraulic conductivity of 100 mm/hour while designing for 100-300 mm/hour.
    - For Monash, we found a higher calibrated value of 190 mm/hour, indicating that actual performance is likely closer to the design intent.
    - If the actual realised saturated hydraulic conductivity is higher, then stormwater treatment performance may be higher than predicted with guideline values. It is recommended sensitivity analysis is used to understand the range of likely outcomes (though guideline design should still be conservative and assume the lower bound of 100 mm/hour to provide a high likelihood of achieving performance requirements)
    - The saturated hydraulic conductivity for the other two assets was highly modified to reflect the specific behaviour of the asset and is not directly useful other than to remind that the parameters in MUSIC need to be understood as a blend between physical and 'black box' parameters. It may be necessary to use physically unrealistic parameters in some cases to adequately represent actual conditions – it is however problematic to determine what these should be in the absence of a specific calibration for a given constructed asset.
    - As an interim approximation for constrained outlet biofilters, the saturated hydraulic conductivity could potentially be constrained to match the outflow rate (calculated in mm/hour for the surface area) for the constrained outlet. This would necessarily have high uncertainty and a greater level of conservatism should be adopted in using the results.

The literature identifies that infiltration is usually the dominant retention pathway where it is allowed to occur with evapotranspiration usually being smaller. The model has high (even very high) sensitivity to the exfiltration rate and this *highlights the importance of this parameter being determined as accurately as possible, ideally using geotechnical testing at the expected depth and in multiple locations for large or significant assets*.

Furthermore, *sensitivity testing should be undertaken for a range of values where stormwater volume reductions are an objective and infiltration is a significant portion of the overall site or precinct water balance.*

Model representation of varying biofilter configurations

We have found that the model performed quite well when simulating an asset configuration for a 'standard' biofilter configuration for which it was designed. This is an asset that includes a filter media of reasonable depth above the outlet, submerged zone and no constraint on outflows.

However, the results indicate a key limitation in representing some asset designs in MUSIC is the lack of functionality to constrain outflows in MUSIC. Some biofilter assets (such as Wicks Reserve) have constrained outlet pipes or a high outlet (Hereford Road). These cannot be directly represented in MUSIC using physically realistic parameters, however it was possible to achieve a numerical calibration to the data by choosing parameter values without physical limitations. For Wicks Reserve, this included using a very low saturated hydraulic conductivity for the filter media to approximate the constrained outflow and a larger extended detention depth to account for storage volume. For Hereford Road, a very low saturated hydraulic conductivity was adopted. We were able to obtain remarkably good results through this approach. However, given the extent of adjustment of the parameters required and implications for how processes are represented, the transferability of the calibrated results to other contexts is questionable and likely would not provide reliable predictions of behaviour and performance.

Current design practice generally assumes a free outlet on biofilters. Therefore, the above could be interpreted as a limitation of the field assets available rather than the model and does not limit its use for common current practice.

On the other hand, there are clear potential benefits to using constrained outlets on biofilters. This can increase stormwater volume retention performance which is a key area of focus. It could also allow biofilters to provide control of flow rates to provide appropriate flow patterns for waterway health as well as effective and more reliable flood mitigation outcomes.

Therefore, there is a clear opportunity for new or improved biofilter models to be developed that can support a constrained outlet. We note that SWMM already has this capability (Lisenbee et al., 2021).

Recommendation: Develop a new/improved biofilter model with constrained outlet

Learnings for modelling and interpretation of modelling results

The following draws on outcomes from the calibrations as well as broader industry experience.

Can we use MUSIC with confidence for modelling stormwater flow volume (percentage MARV) reductions?

Yes. MUSIC is capable of predicting stormwater flow volume reductions with reasonable confidence, provided that it is appropriately parameterised. It is most effective for assets with standard or typical design configurations aligned with the model structure.

Under what conditions should we be cautious about using or relying on MUSIC results for assessment of performance?

For assets that vary from standard configurations, for example those with a constrained outlet, a high outlet, tree pits with large differences in canopy and filter area, green roofs. For these varying configurations, the level of uncertainty in the performance may be much greater.

We may still use the biofilter module for these given limited suitable alternatives, however given the greater uncertainty sensitive testing is more important and we should adopt a higher level of conservatism. The use of alternative models as a cross-check (for example SWMM for stormwater volumes) may also be considered.

