A Data-Driven Decision Model for Combined Sewer Overflow Management using the Low-Impact Development Rapid Assessment Method

Noah Saber-Freedman

A Thesis
In the Department
of
Mechanical and Industrial Engineering

Presented in Partial Fulfillment of the Requirements
For the Degree of
Master of Applied Science (Industrial Engineering) at
Concordia University
Montreal, Quebec, Canada

March 2016

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Noah Saber-Freedman

Entitled: Data-Driven CSO Management Targeting with LIDRA

and submitted in partial fulfillment of the requirements for the degree of

M.A.Sc. Industrial Engineering

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Onur Kuzgunkaya Chair

Dr. Fuzhan Nasiri Examiner

Dr. Ali Akgunduz Examiner

Dr. Ketra Schmitt Supervisor

Approved by _____
Chair of Department or Graduate Program Director

_____
Dean of Faculty

Date April 13, 2016

ABSTRACT

A Data-Driven Decision Model
for Combined Sewer Overflow Management
using the Low-Impact Development Rapid Assessment Method

Noah Saber-Freedman

Mitigating the frequency and severity of Combined Sewer Overflows (CSOs) represents a significant engineering and economic challenge in urban stormwater management (SWM). Low-Impact Development (LID) methods are a decentralized approach for dealing with this challenge. Current methods for estimating CSO mitigation efficacy and informing choices about infrastructure solutions are typically based on simulation of the storm sewer network for municipalities. The recent public availability of rainfall and CSO data represents a potential opportunity to improve the quality of these estimates, as well as reducing the time it takes to generate them.

A novel decision support model is presented which solves a Mixed Integer Program (MIP) formulation of the Low-Impact Development Rapid Assessment (LIDRA) method algorithmically to identify priority catchment areas for intervention with LID infrastructure, as well as the optimal extent of investment, subject to different budgetary constraints. The reliability of the model is improved by means of a Monte Carlo simulation.

This method is demonstrated with an open dataset from the city of Spokane, Washington, but it is generalizable to other municipalities where storm and CSO data is available.

For my father, my teacher:

Dr. Aaron Jaan Saber,

who is now and always very deeply missed.

Table of Contents

Lists

List of Figures

List of Tables

List of Equations

List of Variables

A	=	Area of Watershed [L^2];
A_i	=	The area of the i^{th} shed [L^2];
A_{bed}	=	Area of Rain Garden Bed [L^2];
C	=	Runoff Coefficient, variable with land use [unitless];
C_p	=	Compound Runoff Coefficient, with LID implementation [unitless];
C_{ex}	=	Extant Runoff Coefficient, prior to LID implementation [unitless];
D	=	Storm Depth [T];
d_{bed}	=	Depth of Rain Garden Bed [L];
d_{ij}	=	The depth to CSO for the j^{th} CSO in the i^{th} shed [L];
d'_{ij}	=	The simulated depth-to-CSO for the j^{th} event in the i^{th} shed [L];
d_t	=	Cumulative depth of rainfall preceding a CSO [L];
$d_{t,Cex}$	=	Minimum cumulative depth of rainfall preceding a CSO, with LID implementation [L];
$d_{t,Cp}$	=	Minimum cumulative depth of rainfall preceding a CSO, prior to LID implementation [L];
i	=	Storm Intensity [D/T];
k_i	=	The cost of implementing the solution in the i^{th} shed [$];
K	=	The total budget of the project [$];
κ	=	The cost per unit volume of stormwater storage [$/$L^3$];
m_i	=	The slope of the linear regression between d_t and v for the j^{th} event in the i^{th} shed [L^2];
$N()$	=	The normal distribution function.
T	=	Storm Duration [T];
t	=	Time to Storm [T];

t_c = Equilibrium time for rainfall occurring at the most remote portion of the basin to contribute flow at the outlet [T];

t_{CSO} = Date and Time of CSO Event [T];

t_{storm} = Date and Time of Storm Event [T];

Q_p = Peak Flow [L^3/T];

R = Resolution of the LIDRA model [unitless];

v_{ij} = The volume of the j^{th} CSO in the i^{th} shed [L^3];

v'_{ij} = The simulated volume of the j^{th} event in the i^{th} shed [L^3];

$V_{storage}$ = Volumetric capacity of LID intervention [L^3];

x_i = Binary decision variable indicating if i^{th} CSO-shed is in the solution set;

θ_{ij} = Binary variable indicating if the j^{th} CSO in the i^{th} shed is below the minimum depth to CSO;

δ_i = The minimum depth to CSO in the i^{th} shed [L];

σ_{vi} = The standard deviation of the CSO volumes in the i^{th} shed [L^3];

ϕ = Void Ratio in Rain Garden Bed Soil Medium [unitless].

List of Abbreviations

CS	Combined Sewer
CSO	Combined Sewer Overflow
GLSA	Greedy Local Search Algorithm
IP	Integer Program
LID	Low-Impact Development
LIDRA	Low-Impact Development Rapid Assessment
MIP	Mixed Integer Program
SWM	Stormwater Management
WW	Wastewater
WWTP	Wastewater Treatment Plant

1. INTRODUCTION

1.1 Background

Securing investment in stormwater sewers is challenging for municipalities, and it is particularly difficult to make decisions about which regions of a given municipality should be prioritized for intervention in the absence of sound data. Further, even if abundant data is available, care must still be taken to operationalize said data in the context of informing decisions about municipal infrastructure spending.

Despite their critical importance, sewers are largely invisible to the public, making renovations and improvements a less-than-popular prospect for municipal governments. Indeed, regardless of their necessity, the financial and environmental costs of renovating sewer infrastructure can be seen as unpalatable – or even unnecessarily hazardous – as in the recent case of Montreal's renovation of the southeast interceptor (CBC News, 2015). This unpopularity exacerbates the real environmental and public health costs associated with inefficient urban stormwater and wastewater infrastructure by making discussion of its construction, repair, or improvement taboo. Consequently, this places an increase in importance on reducing infrastructure spending and increasing the efficiency of any notional stormwater management system.

1.2 Problem Statement

Combined Sewer Overflow (CSO) is a frequent issue for Combined Sewer (CS) systems. When a rainfall arrives that exceeds the design capacity for the sewer system, discharges of effluent into a receiving water are a result. These discharges, which contain elevated levels of

contaminants and pathogens when compared to the discharges from dedicated Stormwater Sewers (SS), represent an increased health and environmental risk to municipalities and ecosystems. Figure 1 shows a simple representation of the difference between these two systems:

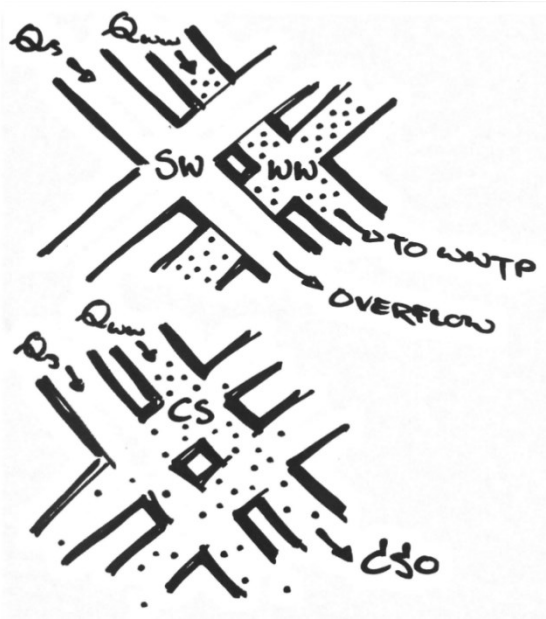


Figure 1: Comparison of Separate and Combined Sewer Systems

This thesis attempts to answer two questions of interest to municipal planning professionals who deal with Combined Sewer (CS) systems:

- **"Where should municipal sewer infrastructure funds be disbursed?"**

  and,

- **"What is the extent of the infrastructure spending that will achieve the best results?"**

Answering these two questions is important for both the efficient management of the health and environmental risks due to CSOs, as well as making the best use of limited municipal infrastructure budgets. These questions are answered by selecting Low-Impact Development (LID) as the type of Stormwater Management (SWM) solution, as the LID approach is a decentralized, cost-effective method for SWM which presents some major advantages over more conventional SWM approaches.

1.3 Thesis Contribution

This thesis demonstrates a method that, given data on the time of onset and volumes of CSOs for a large number of catchment areas under a municipal jurisdiction, and data on the time of onset and depth of storms covering those catchment areas, will identify the catchment area which will respond most favourably to a set infrastructure investment. In addition, the method identifies the optimal extent of investment within that budget cap to minimize waste.

The method works by first solving an Integer Programming (IP) problem to compare storms and CSOs in order to determine a critical parameter for each shed, the minimum depth to CSO. Next, these solutions are used as inputs into a Mixed Integer Programming (MIP) problem which describes the response of a CS system to rainfall, with the output being the maximization of the effects of LID intervention given cost and space constraints. The reliability of the model is tested with Monte Carlo methods and the introduction of random variates as input data.

Such an approach presents two immediate advantages. One of the advantages of this approach is that, unlike most of the sewer infrastructure risk management approaches currently in use, it does not require the surveying and simulation of a complete municipal sewer system,

but only needs information about the stormwater inputs and CSO volume outputs. Consequently, this means that the method can be applied to new municipalities with relative rapidity, making it useful for feasibility studies.

Another advantage is in scope: several decision support models currently in use are able to identify the ideal mix of CSO mitigation solutions for a given catchment area, but there exist very few models that identify the preferred catchment area for intervention from a set of catchment areas are under consideration. By taking a broader view of where infrastructure intervention should be done, rather than which interventions should be done, this method identifies otherwise hidden opportunities for municipalities to maximize the impact of their potentially-limited infrastructure budgets.

2. LITERATURE REVIEW

2.1 Combined Sewer Overflows

CS systems are a municipal sewer system design approach which integrates the domestic wastewater (WW) system with the SWM system for a given municipality. It is an approach which, at least initially, reduces infrastructure costs by avoiding the construction of two disparate piping networks (Field, Sullivan, & Tafuri, 2004), but the relative benefits of the two design approaches are still under debate (Toffol, Engelhard, & Rauch, 2007). With CS systems, there is a risk that when a storm occurs which exceeds their design capacity, a CSO event can follow, typically at a designed outfall into a receiving water body.

Much study has been done on the subject of environmental loading due to CSOs. Urban stormwater runoff typically contains significant concentrations of priority contaminants commonly found in raw sewage, as measured by the Five-Day Biochemical Oxygen Demand (BOD_5), Chemical Oxygen Demand (COD), and Total Suspended Solids (TSS), as well as nutrients such as Ammonia Nitrogen (NH_4-N), and Phosphorus (P) (Lee & Bang, 2000) (Gasperi & al., 2008) (Gupta & Saul, 1996). The latter two substances are an issue as they can cause environmental problems such as eutrophication (Field, Sullivan, & Tafuri, 2004). Taebi and Droste have compared pollution loads in urban runoff and sanitary wastewater (Taebi & Droste, 2004), and CSO discharge can contain elevated levels of nutrients when compared to the discharge of separate sewer systems (Brombach, Weiss, & Fuchs, 2005). There is also the possibility of the presence of pathogens, such as fecal coliforms and *Giardia* (Davis & Cornwell, 2008), being passed directly to receiving waters.

More recently, substances in the category of Contaminants of Emerging Concern, which include endocrine-disrupting substances such as the estrogens from birth control pills excreted in human urine (Richardson & Ternes, 2005) (Battaglin & Kolpin, 2009) are presenting a significant issue for aquatic ecosystems. Finally, there is an increased concern about the effects of changing climate on municipal infrastructure, which may potentially exacerbate the aforementioned issues (Yazdanfar & Sharma, 2015).