Under what conditions can we use MUSIC for this type of modelling?

It is essential that MUSIC is appropriately parameterised to accurately represent stormwater flow volumes. This means that greater attention must be given to:

- Impervious fractions and soil parameters to more accurately estimate catchment flows (ideally with reference to calibrated values for the same or a similar catchment)
- Use of an appropriate and representative 'exfiltration rate', ideally based on geotechnical testing of sub-soil infiltration rates. Even these can significantly under or over-estimate realised outcomes and sensitivity analysis as well as post construction monitoring where possible is recommended.
- Saturated hydraulic conductivity should be modelled at a conservative value (e.g. 100 mm/hour) for assessing treatment performance but sensitivity should also be undertaken at the design saturated hydraulic conductivity (average or ends of the specified range) to understand potential variations in performance. This may be particularly important where outflow rates into a waterway are of interest such as in Western Sydney.
- The PET factor should be revised to best represent the expected conditions. It will usually only be sensitive for assets with a large treatment area relative to catchment (>~5%) or where the effective area over which PET occurs varies significantly from the filter area. The value of 2.1 may be regarded as indicative of likely performance of small to medium sized biofilters and raingardens. This should potentially be reduced for large assets where evapotranspiration from surrounding areas of the asset itself reduces the potential evapotranspiration from a given point.

Recommendations on sensitivity

There are model input parameters with significant uncertainty. Current industry practice is to use the model deterministically using a single input value for these. There is a need for greater awareness of the range of uncertainty in these values and the implications for results. This may be addressed through sensitivity testing where models are run for a range or distribution of input parameters to better understand the range of likely outcomes and the sensitivity of the model to variations in the inputs. Sensitivity testing should be undertaken to test sensitivity of outcomes to a range of input values for parameters with uncertainty (such as saturated hydraulic conductivity and exfiltration rate) to ensure that the asset will still achieve its desired performance objectives (or be within an acceptable shortfall range) should these numbers vary from what was assumed. There is barely any guidance on sensitivity testing available to industry at present.

It is recommended that a guide (or addendum to an existing guideline) is prepared to provide guidance to industry on sensitivity testing required. This should ideally involve input from both researchers and industry to ensure it is both practical and robust.

Recommendations on monitoring and assessment

There is currently little or no monitoring, assessment or field testing to confirm that assets deliver on their design performance. Historically, this has been largely due to cost, primarily for stormwater quality sampling. However, where many large biofilter assets are being constructed, these may involve substantial investments of both public and private funding. The absence of testing and validation mean that these *significant investments may at risk* or not be as well targeted as they could be.

Even a modest investment in monitoring and testing of WSUD assets could yield significant increases in environmental benefits realised and/or cost savings.

There is a need for budgets to be allocated or set aside to fund post construction monitoring and validation of performance. This may involve for example a contribution based on the construction cost or size of development (say 1-3% of capital cost) for each asset constructed towards a pooled fund to support future monitoring and assessment.

This may include monitoring of inflows, outflows (and ideally overflows) to validate stormwater flow volume losses. It would also be beneficial to use this data to calculate infiltration rates to inform future asset design and values for the exfiltration rate parameter.

It is important that the design and as-constructed details of these assets are captured, that the monitoring is undertaken in a robust manner with good quality control and checks and that consistent approaches are used.

It is recommended that a program is established to fund and progress field testing of a proportion of constructed biofilters (and other WSUD assets) to support assessment of performance and improved design guidance for biofilters.

Recommendations related to passively irrigated tree pits

While passively irrigated tree pits were not the specific focus for the calibrations, they are commonly also modelled using the biofilter node in MUSIC. A number of learnings and comments can be made with respect to modelling of these assets may be made.

- The PET factor should be revised to best represent the expected conditions. The default value of 2.1 is based on Carex Appressa in a well-watered biofilter column in a greenhouse and factoring up the underlying areal potential evapotranspiration (APET) (Also called Morton's wet environment evapotranspiration) which is the input data into MUSIC.

- It is likely that lower values may be appropriate for trees and values of 1.85 and 1.5 have tentatively been adopted for high and moderate water use trees respectively. MUSIC represents evapotranspiration as occurring as the filter media area, which would usually correspond to the soil surface area in a passively irrigated tree pit (to ensure soil storage volumes are correct). In many cases the tree may extend well beyond this at maturity. The PET factor may therefore be factored upwards to reflect this. For example:
    - PET factor = Canopy area / filter area * base PET factor = 20 / 15 * 1.85 = 2.4
- The biofilter node is likely adequate to approximate the hydrologic behaviour and performance of a passively irrigated tree pit. However, where the structural configuration of this varies significantly from the standard biofilter structure in MUSIC, there may be much greater uncertainty in the results. This should be taken into consideration in the design by, for example sensitivity testing or building in a factor of safety. This has implications for representing hybrid assets that include structural soils under pavement, adjacent infiltration trenches and other innovations where the uncertainty of results may be greater.

9 Next steps

The following next steps are planned for this study:

- Publish findings in a journal paper
- Present findings and outcomes to stakeholders

The following recommendations are made for next steps and future research arising from the study:

- Establish a program to fund and progress field testing of a proportion of constructed biofilters (and other WSUD assets) to support assessment of performance and improved design guidance for biofilters
- Develop improved models for biofilters supporting a broader range of configurations such as constrained and high outlets, passively irrigated tree pits, adjacent structural soils and green roofs.
- Pursue further calibrations for other sites and models to provide an evidence base to support a range of models and tools that are fit for purpose for predicting performance of biofilters.
- Develop a guide or addendum to an existing guideline providing guidance to industry on what sensitivity testing is required and how this should be undertaken during assessment of biofilter (and more broadly WSUD asset) performance.
- Establish a program and database to allow outcomes of soil infiltration testing to be recorded, allow soil types to be characterised and more broadly inform future design.
- Further assess and refine use of potential evapotranspiration methods (important for assets that are large relative to catchment such as green roofs and vegetated sponges)
- Expand calibration work from this study to include similar assets of interest including passively irrigated street trees, green roofs and permeable paving.
- Evapotranspiration is likely to be more important for large scale assets such as green roofs, wicking beds and vegetated sponges. These assets are in their infancy but have significant potential for volume reductions. A monitoring and model calibration study focussing on these assets and seeking to confirm appropriate models and parameters for evapotranspiration would be valuable to validate their efficacy and support future adoption.

10 Acknowledgements

We give our thanks to all the data providers who generously provided the underlying monitoring data to support this study including Tim Fletcher, Belinda Hatt, Jeremie Bonneau and Peter Poelsma as well as relevant colleagues. The study would not have been possible without this data. Also Kefeng Zhang and Anna Lintern who generously made their statistical model available and provided support.

We also acknowledge the support of Mohammed Sabbagh and his supervisor Brandon Winfrey from Monash University. Mohammed undertook significant additional calibration efforts and these as well as early drafts of journal papers have helped inform this work.

We also thank our client for their long-standing and unwavering support throughout the long journey this project has involved and for their vision and willingness to challenge business as usual approaches and seek to develop greater understanding of the real performance of WSUD assets.

11 References

Beven J., K. (2001). *Rainfall-Runoff Modelling: The Primer*. John Wiley & Sons Ltd.

Bonneau, J., Fletcher, T. D., Costelloe, J. F., Poelsma, P. J., James, R. B., & Burns, M. J. (2020). The hydrologic, water quality and flow regime performance of a bioretention basin in Melbourne, Australia. *Urban Water Journal*. https://doi.org/10.1080/1573062X.2020.1769688

Browne, D., & Deletic, A. (2011). Predicting the performance of stormwater infiltration systems. In *Department of Civil Engineering: Vol. Doctor of*.

Browne, D., Deletic, A., Mudd, G. M., & Fletcher, T. D. (2013). A two-dimensional model of hydraulic performance of stormwater infiltration systems. *Hydrological Processes*, *27*(19). https://doi.org/10.1002/hyp.9373

Browne, D., Mcarthur, J., & Luxton, J. (2020). *Green infrastructure stormwater retention performance*.

Burns, M., Fletcher, T., Duncan, H., Hatt, B., Ladson, A., & Walsh, C. J. (2012). *The stormwater retention performance of rainwater tanks at the landparcel scale*.

Clearwater. (2012). *Wicks Reserve Bio-Infiltration System*.