In short, the environmental contaminant loading due to urbanization is detected in water quality, and this loading represents a significant infrastructure engineering challenge (Elliott & Trowsdale, 2007).

2.2 LID Systems

Traditional SWM approaches involve the use of centralized engineering solutions (Field, Sullivan, & Tafuri, 2004) but recent work has seen the rise of LID methods for dealing with stormwater. The premise of LID - also called Sustainable Urban Drainage Systems (SUDS) (Elliott & Trowsdale, 2007) - is to mimic the pre-development hydrological properties of a region (Dietz & Clausen, 2008) (Montalto, et al., 2007).

Overall, LID decreases peak discharge depth and volume, while increasing the lag time and runoff threshold when compared with more traditional SWM approaches (Hood, Clausen, & Warner, 2007). There is some evidence that some LID methods, like constructed wetlands, may remove some of the aforementioned emerging contaminants (Cahill, 2012). This suggests that there may be a potential gain in efficacy from adopting alternative urban stormwater/wastewater management approaches (Matamoros, Garcia, & Bayona, 2008). LID design typically

emphasizes distributed interventions like rainwater catchment cisterns, green roofs, permeable pavement, and rain gardens (Montalto, et al., 2007).

Rainwater catchment cisterns are storage vessels which buffer the volumetric capacity of the sewer system. They detain water on its path from the roofs of buildings to the CS system. The stored water can also be used for irrigation, or other uses (Jones & Hunt, 2009). Cisterns can provide a ready source of non-potable water for reuse, which is particularly important in regions where water must be imported (Appan, 1999).

Green roofs, also known as “vegetated roof systems”, consist of vegetation that is grown on the roofs (Cahill, 2012). Green roofs are a LID SWM approach with the advantage of transforming contaminants as well as retain stormwater (Mulligan, 2002). They also have the advantage of contributing to the thermal control of the building due to the evaporative cooling of transpiration via the vegetation (Cahill, 2012). Installing a green roof, however, can present a significant engineering challenge, as the weight of the vegetation can contribute substantially to the structural load on the building.

Permeable pavement is a LID SWM method consisting of a pervious medium above a storage reservoir (Cahill, 2012). For smaller storms, water avoids the sewer system completely, but an overflow control structure is important for avoiding ponding on the surface of the road during larger storms.

Rain gardens are a distributed stormwater detention method that reduce runoff volume and mitigate peak discharge rates (Cahill, 2012), and, like green roofs, they have the added bonus of being a bioremediation method, transforming contaminants into less harmful forms in the soil matrix (Mulligan, 2002). A typical design for a rain garden consists of a bed of planting mix,

typically 18 inches in depth and sloped on the sides, above a drainage bed of gravel or washed sand (Cahill, 2012). A riser with a domed grate and a sump is typically installed to put an upper limit on the ponding depth. Plants are grown in the soil medium in order to keep the soil in place, to conceal the water that pools in the garden, and to improve the extent of transformation of undesirable contaminants that are washed into the garden. One of the advantages of rain gardens is that they are rather pretty to look at, and this gives them a measure of flexibility as an infrastructure intervention. Rain gardens constructed as ditches with a high length-to-width ratio are frequently referred to as bioswales. A diagram of a rain garden can be seen in Figure 2 (Oregon State University, 2016). For the purposes of the method described here, rain gardens are used to estimate the cost of the project, as well as the total land surface area to be repurposed.

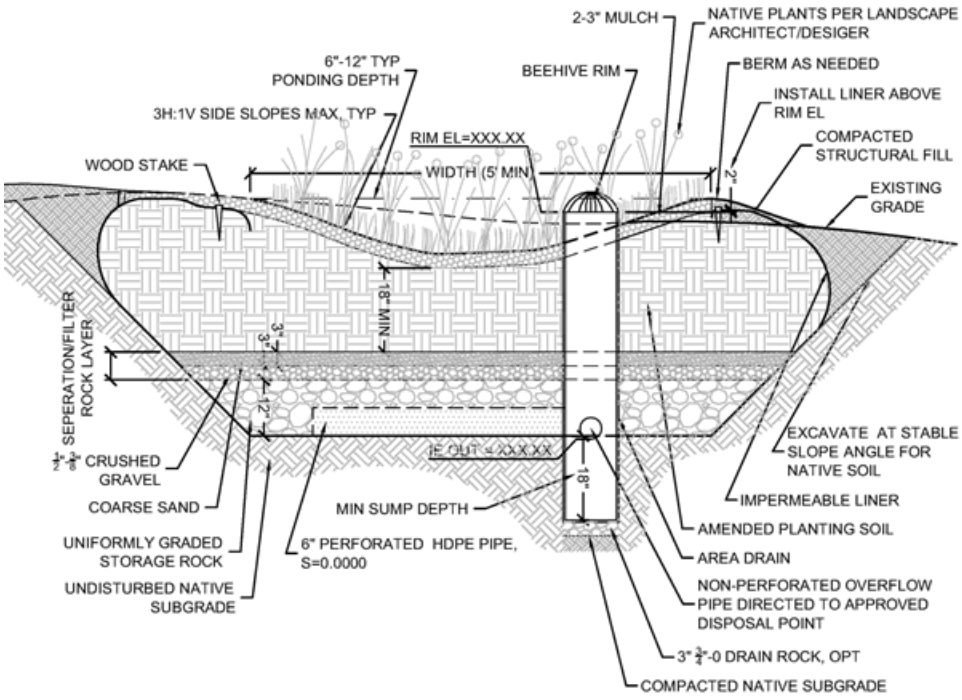


Figure 2: Design of a Rain Garden

2.3 LID Modelling

A number of modelling approaches exist for assessing the efficacy of LID approaches for a region (Yazdanfar & Sharma, 2015), including Chiew and McMahon's Model for Urban Stormwater Improvement Conceptualisation (MUSIC), Palmstrom and Walker's P8-UCM, the Probabilistic Urban Rainwater and Wastewater Reuse Simulator (PURRS) of Coombes, Haith's RUNQUAL, the Source Loading and Management Model (SLAMM) of Pitt, and the Low-Impact Development Rapid Assessment method (LIDRA). The latter method, developed by Montalto et al. (Montalto, et al., 2007), seeks to evaluate the effect of LID SWM solutions for dealing with CSOs.

LIDRA is based on the well-known Rational Method of hydrology (Bedient, Huber, & Vieux, 2008) (Yazdanfar & Sharma, 2015) (Montalto, et al., 2007), and parameterizes the effect of LID technology application as a change in the dimensionless Runoff Coefficient C, which takes a value between 0 and 1. Higher values of C indicate a reduced permeability – and, therefore, greater runoff volumes. Importantly, LIDRA's governing equations feature the parameter d_t, which corresponds to "the cumulative depth of rainfall preceding a CSO" (Montalto, et al., 2007). Typically, CSO mitigation estimates are based on a simulated sewer response to a known rainfall. This means that the sewer system for a given municipality must be modelled by a pipe flow simulation software package like EPANET (USEPA, 2015) in order to produce values for d_t from known or simulated rainfall inputs (Montalto, et al., 2007) (Toffol, Engelhard, & Rauch, 2007). Estimating d_t, as described here, is faster than modelling the entire system when storm and CSO data are abundant.

Even with the availability of abundant data, finding this critical value d_t is not always a straightforward affair; even though many municipalities keep excellent records of rainfall and CSO volumetric flow rates and durations, these records are often decoupled from one another and therefore a critical independent variable used in the LIDRA method may not be readily available. Since a definitive causal relationship between storms and CSOs is not known for each CSO in the set of events, and therefore the value d_t is not immediately available, a Greedy Local Search Algorithm (GLSA) is used to derive a value of d_t for each CSO in the record so that the minimum d_t, indicated by $d_{t,Cex}$, can then be determined for each CSO-shed.

The following approach is presented with the caveat that correlation does not indicate causation, but that correlationary data nevertheless retains some predictive power.

2.4 Costs and Benefits

Houle et al. rightly point out that the effective communication of LID cost is critical to ensuring its implementation (Houle & al., 2013). Their analysis of the cost-benefit relationship for LID methods includes volumetric storage and retention, as well as a discussion of contaminant removal for TSS, phosphorus, and nitrogen. They compare maintenance costs for LID with maintenance costs for conventional SWM and show that maintenance costs were lower for “bioretention and subsurface gravel wetland” (Houle & al., 2013), which motivates the selection of the rain garden solution as the means by which system costs are estimated in this work.

Ariratnam and MacLeod note that when it comes to big infrastructure projects, everybody wants to build and nobody wants to maintain (Ariratnam & MacLeod, 2002). They perform an economic analysis with benefit-cost ratio with the inclusion of the effect of interest, and the

expected probability of deficiency. This probability of deficiency, first introduced in a report prepared for the city of Edmonton and openly discussed in the literature, is a statistical measure. This measure has some advantages, as it makes it possible to estimate the rate at which the sewer system must be repaired, but it doesn't give good localization for where the failures are likely to occur. Wirahadikusumah et al. discuss methods for assessing sewers for rehabilitation, discussing sophisticated survey methods involving robots and ground-penetrating radar (Wirahadikusumah & al., Assessment technologies for sewer system rehabilitation, 1998), approaches which require time and skilled operators. Costs are not indicated in the aforementioned work.

Modelling sewer infrastructure can be a complex and costly affair for large networks, and gaining access to information about the condition of sewer pipes can be a significant project. Cahill presents some straightforward cost estimates for LID infrastructure design purposes (Cahill, 2012). These cost estimates are useful for estimates such as the one described in this paper. The use of rapid assessment methods like LIDRA is intended to save time and money for municipalities that have access to good-quality data about their CSO parameters. Finally, it should be reiterated that LID presents significant advantages when compared with conventional SWM methods both in terms of efficacy and cost (Houle & al., 2013).

2.5 Decision Models

Budgetary constraints on municipalities necessitate judicious CSO mitigation infrastructure spending. It therefore follows that prioritization of CSO-sheds is a useful tool for municipal infrastructure policy and planning. There exist many decision support models for prioritizing infrastructure investment for existing, conventional sewers (Halfawy, Dridi, & Baker, 2008) (Wirahadikusumah, Abraham, & Castello, Markov decision process for sewer

rehabilitation, 1999) (Fenner & Sweeting, 1999), but scholarly literature describes comparatively few decision support tools available for LID infrastructure.

Ahammed, Hewa, and Argue discuss a model (Ahammed, Hewa, & Argue, 2012) comparing LID technologies with Analytic Hierarchy Process (Saaty, 1990), a decision model based on expert judgment. This approach, however, doesn't make recommendations about specific geographic regions for infrastructure investment. Nielson and Turney presented a conference paper (Nielson & Turney, 2010) discussing optimization analyses for CSO control with green infrastructure in the city of Indianapolis, IN. Again, this method focusses comparing different solutions within one watershed. Finally, Sample et al. discuss decision support systems for SWM from a Geographic Information System (GIS) standpoint (Sample & al., 2001), and uses a linear program to evaluate LID SWM for a hypothetical study area. Lee, Heaney, and Lai perform a similar analysis (Lee, Heaney, & Lai, 2016).

There seems to be a need for methods that rapidly compare CSO-sheds for LID infrastructure investment priority instead of simply comparing different LID technologies within a single catchment area, use real data instead of simulation modelling or expert opinion as decision variables, and treat real regions of interest instead of hypothetical study areas. It is hoped that the approach described in this thesis serves to fill that gap.

2.6 Algorithmic Solutions of IP and MIP Problems

There are many different forms that IP and MIP can take, and some of the forms are common and well-studied. The maximum cardinality matching problem, for example, is a

common IP problem where the solution is a matching, M, on a graph, G, consisting of vertices V, and edges E (Wolsey, 1998). Examples of M are shown in Figure 3.

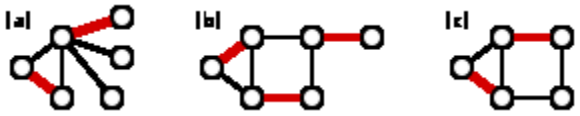


Figure 3: Three Examples of a Matching M on a Graph G

An interesting property of the maximum cardinality matching problem is that it – and its inverse, the minimum cardinality matching problem – is a problem in the set $\mathcal{P}$, the set of problems for which an algorithm exists that will solve the problem to optimality in polynomial time (Wolsey, 1998). Some graphs can be divided into two sets of vertices which do not have edges between them. These types of graphs are called bipartite graphs, and are also well-studied (Glover, 1967). Graphs with more than one partition can also be constructed.

Nonlinear MIPs, however, are in the set $\mathcal{NP}$, for which algorithms that solve the problem to optimality in polynomial time have not been shown to exist (Wolsey, 1998). Consequently, for problems in the set $\mathcal{NP}$, a solution can be found algorithmically, but there is no way to know if that solution is a global optimum.

There are several types of algorithms for solving IP and MIP problems (Wolsey, 1998). With a judicious selection of algorithm – informed by the structure of the problem – very strong solutions can be found in a relatively short period of time. For problems in the set $\mathcal{P}$, such as the maximum cardinality matching problem on a bipartite graph, there exist algorithms that can find the global optimum in polynomial time by examining local optimality criteria. For problems not in

$\mathcal{P}$, however, finding a global optimum is not a certainty, and creative methods must be used to determine a solution.

Greedy Local Search Algorithms (GLSA) start with an incumbent solution, typically the trivial solution:

$$x_i = 0 : \forall x \in X$$

where:

x = An integer decision variable.

Equation 1: The Trivial Solution to an Integer Programming Problem

and proceed by checking to see whether an adjacent solution in the solution space improves the value of the objective function over the incumbent. If such a solution exists, the new solution is accepted as the new candidate solution for the new optimal solution (Selman & Kautz, 1993). The algorithm checks again to see if an adjacent solution exists that improves the objective function, and so on. Left to run for an arbitrarily long period of time, the GLSA will eventually stop when there exist no adjacent solutions that improve the value of the objective function over the incumbent. If the problem is in the set $\mathcal{P}$, then the solution is a global optimum (Wolsey, 1998).

MIPs of the kind featured in this thesis are, as stated, not in $\mathcal{P}$. Therefore, different algorithms must be used to determine feasible solutions and optima. One such approach, which can be used for MIPs of is to simply enumerate a subset of feasible solutions and test each one (Revelle, Whitlach, & Wright, 2004). This is time consuming, but feasible when the datasets are not too large. Consider that, the data spans S CSO events in T CSO-sheds. Further, if the

continuous variable C_p is discretized by breaking it up into R portions, then the number of possible solutions is, at, most,

$$|\tilde{Q}| = R * \sum_{i}^{T} S_i$$

Where:

$|\tilde{Q}|$ = The cardinality of the candidate solution set $\tilde{Q}$ that solves the LIDRA MIP;

R = The resolution of the search;

$|S|$ = The number of CSO events in the i^{th} shed;

T = The number of CSO sheds.

Equation 2: The Cardinality of a Candidate Solution Set

In the case of the data used in this thesis, the resolution is 100, and the total number of CSO events is 3513, so the upper bound on the cardinality of the set of candidate solutions is 351300. This is a number of variables that most modern personal computers to can manage quickly, making this a useful approach for many municipalities where information technology budgets are limited.

Finally, it should be noted that although the IP and MIP problems and their solution algorithms are discussed separately, they need not be constructed or solved separately. Splitting the matching problem and the decision model into two problems was done in the interests of modularity.

2.7 Monte Carlo Methods

If the data is subject, as it is in the case of this thesis, to some kind of stochastic process, then it is possible that decisions made on the basis of the information to date will not be optimal for future scenarios. This attempt to reduce the sensitivity of the model to outliers in the CSO and storm input data is what makes it robust. Since the LIDRA MIP is solved by an exhaustive search of a discretized candidate solution space, care must be taken to ensure that the model is not sensitive to outliers in the data, as oversensitivity to outliers in the storm and CSO inputs can cause one single freak event to dominate the selection of the model outputs. Such a sensitivity mean the model outputs no longer represent the 'true' best option. Improving the ability of the model to withstand the effect of outliers, and in turn, to provide better results, is accomplished by applying Monte Carlo methods, also known as Monte Carlo simulation.

Monte Carlo methods compare results based on repeated random sampling of data, and as such, give useful insight when the data or the model has some inherent uncertainty (Kammen & Hassenzahl, 1999). In cases where the data has an element of uncertainty associated with it – in this case, the weather, or the production of wastewater in households – it can be very useful to have information about the shape of the distribution of possible outputs for the decision models that depend on that data. By "playing the game" (Clemen & Reilly, 2001) multiple times, such a distribution can be developed.

The value of applying a Monte Carlo method can be the difference between making a decision based on signal and noise.

3. EXPLORATORY DATA ANALYSIS

In 2008, the city of Spokane, WA, published the Spokane Regional Stormwater Manual (SRSM). The document states that it is intended to "protect water quality, prevent adverse impacts from flooding, and control stormwater runoff to levels equivalent to those that occurred prior to development" (City of Spokane, City of Spokane Valley, Spokane County, 2008). The SRSM goes on to describe design guidelines for several LID CSO-reduction methods specially tailored to the needs of Spokane. In addition to this document, Spokane has done an excellent job of cataloguing their rainfall and CSO data since 2001, resulting in a rich dataset of 763 CSO-producing storms spanning over 13 years as of this writing.

The data was initially made available in excel spreadsheet format and in comma-separated value format. The availability of a policy and engineering design document, in addition to the excellent data on storms and CSOs, make Spokane an ideal candidate for developing methods of the kind described in the present report. The storm data consists of storm depth measurements from rain gauges for storms where CSOs were known to occur, and the CSO data contains measurements of total volume (in gallons), volumetric flow rate (in gallons per minute), and duration (in minutes) of the various CSOs for the 27 CSO-sheds in the area. The geographic delineations of these CSO-sheds can be seen in Figure 6. It should be noted that two CSO-sheds with less than 5 recorded CSO events were excluded, bringing the total number of CSO-sheds in the record to 24 and the total number of CSO events in the record to 3325. Finally, including the CSO-shed Area in the computations lowered the number of sheds to 21 and the number of CSO events to 3198.

3.1 Storm Data

The storm dataset contains a set of 763 CSO-causing storms recorded between December 21, 2001 at 11:03 and December 19, 2012 at 15:09. The maximum recording from any rain gauge for any storm was 83.52 inches – a very large value, which may be present in the data due to mismeasurement or a typographical error in the data entry. Average (across rain gauges) depth of CSO-causing rainfall had a minimum value of 0.020 inches, and a maximum of 7.593 in., with a mean value of 0.2677 in. and a standard deviation of 4.604 in.

It is important to note that the markedly low minimum value of the CSO-causing storms – as low as 0.01in. in many of the cases – could be due to CSOs triggered not by precipitation but by the spring melt. This case is reflected in the data on precipitation rates: The minimum rate of precipitation is recorded at 0.0001 hundredths of an inch per hour, with a maximum of 151.85 in./100*hr. and a mean precipitation rate of 0.2135 in./100*hr. The corresponding standard deviation for precipitation rate was computed to be 5.497 in./100*hr.

The duration of CSO-causing storms had a mean value of 2.526 days, spanning a range from 1.001 days to 20.170 days, with a standard deviation of 1.907 days. Density histograms of the base-ten logarithm of storm depth, duration, and intensity are indicated in Figures 4, 5, and 6, respectively.

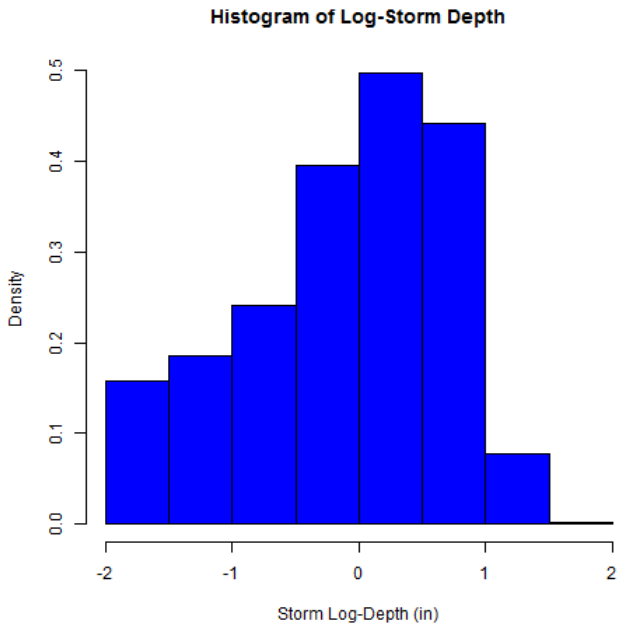


Figure 4: Histogram of Storm Depth (Log-10)

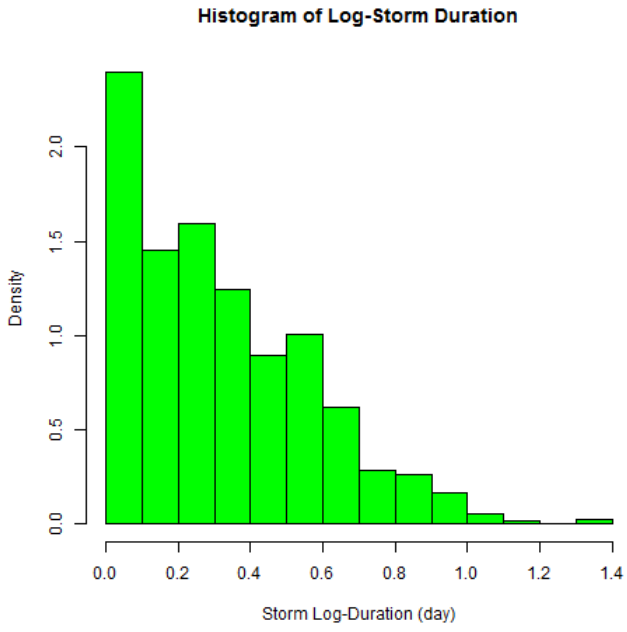


Figure 5: Histogram of Storm Duration (Log-10)

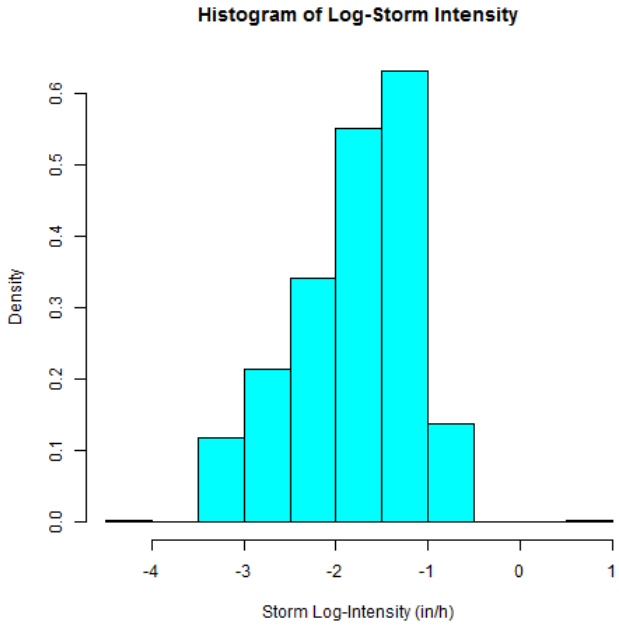


Figure 6: Histogram of Storm Intensity (Log-10)

3.2 CSO Data

Table 1 indicates the mean, standard deviation, and maximum values of CSO flows for each CSO shed. Note that for CSO-shed #20 (CSO20 in the tables), not enough data was available to generate a useful summary. In these cases, the R script which produced these summary tables produced non-numeric outputs. These non-numeric outputs have been replaced by a dash character in the tables.