Davis, A. P., Traver, R. G., Hunt, W. F., Lee, R., Brown, R. A., & Olszewski, J. M. (2012). Hydrologic Performance of Bioretention Storm-Water Control Measures. *Journal of Hydrologic Engineering*. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000467

DeBusk, K., Hunt, W., & Line, D. (2010). Bioretention Outflow: Does It Mimic Non-Urban Watershed Shallow Interflow? In *Journal of Hydrologic Engineering* (Vol. 16). https://doi.org/10.1061/41114(371)313

Doherty, J. (2015). Calibration and Uncertainty Analysis for Complex Environmental Models. *Groundwater*, 227.

Doherty, J. (2024). *Home | PEST*. https://pesthomepage.org/

Dotto, C. B. S., Deletic, A., & Fletcher, T. D. (2009). Analysis of parameter uncertainty of a flow and quality stormwater model. *Water Science and Technology*, *60*(3), 717–725. https://doi.org/10.2166/wst.2009.434

Dotto, C. B. S., Deletic, A., McCarthy, D. T., & Fletcher, T. D. (2011). Calibration and Sensitivity Analysis of Urban Drainage Models: Music Rainfall/Runoff Module and a Simple Stormwater Quality Model. *Australasian Journal of Water Resources*, *15*(1), 85–94. https://doi.org/10.1080/13241583.2011.11465392

Duncan, H. P., Fletcher, T. D., Vietz, G., Urrutiaguer, M., Duncan, H., Fletcher, T., Bos, D., & Walsh, C. (2014). *The feasibility of maintaining ecologically and geomorphically important elements of the natural flow regime in the context of a superabundance of flow-Stage 1: Kororoit Creek Study*.

EPA. (2021). *Urban stormwater management guidance*. https://www.epa.vic.gov.au/about-epa/publications/1739-1?utm_medium=email&utm_campaign=Urban+stormwater+response+to+comments&utm_content=epa.vic.gov.au%2Fabout-epa%2Fpublications%2F1739-1&utm_source=www.vision6.com.au

EPA Victoria. (2021). *Urban stormwater management guidelines*. https://www.epa.vic.gov.au/-/media/epa/files/publications/urban-stormwater-management-1739-1.pdf?la=en&hash=297A34E91AFF7A774C472B9F338063C8

eWater. (2024). *MUSIC Manual*.

FAWB. (2009). *Adoption Guidelines for Stormwater Biofiltration Systems*. Facility for Advancing Water Biofiltration, Monash University.

Fowdar, H. S., Neo, T. H., Ong, S. L., Hu, J., & McCarthy, D. T. (2022). Performance analysis of a stormwater green infrastructure model for flow and water quality predictions. *Journal of Environmental Management*, *316*, 115259. https://doi.org/10.1016/J.JENVMAN.2022.115259

Hatt, B. E., Fletcher, T. D., & Deletic, A. (2009). Pollutant removal performance of field-scale stormwater biofiltration systems. *Water Science and Technology*. https://doi.org/10.2166/wst.2009.173

Hennekam, B. (2021). *Validation of the Model for Urban Stormwater Improvement Conceptualisation*. University of South Australia.

Hoban, A., & Gambirazio, C. (2018). Implications of recent research on stormwater quality targets and practices. In Sydney (Ed.), *Stormwater Industry Association of Australia*. Stormwater Australia. https://www.stormwater.asn.au/75-conference-papers

Hoban, A., & Gambirazio, C. (2021). Bioretention performance: a review of field studies. *Https://Doi.Org/10.1080/13241583.2021.1984190*, *25*(2), 202–210. https://doi.org/10.1080/13241583.2021.1984190

Imteaz, M. A., Ahsan, A., Rahman, A., & Mekanik, F. (2013). Modelling stormwater treatment systems using MUSIC: Accuracy. *Resources, Conservation and Recycling*, *71*, 15–21. https://doi.org/10.1016/J.RESCONREC.2012.11.007

Lintern, A., Daly, E., & Deletic, A. (2012). *Verifying a stormwater biofiltration model*.