Table 1: CSO Summary Statistics

		Mean of Recorded Values			Standard Deviation of Recorded Values			Maximum of Recorded Values		
CSO Shed	Count	Total Flow (gal)	Duration (min)	Flow Rate (gpm)	Total Flow (gal)	Duration (min)	Flow Rate (gpm)	Total Flow (gal)	Duration (min)	Flow Rate (gpm)
CSO02	18	2334	125	35.37	6321	134.3	97.88	27159	540	418
CSO03C	22	6088	167.4	41.61	8061	123.1	47.28	33757	480	174
CSO06	299	166400	280.5	680.4	270400	424.9	939.9	1779061	4500	6330
CSO07	125	27480	151	298	53930	168.1	591.1	443561	940	4470
CSO10	109	13570	135.8	105.6	25760	136.3	208.1	113746	725	1250
CSO12	324	122800	272.3	533.7	227100	349.7	1051	2242467	2220	10900
CSO14	152	7887	379	33.1	18780	505.6	100.3	130749	2990	811

CSO15	96	19950	151.1	120.5	43600	177	203.8	307765	1025	1310
CSO16B	56	31960	201.7	189.6	55360	439.6	239.6	311780	3030	1030
CSO19	5	134.7	363.3	1.823	59.18	288.6	2.82	203	595	5.08
CSO20	2	77740	45	1730	-	-	-	77739	45	1730
CSO22B	15	10340	35.77	486.8	28520	28.49	1427	104476	100	5220
CSO23	198	67380	157.7	383.8	99070	197.3	334	618655	1755	1440
CSO24A	226	377700	290.9	1039	847200	382.2	1749	7875001	2930	13100
CSO24B	114	8849	344.9	59.38	31860	660.1	170	238168	3755	1360
CSO25	202	20550	110.5	236.1	37430	132.4	388.3	214474	795	3350
CSO26	280	658600	185.1	2973	1118000	225.3	2885	8933946	1710	20700
CSO33A	89	4212	112.5	45.73	8993	132	56.33	76628	595	301
CSO33B	78	882400	58.05	12710	1295000	57.37	13000	8159760	345	58400
CSO33C	79	6491	96.92	77.1	17740	142.4	130.3	142477	810	756
CSO33D	268	18260	386	68.4	42920	725.3	79.52	515733	6885	436
CSO34	217	662300	189.1	2798	1180000	242.5	3032	8444710	1985	15500
CSO38	114	8084	351.4	28.8	34800	494.2	69.57	272392	3710	377
CSO39	45	13240	178.1	187.8	22940	276.1	219.5	123636	1310	727
CSO40	102	8338	199.4	51.74	22280	292	103.4	142766	1675	732
CSO41	131	29250	160.3	209	69840	383.1	298.6	544853	4085	2240
CSO42	11	183	114.4	6.267	36.17	144.6	1.981	221	455	7.45

3.3 CSO-Shed Area Data

Measurements for the different CSO-sheds were made by importing a PDF document of the CSO-shed delineation for the city of Spokane (Figure 2) into AutoCAD, fixing the scale to match the scale of the drawing, tracing the CSO-shed boundaries with the line tool, and measuring the areas. This gives the areas in Table 2.

Table 2: CSO-Shed Areas

	x	Area (ha.)
1	CSO02	26.05
2	CSO06	170.39
3	CSO07	42.70
4	CSO10	19.53
5	CSO12	126.38
6	CSO14	25.11
7	CSO15	43.72
8	CSO23	57.91
9	CSO24A	658.09
10	CSO24B	25.21
11	CSO25	7.44
12	CSO26	215.64
13	CSO33A	23.61

14	CSO33B	392.44
15	CSO33C	5.53
16	CSO33D	17.26
17	CSO34	687.43
18	CSO38	25.34
19	CSO39	17.84
20	CSO40	19.93
21	CSO41	31.43
22	CSO42	10.91

The SRSM gives the value of C_{ex} as 0.36 for the entire system (City of Spokane, City of Spokane Valley, Spokane County, 2008).

A map of the delineations of the CSO sheds in Spokane is visible in Figure 7:

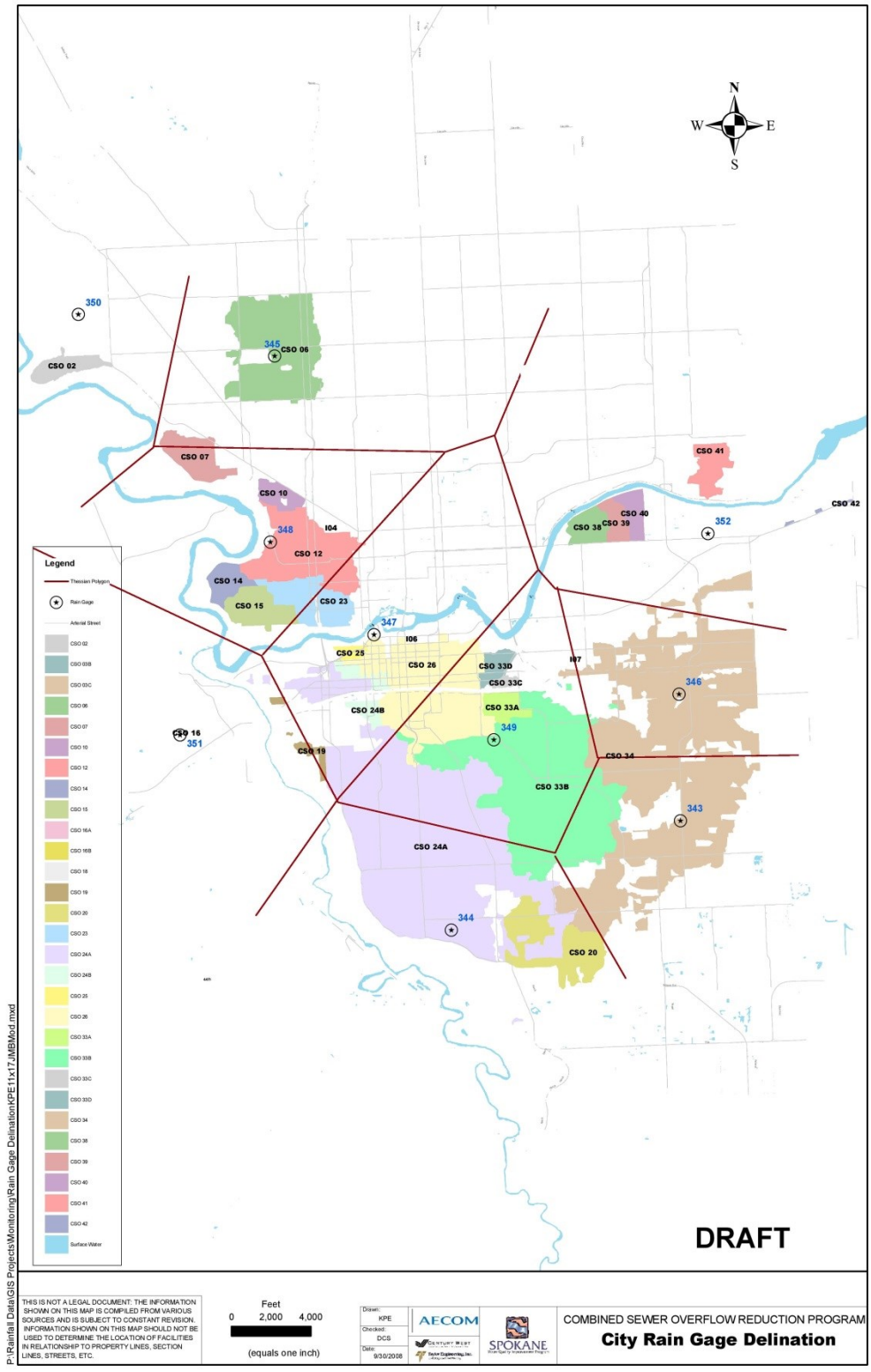


Figure 7: Rain Gauge and CSO-Shed Delineation for Spokane, WA.

3.4 Budget Data

Finally, to frame the scale of the difficulty in making decisions about how sewer infrastructure expenditure is to be allocated, Figure 8 indicates the utilities division of the 2016 City of Spokane Program Budget (City of Spokane, 2016). The total budgetary allocation for Environmental Projects, Wastewater Capital Projects, Wastewater Collections and Maintenance, Wastewater Management Riverside Park Water Reclamation Facility, Water & Hydroelectrical Services, Water/Wastewater Debt Service Fund, and the Water/Wastewater Revenue Bond Fund amounts to a total of $165.375M.

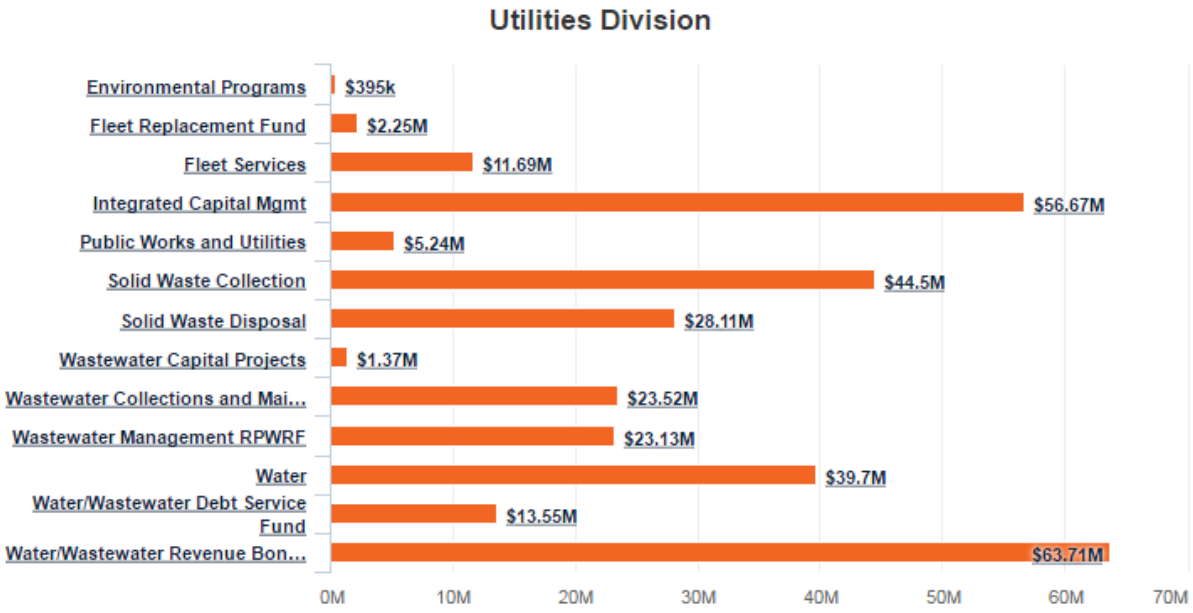


Figure 8: Spokane Municipal Budget, 2016 - Utilities Division

The research question is hereby restated: given the data described in this section, where should infrastructure planning professionals spend their LID investment dollars, and how many of those dollars should they spend?

4. METHODS

The prioritization of CSO-sheds for LID targeting is based on LIDRA. In order to apply LIDRA, however, a missing parameter must first be determined. This is done by applying a GLSA to the storm and CSO datasets. All computation was done in the R scripting language. The methods section begins by describing LIDRA, and follows by describing the GLSA used to generate its inputs.

4.1 Low-Impact Development Rapid Assessment (LIDRA)

LIDRA is based on the rational method of hydrology, a widely-used approach for modelling runoff. It is based on the following equation, which is written for a specific watershed and storm:

$$Q_p = CiA$$

Where:

Q_p = Peak Flow [L^3/T];
C = Runoff Coefficient, variable with land use [unitless];
i = Intensity of rainfall of chosen frequency for a duration equal to time of concentration t_c [L/T];
t_c = Equilibrium time for rainfall occurring at the most remote portion of the basin to contribute flow at the outlet [T];
A = Area of Watershed [L^2].

Equation 3: The Rational Method for Predicting Surface Flow from a Rainfall

This equation represents volumetric flow through a control volume with the input i entering at the top, the output Q exiting through the sides. For the sake of completeness, an additional

complementary output Q' can be described, exiting through the bottom, and having the properties:

$$Q' = (1 - C)iA$$

Equation 4: The Rational Method (Complement Groundwater Flow)

A mass flow diagram for computing Q via the rational method is given in Figure 9:

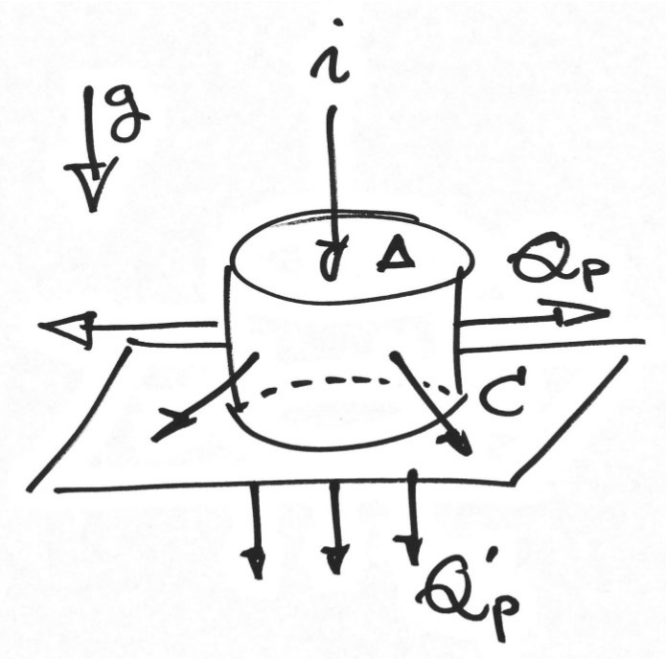


Figure 9: The Rational Method

The parameter of interest here is C, which represents the relative quantity of water flowing over the ground (Bedient, Huber, & Vieux, 2008), and has a value between 0 and 1, by definition. LIDRA parameterizes the extent of a change in LID stormwater management applications as a change in the value C, where "C_{ex} is the dimensionless runoff coefficient corresponding to the existing level of imperviousness of the CSO-shed", and "C_p is the composite runoff coefficient corresponding to a potential level of LID implementation in the

sewershed". An engineer should hope that C_p winds up being less than C_{ex}. Figure 10 is a graphical representation of the effect of a change in C on Q:

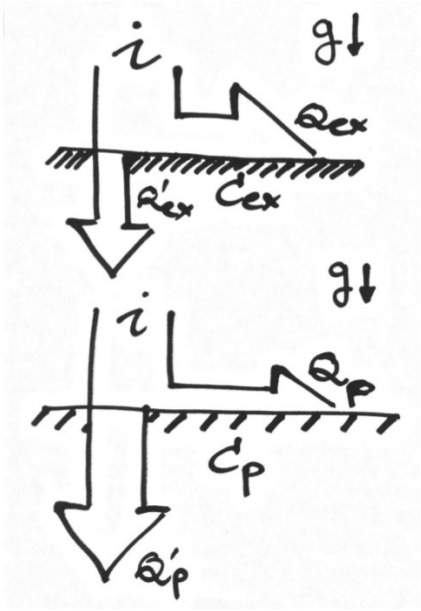


Figure 10: The Effect of a Change in the Runoff Coefficient C on Runoff Q

Next, LIDRA substitutes the parameter i for a ratio of depth to time, as follows:

$$Q_t = C_{ex} \frac{d_{t,Cex}}{t} A$$

Where:

Q_t = Peak runoff flow rate caused by rainfall of duration t and depth $d_{t,Cex}$ [L^3/T];

$d_{t,Cex}$ = Cumulative depth of rainfall preceding a CSO [L].

Equation 5: LIDRA Equation for Flow in the Initial Condition

Isolating $d_{t,Cex}$ yields the following expression:

$$d_{t,Cex} = \frac{Q_t t}{AC_{ex}}$$

Equation 6: LIDRA Equation for Depth in the Initial Condition

And therefore:

$$d_{t,Cp} = \frac{Q_t t}{AC_p}$$

Equation 7: LIDRA Equation for Depth in the Final Condition

Finally, LIDRA expresses $d_{t,Cp}$ as a function of the ratio of values of C:

$$d_{t,Cp} = \frac{C_{ex}}{C_p} d_{t,Cex}$$

Equation 8: Modified LIDRA Equation for Depth in the Final Condition

It is Equation 6 which is of use here, as once C_{ex} and $d_{t,Cex}$ are known, then $d_{t,Cp}$ can be computed simply by varying the value C_p.

For every existing Storm-CSO pair, a value of $d_{t,Cex}$ can be determined. In the method described in this paper, however, the parameter of greatest interest is the minimum $d_{t,Cex}$ value in each CSO-shed. The intention is to observe how d_t rises when C decreases, and then to observe how many CSOs with the property:

$$d_t < d_{t,Cp}$$

Equation 9: CSO Mitigation Condition

are excluded from the record. This estimated reduction in CSO events Δf corresponds to an estimated reduction in the CSO volume ΔV_{CSO}, and this ΔV_{CSO} reduction value is of the greatest interest to urban SWM professionals (Cahill, 2012).

That said, before LIDRA can be implemented, the parameter $d_{t,Cex}$ must be determined for each CSO-shed in Spokane. This is described in the following sections.

4.2 Formulation of the Time-To-CSO IP

The parameter $d_{t,Cex}$ is not measured directly during storm or CSO events, and must be determined in some way. This was accomplished using an Integer Programming (IP) problem, which matches storms with the CSO events that preceded them. The IP takes the form of the minimum cardinality matching problem on a bipartite digraph.

Let the graph $G(V_c, V_s, E_t)$ denote a bipartite digraph, where the vertices V_c are individual CSO events, the vertices V_s are individual storms, and E_t are the edges between them, weighted by t, the time between events. The graph has the particular property that no edges E_t exist between CSOs and storms that occur afterwards. A solution of this IP is a matching, M, on the set of edges E_t that represent links between the CSOs and the storms which occur immediately prior.

$$minimize: Y = \sum_{i,j} t_{ij} x_{ij}$$

$$subject\ to:$$

$$\sum_i x_{ij} = 1 : \forall i \in V_c$$

$$x_{ij} \in \mathbb{B}$$

Where:

t_{ij} = The time between the i[th] CSO and the j[th] storm [T];

x_{ij} = Binary variable indicating if the edge $e_{ij} \in E_t$, from the i[th] CSO to the j[th] storm is in the solution set M;

Figure 10 shows a bipartite digraph for which the IP is formulated:

Equation 10: The Time-to-CSO IP

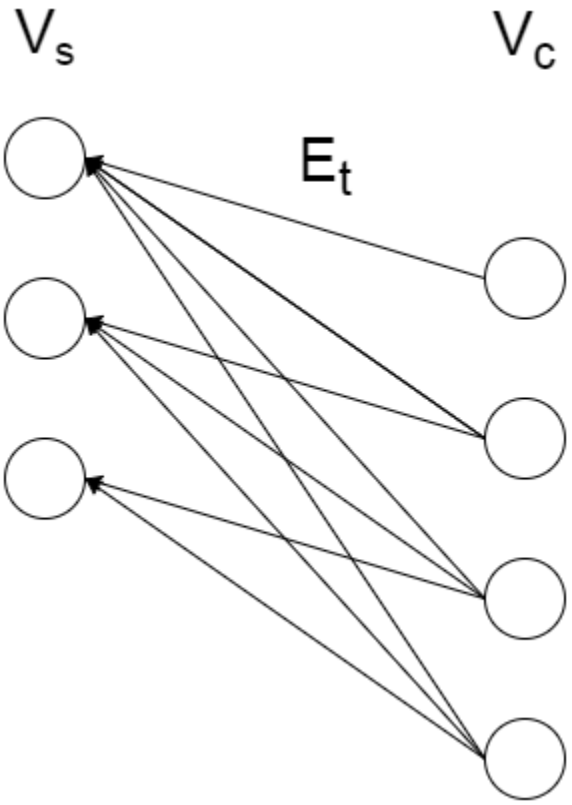


Figure 11: A Bipartite Graph

Matchings have been shown to be part of the set of solutions P, for which a global optimum can be found in polynomial time. Note that not every storm need cause a CSO, only that each CSO must have a storm. This is exceptionally good news, as it means that a Greedy Local-Search Algorithm (GLSA) will find a global optimum for input. The IP is all-but-trivial, as the matching M will consist of only those edges with an end in V_{ci} where t_{ij} is minimum with respect to all the other edges with an end on that same vertex. In fact, the only real task here is defining the edges on the graph, as, for a well-defined bipartite digraph, where the vertices in the problem space lead backwards in time from CSOs to storms, any constraint indicating that a CSO cannot be caused by a storm that came after it is redundant. Once the time between CSOs and their preceding storms has been determined, computing the depth to CSO is a straightforward operation, as will be shown.

4.3 Algorithmic Solution of the Time-to-CSO IP

As mentioned, determination of the parameter d_t is not always a straightforward affair. Here, an algorithm is presented which pairs CSOs with storms and computes d_t. First, some basic assumptions about the nature of the data are discussed, then the GLSA is introduced.

The guiding assumptions behind the attempt to pair CSOs with storms are: the start time for a storm must precede the start time for a CSO, all storms were large enough to cover the entirety of the CSO-shed where measurements were taken, and that the rainfall was constant over the duration of the storm. Taking these assumptions in hand results in a simplified model, but the results still yield usable values for d_t.

Generally speaking, greedy algorithms are algorithms that “always [take] the best immediate, or local, solution while finding an answer. Greedy algorithms find the overall, or

globally, optimal solution for some optimization problems, but may find less-than-optimal solutions for some instances of other problems." (Black, 2005). In this case, the GLSA is implemented as follows:

First, as discussed in Section 3.2, each CSO is paired with the storm immediately prior. Once each storm a storm is paired with each CSO event, the parameters for the LIDRA method (Montalto, et al., 2007) are computed as follows: time to CSO t is computed as the difference between the time of storm onset and the time that the CSO began.

$$t = t_{CSO} - t_{storm}$$

Where:

t = Time to CSO [T];

t_{CSO} = Date and Time of CSO [T];

t_{storm} = Date and Time of Storm [T].

Equation 11: Time to CSO

Next, the depth to CSO d_t is computed by linear interpolation as follows:

$$d_t = t\frac{D}{T} = ti$$

Where:

T = Storm Duration [T];

D = Storm Depth [T];

i = Storm Intensity [L/T].

Equation 12: Computation of dt by Linear Interpolation

Finally, the lowest value of d_t is identified for each shed, and these values become the values $d_{t,Cex}$ for each shed.