Lisenbee, W. A., Hathaway, J. M., Burns, M. J., & Fletcher, T. D. (2021). Modeling bioretention stormwater systems: Current models and future research needs. *Environmental Modelling & Software*, 105146. https://doi.org/10.1016/J.ENVSOFT.2021.105146

Liu, J., Sample, D. J., Bell, C., & Guan, Y. (2014). Review and research needs of bioretention used for the treatment of urban stormwater. In *Water (Switzerland)* (Vol. 6, Issue 4, pp. 1069–1099). MDPI AG. https://doi.org/10.3390/w6041069

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

NSW Department of Planning and Environment. (2022). *Western Sydney Aerotropolis Development Control Plan 2022*. https://shared-drupal-s3fs.s3.ap-southeast-2.amazonaws.com/master-test/fapub_pdf/NSW+Planning+Portal+Exhibitions/Western+Sydney+Aerotropolis+Development+Control+Plan+-+Phase+2.pdf

NSW DPIE. (2022). *Western Sydney Aerotropolis Development Control Plan 2022*. https://pp.planningportal.nsw.gov.au/western-sydney-aerotropolis-DCP

Parker, N. (2010). Assessing the effectiveness of Water Sensitive Urban Design in SouthEast Queensland. In *Faculty of Built Environment and Engineering*.

Peng, Z., & Stovin, V. (2017). Independent validation of the SWMM green roof module. *Journal of Hydrologic Engineering*. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001558

Platz, M., Simon, M., & Tryby, M. (2020a). Testing of the Storm Water Management Model Low Impact Development Modules. *Journal of the American Water Resources Association (JAWR)*, *56*(2), 283–296. https://onlinelibrary.wiley.com/doi/epdf/10.1111/1752-1688.12832

Platz, M., Simon, M., & Tryby, M. (2020b). Testing of the Storm Water Management Model Low Impact Development Modules. *Journal of the American Water Resources Association*, *56*(2), 283–296. https://doi.org/10.1111/1752-1688.12832

Poelsma, P. J., Fletcher, T. D., & Burns, M. J. (2013). Restoring natural flow regimes : the importance of multiple scales La restauration d ' un régime d ' écoulement naturel : la mise en œuvre à plusieurs échelles. *Novatech 2013*, 1–10. https://watersensitivecities.org.au/content/restoring-natural-flow-regimes-the-importance-multiple-scales/

Poresky, A., Bracken, C., Strecker, E., & Clary, J. (2012). *Expanded Analysis of Volume Reduction in Bioretention BMPs* (Issue January 2011). http://www.bmpdatabase.org/

Randelovic, A., Zhang, K., Jacimovic, N., McCarthy, D., & Deletic, A. (2016). Stormwater biofilter treatment model (MPiRe) for selected micro-pollutants. *Water Research*. https://doi.org/10.1016/j.watres.2015.11.046

Spraakman, S., Rodgers, T. F. M., Monri-Fung, H., Nowicki, A., Diamond, M. L., Passeport, E., Thuna, M., & Drake, J. (2020). A need for standardized reporting: A scoping review of bioretention research 2000–2019. *Water (Switzerland)*, *12*(11), 1–35. https://doi.org/10.3390/w12113122

State of New South Wales and Department of Planning and Environment. (2022). *Technical guidance for achieving Wianamatta-South Creek stormwater management targets*. www.environment.nsw.gov.au

Sydney Water, & E2Designlab. (2019). *Stormwater Retention and Detention for WSUD Stormwater Retention and Detention for WSUD | E2Designlab*.

Tanner, C., Winfrey, B., Fletcher, T., O'Neill, A., & Leinster, S. (2020a). *Review of Recent Papers About Stormwater Quality Management Targets and Practices in Queensland 17 Review of recent papers about stormwater quality management targets and practices in Queensland Review of Recent Papers About Stormwater Quality Management Targets and Practices in Queensland*. www.watersensitivecities.org.au

Tanner, C., Winfrey, B., Fletcher, T., O'Neill, A., & Leinster, S. (2020b). *Review of Recent Papers About Stormwater Quality Management Targets and Practices in Queensland 17 Review of recent papers about stormwater quality management targets and practices in Queensland Review of Recent Papers About Stormwater Quality Management Targets and Practices in Queensland*. www.watersensitivecities.org.au

Victoria Stormwater Committee. (1999). *Urban Stormwater Best Practice Environmental Management Guidelines*.

Walsh, C. J., Fletcher, T. D., & Burns, M. J. (2012). Urban Stormwater Runoff: A New Class of Environmental Flow Problem. *PLOS ONE*, *7*(9), e45814. https://doi.org/10.1371/journal.pone.0045814

WBM, B. (2015). *NSW MUSIC Modelling Guidelines* (Issue August).