A flowchart representing the algorithm can be found in Figure 12:

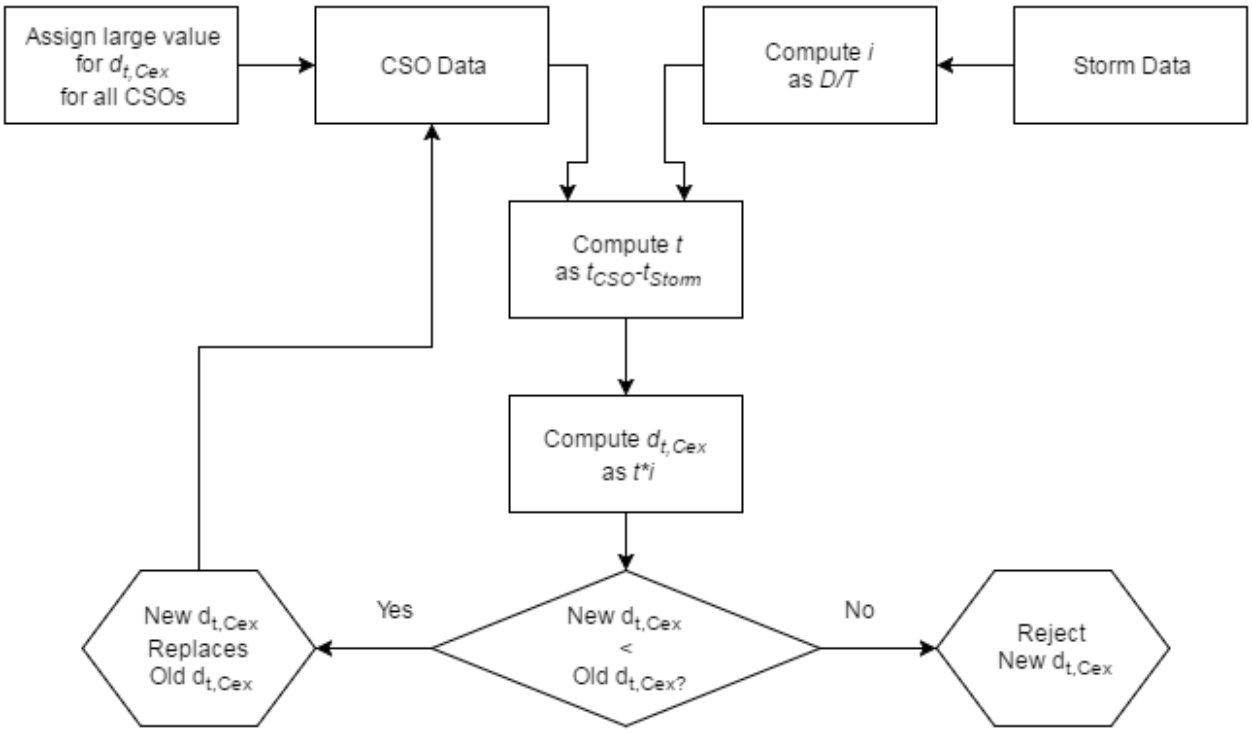


Figure 12: Flowchart of the Greedy Local Search Algorithm

4.4 Formulation of the LIDRA MIP

Finding the optimal shed and extent of investment involved formulating a MIP to maximize the amount of CSO volume reduced over the available record. This MIP was later solved algorithmically. The representation of the problem as an MIP is given as:

$$maximize\colon Z = \sum_i x_i \sum_{i,j} \theta_{ij} v_{ij}$$

$$subject\ to\colon$$

$$\sum_i x_i k_i \leq K$$

$$\sum_i x_i = 1$$

$$s * max(\theta_{ij} v_{ij}) \leq A_i$$

$$k_i = \kappa * max(\theta_{ij} v_{ij})$$

$$\theta_{ij} = 1 \; if \; d_{ij} \leq \delta_i, \; 0 \; otherwise$$

$$\delta_i = \frac{C_{ex}}{C_p} min(d_i)$$

Where:

A_i = The area of the i[th] shed [L^2];

C_{ex} = The runoff coefficient, prior to intervention [unitless];

C_p = The runoff coefficient, after intervention [unitless];

d_{ij} = The depth to CSO for the j[th] CSO in the i[th] shed [L];

K = The total budget of the project [$];

k_i = The cost of implementing the solution in the i[th] shed [$];

s = The area per unit volume of stormwater storage [L^2/L^3];

v_{ij} = The volume of the j[th] CSO in the i[th] shed [L^3];

x_i = Binary decision variable indicating if i[th] CSO-shed is in the solution set;

δ_i = The minimum depth to CSO in the i[th] shed [L];

θ_{ij} = Binary variable indicating if the j[th] CSO in the i[th] shed is below the minimum depth to CSO;

κ = The cost per unit volume of stormwater storage [$/$L^3$];

$$v, C, K, A, u, \; d, \delta \; \epsilon \; \mathbb{R}$$

$$x, \theta \; \epsilon \; \mathbb{B}$$

Equation 13: The LIDRA MIP

The optimization function is the total volume of CSO reduction. In the case that the nontrivial solution exists, x_i is 1 where i denotes the optimal CSO shed for LID investment, and the total CSO volume captured by LID infrastructure is $\sum_{i,j} \theta_{ij} v_{ij}$.

The first constraint is a cost constraint on the problem, indicating that any proposed LID intervention does not exceed a set budget, indicated by K. Here, k_i, the total cost of CSO storage in the ith shed, is given by the unit cost of CSO storage, denoted by the constant κ, and the largest measured CSO in captured by the intervention, given by $\max(\theta_{ij} v_{ij})$.

The following constraint is a constraint on the area occupied by intervention. This constraint indicates that the total area devoted to CSO storage, computed by the product of the storage coefficient s and the largest volume of CSO stored in the ith shed, cannot exceed the area of the ith CSO-shed, indicated by A_i.

θ_{ij} is a binary decision variable indicating whether the jth CSO in the ith shed is captured by LID intervention. By definition, the jth CSO in the ith shed is captured if the depth to CSO for that event, d_{ij}, is less than the minimum depth to CSO after intervention. The minimum depth to CSO for the ith shed after intervention, δ_i, is the product of the minimum depth prior to intervention, $\min(d_i)$, and the ratio of the unitless runoff coefficient prior to intervention, C_{ex}, to the unitless runoff coefficient after intervention, C_p. Note that since $C_{ex} > C_p$, $\delta_i > \min(d_i)$.

The final constraint forces the system to select only one shed for intervention. If the constraint is relaxed, the solution may include a set of sheds for infrastructure investment. This would cause the final project budget to approach its cap K, and the total volume Z would increase, but it is possible that the cost-benefit ratio of the project would decrease.

Solution of the MIP yields a single CSO shed for targeting, the maximum capacity of CSO that the new system will accommodate, the cost to install the system, and the land area devoted

to LID infrastructure. As will be shown in the results, this MIP problem is sensitive to changes in the budget, K, and it is robust with respect to the CSO volumes v_{ij} and their corresponding depths d_{ij}.

4.5 Algorithmic Solution of the LIDRA MIP

An algorithm is used to generate a solution to the MIP described in Section 3.4. The algorithm maps the problem space numerically, by computing a large number of potential solutions to the MIP, and then selecting the best one via exhaustive search of the enumerated candidate solutions. The first step is to define the resolution of the search, and subsequently to compute a range of values of C_p as fractions of C_{ex} in the following fashion:

$$C_p = \frac{1}{R} C_{ex}$$

Where:

R = The resolution (User-Defined, Unitless)
The implementation described here uses a value of 20.

Equation 14: Computation of Cp

Figure 13 is a graphical representation of the discretized solution space, where solutions exist at the points.

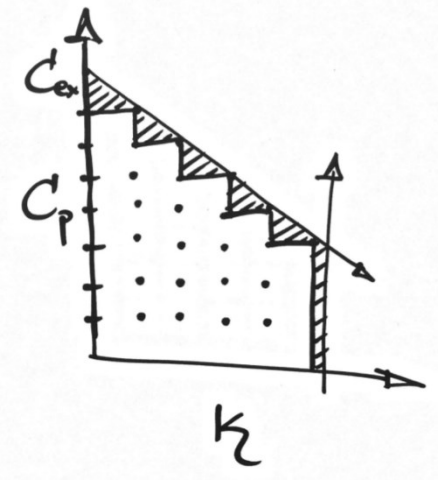


Figure 13: Discretized MIP Solution Space (Two Decision Variables Shown)

Values of $d_{t,Cp}$ are subsequently computed for each CSO-Shed using their previously-determined values of $d_{t,Cex}$ and Equation 6. CSO events with smaller depth-to-CSO values are then discarded as in Equation 7, and the difference in the number of CSO events, the total discharge volume, and the total CSO duration is recorded at each level of C_p. Table 3 indicates some values of C_p tested in the method.

Table 3: Runoff Coefficients

	%Cex	Cp
1	1	0.36
2	0.95	0.342
3	0.9	0.324
4	0.85	0.306
5	0.8	0.288
6	0.75	0.27
7	0.7	0.252
8	0.65	0.234
9	0.6	0.216
10	0.55	0.198
11	0.5	0.18
12	0.45	0.162
13	0.4	0.144
14	0.35	0.126

15	0.3	0.108
16	0.25	0.09
17	0.2	0.072
18	0.15	0.054
19	0.1	0.036
20	0.05	0.018

Cost and design parameters for rain gardens were used to estimate the cost of infrastructure intervention.

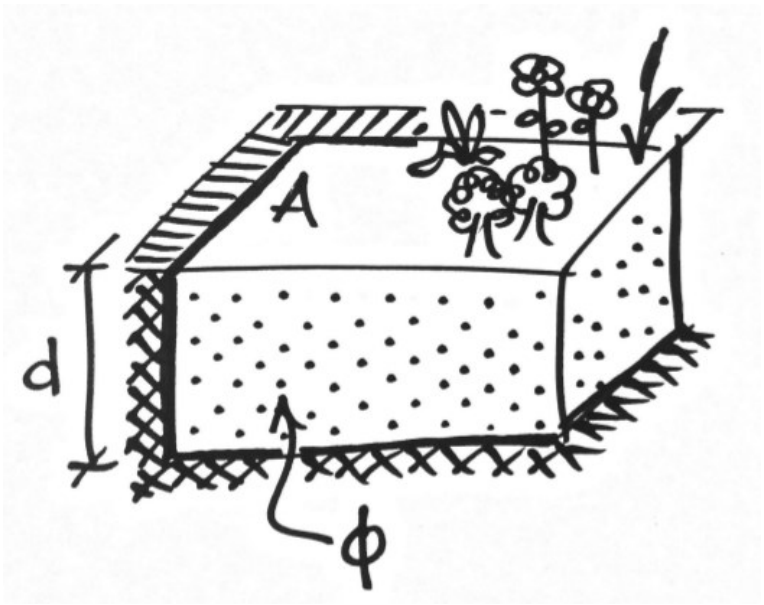


Figure 14: Idealized Rain Garden

Cahill gives a cost estimate for a rain garden between $12.50 and $17.50 per square foot (Cahill, 2012), so the mean value of $15/sq.ft was used for these computations. Cahill also recommends that rain gardens be dug to a depth of between 2 feet and 3 feet, so a mean value of 2.5 feet was used for the depth. Cahill also recommends that soil and gravel medium with average void space 0.4 be used. Therefore, the cost per cubic volume of stormwater storage is estimated as:

$$V_{storage} = A_{bed} * d_{bed} * \phi$$

$$V_{storage} = 30" * 0.0254 \frac{m}{in.} * 0.4 * A_{bed}$$

$$V_{storage} = 0.3048\, m * A_{bed}$$

$$A_{bed} = \frac{V_{storage}}{0.3048} = V_{storage} * 3.281 \frac{m^2}{m^3}$$

and:

$$Cost = A_{bed} * 15 \frac{\$}{ft^2} * 10.76 \frac{ft^2}{m^2} = V_{storage} * 3.281 \frac{m^2}{m^3} * 161.46 \frac{\$}{m^2}$$

$$Cost = V_{storage} * 529.75 \frac{\$}{m^3}$$

Where:

$V_{storage}$	=	Volume of stormwater to be stored [L^3];
A_{bed}	=	Area of land to be converted to rain garden [L^2];
d_{bed}	=	Depth of rain garden [L];
ϕ	=	Soil void ratio, estimated at 0.4.

Equation 15: Derivation of Stormwater Storage Cost

An idealized rain garden is shown in figure 14.

4.6 Benefit-Cost Analysis

Finally, the benefit-cost ratio computed for each solution in the space by the following ratio:

$$BCR = \frac{\Delta V}{Cost}$$

Where:

BCR = Benefit-Cost Ratio;

ΔV = Total change in CSO Volume over the period of data collection [L^3];

$Cost$ = The cost of system implementation [$].

Equation 16: Benefit-Cost Analysis

The BCR is computed for every CSO-shed at its optimal value of C_p. A normalized BCR is also computed, where:

$$nBCR = \frac{BCR}{\max(BCR)}$$

Note that, at optimality, $nBCR$ is unity.

4.7 Demonstration of Robustness with Monte Carlo Methods

The advantage of applying Monte Carlo methods is twofold: in the case of large datasets, such as the one from Spokane, the method can be tested to see if its outputs are robust. The following procedure was done to assess the behavior of the model when randomness was introduced.

The gamma distribution is a common distribution used in hydrology to model the frequency of storms of a certain size. Following this logic, a gamma distribution function was fit to the parameter $d_{t,Cp}$ using the MASS package (Ripley, 2016) for the R scripting language. The shape and rate constants for these gamma distribution functions are shown in Table 4:

Table 4: Shape and Rate Constants for GDF fitting dtCp

Shed	Shape Constant	Rate Constant
CSO02	0.462150949228492	1.37755922194204
CSO06	0.734669250585028	1.64646503700817
CSO07	0.781042923879469	1.22078542991454
CSO10	0.652194978664841	0.937738400912821
CSO12	0.725411291705688	1.48510788228932
CSO14	0.688839519200932	1.07761305999719
CSO15	0.684094226631234	1.03531642264169
CSO23	0.679180921213085	1.15263582632527
CSO24A	0.821982645042825	1.57205695357716
CSO24B	0.649652756325846	0.883539842310734
CSO25	0.667764905673663	1.03151324658415
CSO26	0.683376134357938	1.16442030383645
CSO33A	0.861234009424313	0.828077647638232
CSO33B	0.967471244690295	1.35375133456719
CSO33C	0.73576420003766	0.796064918317537
CSO33D	0.699900185069065	1.16589023161657
CSO34	0.921408578259018	1.90456382010835
CSO38	0.83709176599359	1.26507118628418
CSO39	0.6105406208408	0.688687292092215
CSO40	0.551970686123125	1.02294241448893
CSO41	0.747928960385467	1.00875203194555

Next, the CSO volume V_{CSO} was seen to be predicted by the depth to CSO $d_{t,Cp}$ by a linear regression that passed through the origin, with a coefficient of correlation R^2 on the order of about 0.35 for all CSO sheds. The slopes of the regression lines for each shed are:

Table 5: Slope of the Linear Regression comparing dt and V

	Shed	Slope
1	**CSO02**	2333.61111111111
2	**CSO06**	166873.802047782
3	**CSO07**	26598.243902439
4	**CSO10**	13691.4299065421
5	**CSO12**	122837.928125
6	**CSO14**	7887.35810810811
7	**CSO15**	20007.9891304348
8	**CSO23**	65741.0816326531
9	**CSO24A**	377715.696428572
10	**CSO24B**	8849.23636363636
11	**CSO25**	20717.5204081633
12	**CSO26**	651840.97080292
13	**CSO33A**	4143.96470588235
14	**CSO33B**	886743.386666667

15	**CSO33C**	6490.82051282051
16	**CSO33D**	18204.3320754717
17	**CSO34**	662532.683962264
18	**CSO38**	8005.06363636364
19	**CSO39**	13241.6136363636
20	**CSO40**	8420.12121212121
21	**CSO41**	29245.503875969

Finally, randomness was introduced to the system by generating a new set of random variates v' from d_t with a normal distribution using the standard deviation of the original v values seen in Table 5 and the following equation:

$$v'_{ij} = N((m_i * d'_{ij}), \sigma_{vi})$$

Where:

v'_{ij} = The simulated volume of the j[th] event in the i[th] shed [L^3];

m_i = The slope of the linear regression between d_t and v for the j[th] event in the i[th] shed [L^2];

d'_{ij} = The simulated depth-to-CSO for the j[th] event in the i[th] shed [L];

σ_{vi} = The standard deviation of the CSO volumes in the i[th] shed [L^3];

$N()$ = The normal distribution function.

Equation 17: Generation of Simulated CSO Volumes

As the number of CSO events on record is different for each CSO shed, the number of d_t and v' random variate pairs computed was equal to the original number of CSOs recorded for each shed during the period of observation. The difference in the means of the random variates and the original values was close to 0, indicating that the method for producing v' described in this section was an unbiased estimator of v. Figure 15 gives a representation of how this looks:

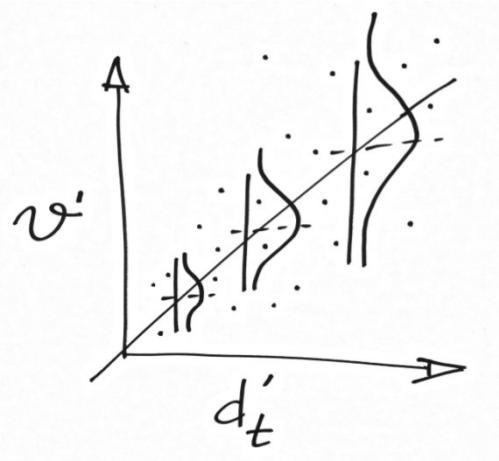


Figure 15: Estimation of CSO Volume from Depth-To-CSO, with randomness

At this point, a Monte Carlo process was initiated. The model was run 100 times at different budget levels, and the frequency that each shed was selected as the optimum shed was recorded, and the resulting value of the objective function was recorded. This led to some surprising results, as will be discussed.

5. RESULTS

5.1 Minimum Depth-to-CSO

The very first insight gained by the method is the estimation of the parameter $d_{t,Cex}$, the minimum depth to CSO prior to intervention. The distribution of those initial values can be seen in the histogram in Figure 16, and some summary statistics are available in table 5. Note that most of the CSO-sheds have a very low value for $d_{t,Cex}$, indicating that in most cases, the effect of LID implementation will be observed after only a modest investment. This is promising news for any infrastructure planner on a limited budget.

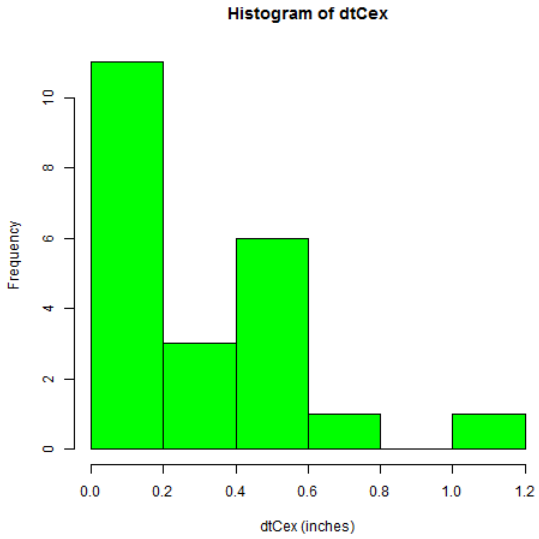


Figure 16: Histogram of dtCex

Table 6: Summary Statistics for dtCex

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
0	0.001680	0.008392	0.027	0.01967	0.4302

5.2 System Behaviour, Deterministic Case

Figure 17 indicates the variation of CSO volume with C_p.

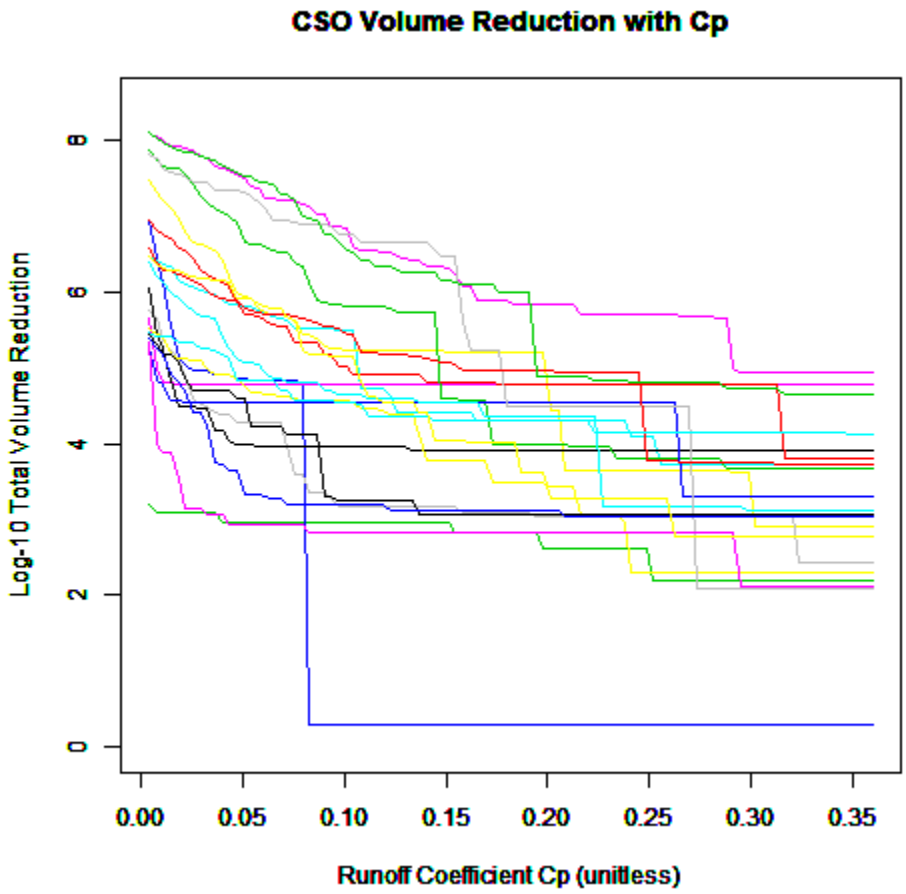


Figure 17: Change in Volume Reduction with Cp

Figure 18 indicates how the maximum CSO volume captured changes with each shed. This

Note that in Figures 17 and 18, both the total and the maximum captured V_{CSO} increase as C_p decreases.