Zhang, K., Liu, Y., Deletic, A., McCarthy, D. T., Hatt, B. E., Payne, E. G. I., Chandrasena, G., Li, Y., Pham, T., Jamali, B., Daly, E., Fletcher, T. D., & Lintern, A. (2021). The impact of stormwater biofilter design and operational variables on nutrient removal - a statistical modelling approach. *Water Research*, *188*, 116486. https://doi.org/10.1016/j.watres.2020.116486

Zhang, K., Randelovic, A., Page, D., McCarthy, D. T., & Deletic, A. (2014). The validation of stormwater biofilters for micropollutant removal using in situ challenge tests. *Ecological Engineering*, *67*, 1–10. https://doi.org/10.1016/j.ecoleng.2014.03.004

12 Appendix 1

Table 10 Field monitored biofilter asset data (1 of 3)

Paper	**Bonneau2020a**	**Davis2008a**	**Davis2012a**	**Davis2012a**	**Davis2012a**	**deMacedo2019**	**Hatt2009a**	**Hatt2009a**	**Hatt2012**	**Hunt2006a**
Location	Melbourne, VIC	Maryland, US	Maryland, US	North Carolina, US	Villanova, US	San Carlos, Brazil	McDowall, Qld	Monash University, Vic	Clifton Hill, Melbourne, Victoria	NC, USA
Asset	Wicks Reserve Biofilter	1	Maryland Bioretention	North Carolina Bioretention	Villanova bioretention	1	McDowall	Monash car park biofilter	Clifton Hill Biofilter	G2
No. events	96	49	124	364	124	14	4	24		48
Period (months)	36	24				36	N/A	6	9	12
Catchment area (m^2)	330,000	2,400	1,836	2,200	5,261	23,000	1,000	4,500	73,000	2,000
Impervious fraction (%)	15%	100%	85%	76%	50%	25%	100%	100%	40.00%	
Area (m^2)	1800	52.8	102	146	149	60.63	20	45	200	100
Ks (mm/hour)		-				5.83				-
EDD (m)	0.35		0.3	0.16	0.25		0.2	0.25	0.175	-
Filter media depth (m)	0.35	1	0.9	1.1	1.2	0.5	0.4	0.5	0.5	1.2
Transition layer depth (m)	0.15	-	-	-	-	0.7	0.1	-	0.1	0
Drainage layer depth (m)	0.3	-		0.7		2	0.2	0.15	0.15	0

Paper	Bonneau2020a	Davis2008a	Davis2012a	Davis2012a	Davis2012a	deMacedo2019	Hatt2009a	Hatt2009a	Hatt2012	Hunt2006a
Underdrainage	Y	Y	Y	Y	N	N	Y	Y	Y	Y
Filter media	Loamy sand	Sand, topsoil, organics	Sandy Loam	Sand	Sandy Loam	Natural soil (medium sand, fine sand and dark clay)	Sandy Loam	Sandy Loam	Loamy sand, perlite, vermiculite and topsoil	-
Submerged zone depth (m)	0.5	0	0	0.7	0	2.7	0	0		0
Lined	N	N	N	N	N	N	Y	Y	Y	N
Underlying soil	Heavy clay	-	Clay	Sandy Loam	Loam (50% sand 20% clay)		-	N/A		Clay Loam
Rainfall (mm/year)	730	1070	1070	1140	1040	1362	1140	680	650	1096
Stormwater retention (%)	35%	75.50%	77.50%	86.40%	51.60%	70%	20.10%	33.00%	15%	78.20%

Table 11 Field monitored biofilter asset data (2 of 3)

Paper	Lucke2015a	Mahmoud2019	McKenzie-McHarg2008a	Parker2010a	Passeport2009a	Passeport2009a	Peljo2016a	Peljo2016a	Peljo2016a	Peljo2016a
Location	Caloundra, Sunshine Coast, QLD		Saturn Crescent, Qld	Qld	Ohio, USA	Ohio, USA	Bells Reach, Caloundra, Qld	Bells Reach, Caloundra, Qld	Bells Reach, Caloundra, Qld	Bells Reach, Caloundra, Qld
Asset		1	Saturn Crescent	1	N	S	Bioretention A	Bioretention B	Bioretention C	Bioretention D
No. events	12	45	4	22	16	13	1	1	1	1
Period (months)	5	13	N/A		12	12	N/A	N/A	N/A	N/A
Catchment area (m^2)	55	1,619		6,530	3,450	3,450	1,550	320	1,210	290
Impervious fraction (%)	100%	100%		52%	40%	40%	NR	NR	NR	NR
Area (m^2)	9.1	55	20	250	102	102	10.9	14.5	13.5	15.8
Ks (mm/hour)	180	130		-	-	-		106		229
EDD (m)	0.1	0.3		0.1	0.23	0.23	0.2	0.2	0.2	0.2
Filter media depth (m)	0.9	0.76		0.8	0.6	0.9	0.6	0.6	0.6	0.6
Transition layer depth (m)	0.1			0.2	-	-	0.1	0.1	0.1	0.1

Paper	Lucke2015a	Mahmoud2019	McKenzie-McHarg2008a	Parker2010a	Passeport2009a	Passeport2009a	Peljo2016a	Peljo2016a	Peljo2016a	Peljo2016a
Drainage layer depth (m)	0.2	0.15-0.23		0	0.15	0.15	0.2	0.2	0.2	0.2
Underdrainage	Y	Y		Y	Y	Y	Y	Y	Y	Y
Filter media	Sandy loam	Washed sand		-	Slate fines, sand and organics	Slate fines, sand and organics	Sand	Sand	Sand	Sand
Submerged zone depth (m)	0	0		0	0.45	0.75				
Lined	Y	Y		N	N	N	N	N	N	N
Underlying soil				Silts	Loamy clay	Sandy Loam	Sand/sandy loam	Sand/sandy loam	Sand/sandy loam	Sand/sandy loam
Rainfall (mm/year)	1140	526.5	1140	1320	1140	1140	1686	1686	1686	1686
Stormwater retention (%)	61%	78%	23%	42%	18%	14%	69%	39%	72%	87%

Table 12 Field monitored biofilter asset data (3 of 3)

Paper	Poelsma2013	Shrestha2018	Shrestha2018	Shrestha2018	Shrestha2018	Shrestha2018	Trowsdale2011a	Winston2016a	Winston2016a
Location	Mt Evelyn, VIC	Shrestha2018_1_Bioretention	Shrestha2018_2_Bioretention	Shrestha2018_3_Bioretention	Shrestha2018_4_Bioretention	Shrestha2018_5_Bioretention	Paul Matthews Rd, Auckland, NZ	Ohio, USA	Ohio, USA
Asset	Hereford Road Biofilter	1	2	3	4	5	Paul Matthews Road Biofilter	HA Bioretention	UC Bioretention
No. events	196	17	37	35	16	16	12	90	90
Period (months)	9	15	15	15	15	15		13	7
Catchment area (m^2)	9,800						18,000	4,600	3,600
Impervious fraction (%)	100%						86%	58%	77%
Area (m^2)	100	3.7	3.72	3.72	3.7	3.7	200	136	182
Ks (mm/hour)								100	168
EDD (m)	0.3	0.15	0.15	0.15	0.15	0.15		0.39	0.3
Filter media depth (m)	0.4	0.61	0.61	0.61	0.61	0.61		0.84	0.6
Transition layer depth (m)	0.2	0.076	0.076	0.076	0.076	0.076		0.15	0.15

Paper	Poelsma2013	Shrestha2018	Shrestha2018	Shrestha2018	Shrestha2018	Shrestha2018	Trowsdale2011a	Winston2016a	Winston2016a
Location	Mt Evelyn, VIC	Shrestha2018_1_Bioretention	Shrestha2018_2_Bioretention	Shrestha2018_3_Bioretention	Shrestha2018_4_Bioretention	Shrestha2018_5_Bioretention	Paul Matthews Rd, Auckland, NZ	Ohio, USA	Ohio, USA
Asset	Hereford Road Biofilter	1	2	3	4	5	Paul Matthews Road Biofilter	HA Bioretention	UC Bioretention
Drainage layer depth (m)	0.4	0.23	0.23	0.23	0.23	0.23		0.3	0.3
Underdrainage	Y	Y	Y	Y	Y	Y		Y	Y
Filter media	Loamy sand	Sand compost mix over sand							
Submerged zone depth (m)	0.9							0.38	0.6
Lined	N	Y	Y	Y	Y	Y		N	N
Underlying soil	Heavy clay								
Rainfall (mm/year)	997	934	934	934	934	934		1,010	1,010
Stormwater retention (%)	8%	80%	82%	77%	48%	86%	58%	39%	59%