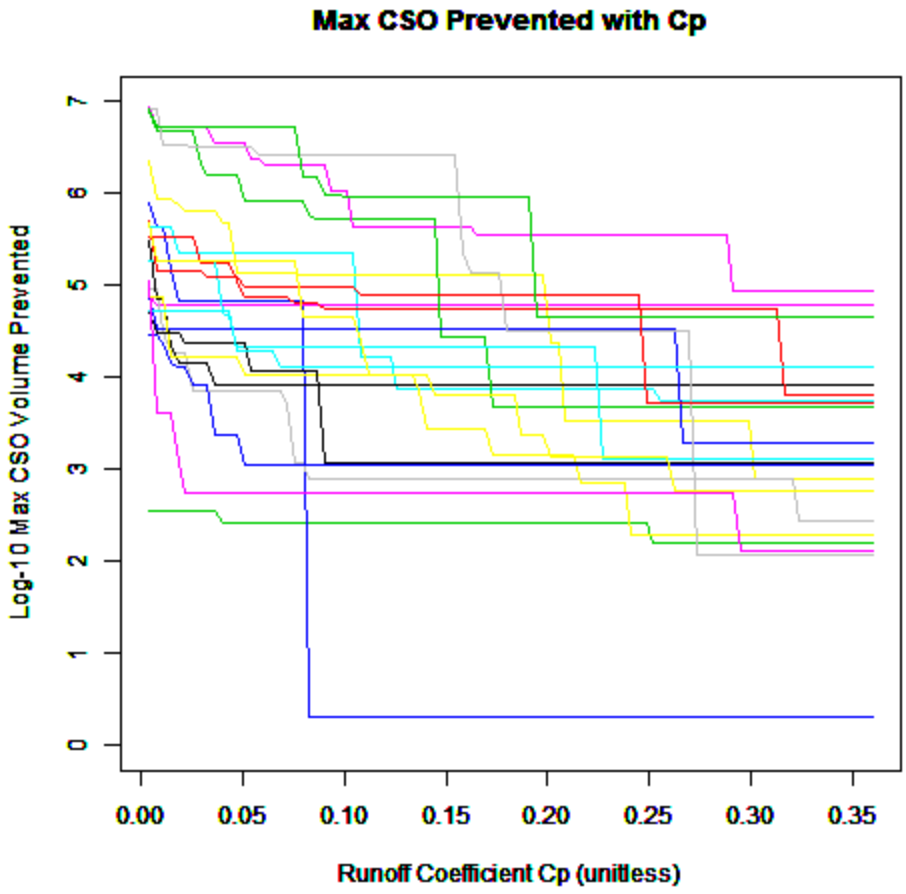


Figure 18: Max CSO Volume Prevented with Cp

The height of the curve on the graphs for cost of implementation and area occupied by LID infrastructure are scalar multiples of the heights of the curves on the graphs in Figure 18, and are therefore not included. Figures 24 through 45 give a much clearer picture of the effects of LID application, indicating the predicted change in cumulative CSO volume and frequency over the record with C_p for each CSO-shed in the analysis. These figures can be found in Appendix A. Note that figures with more discontinuities in the curve indicate a greater number of CSOs in the region over the course of the analysis. The curves clearly show the relationship that as C_p decreases, the CSO volume over the course of the record will also decrease.

Scatterplots of the the cost-benefit ratio for reduction in the various CSO sheds are shown in Figures 19 and 20 for both volume and frequency. Both scatterplots are produced in the environment where LID funds are abundant, with a value of $100M.

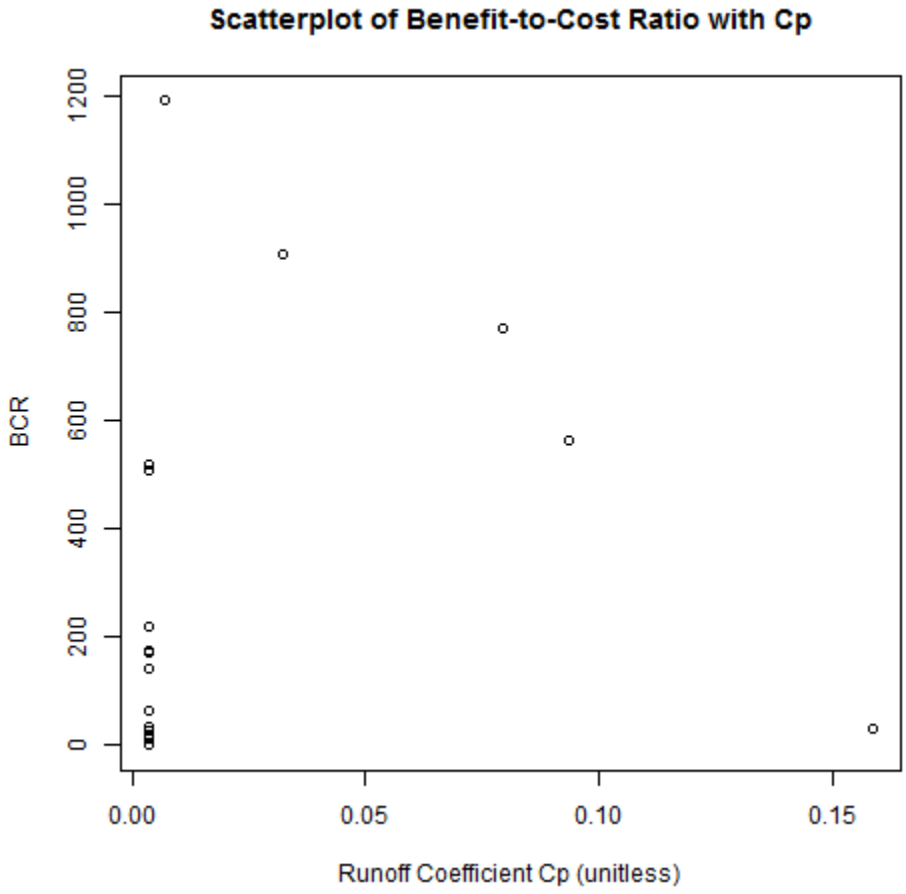


Figure 19: Scatterplot of Benefit-to-Cost Ratio (Volume)

Note that most of the BCR values are close to the origin. The point at the top left is the optimum shed CSO12, using $42.9M to stop 20.4×10^6 m^3 of CSO from reaching the receiving water. The point at the bottom right is CSO33B, using $11.1M to reduce the total flow by a less-impressive 393 153 m^3.

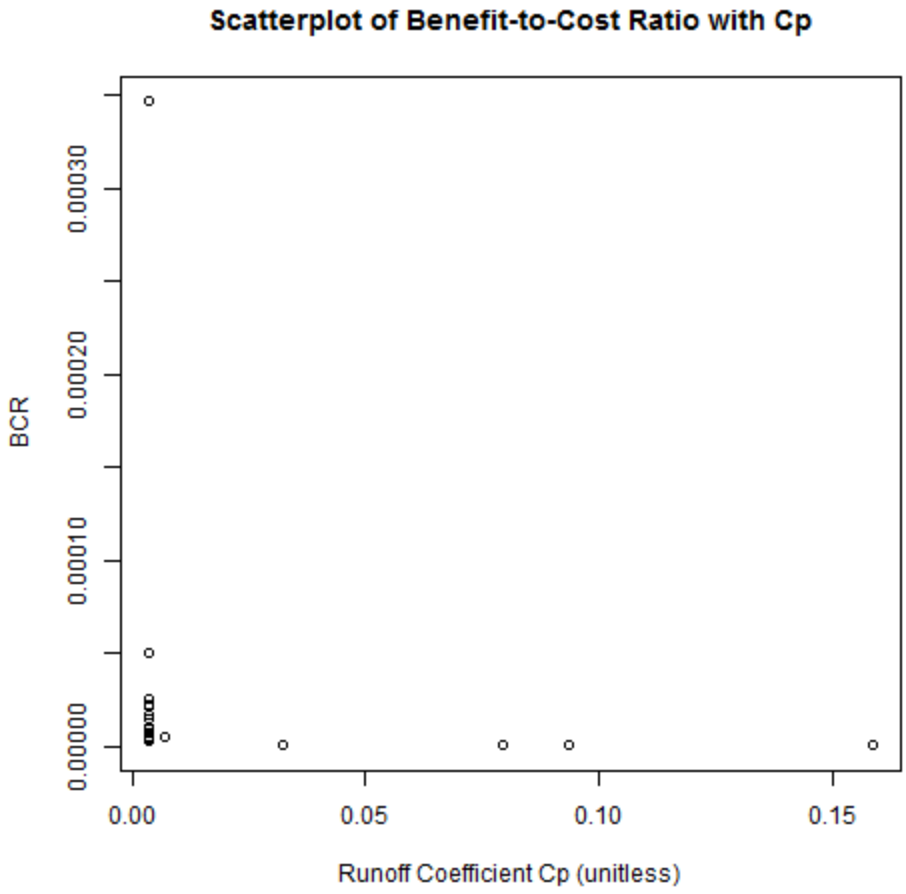


Figure 20: Scatterplot of Benefit-to-Cost with Cp (Frequency)

Reducing the frequency of CSO events is not nearly as interesting as reducing the total volume, but Figure 20 still serves some illustrative purposes. The winning shed, in this case, is CSO02, spending $17.3k to reduce the total flow by 6 events from 18, or exactly 33%. CSO33B is in the bottom-right, spending $11.1M to reduce the total number of CSO events by 72 from 196, or just under 37%.

5.3 Solutions at Optimality, Deterministic Case

The selection of the optimal shed and the extent of intervention was sensitive to changes in the budget K. The solution at optimality for different values of K will be given in the following table. The resolution of the search R was fixed at a value of 100:

Table 7: Optimality in the Deterministic Case

Budget	$1M	$2M	$3M	$4M	$5M	$7.5M	$10M
Shed	CSO14	CSO38	CSO39	CSO23	CSO33D	CSO33D	CSO33D
BCR	9.93	14.19	17.62	40.57	47.90	147.10	147.1
C_p	0.0144	0.0036	0.0036	0.054	0.0504	0.0072	0.0072
$V_{storage}\ (m^3)$	18 550	29 235	52 099	74 502	95 863	145 998	145998
$\Delta V_{CSO}(m^3)$	170 024	245 868	30 5050	52 8961	624 507	2 518 464	2518464
$A_{garden}(m^2)$	5654.04	8910.83	15 879.77	22 708.21	29 219.04	44 500	44500
Cost ($)	912 618	1 438 296	2 563 154	3 665 332	4 716 245	7 182 775	7 182 775

Note the dependency on the budget K up to $5M, at which point CSO33D dominates.

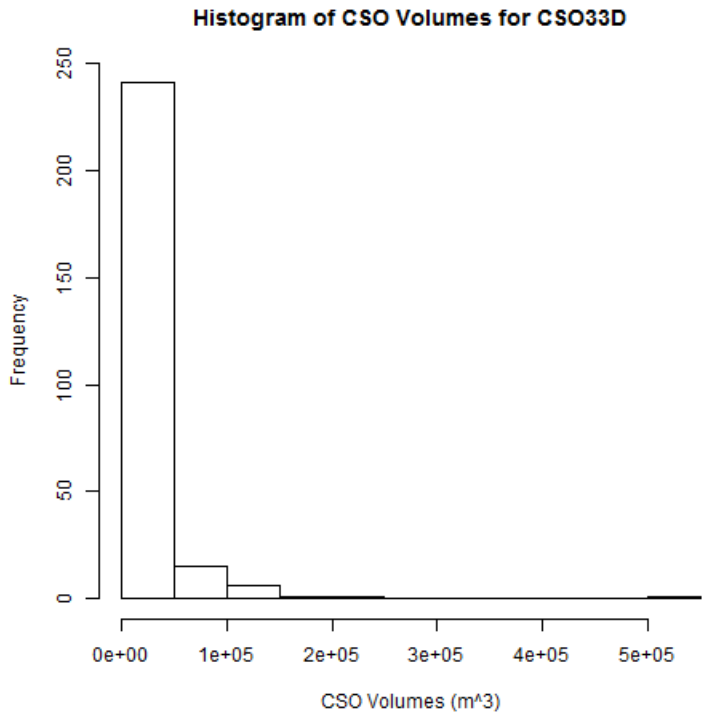


Figure 21: Histogram of CSO Volumes for CSO33D

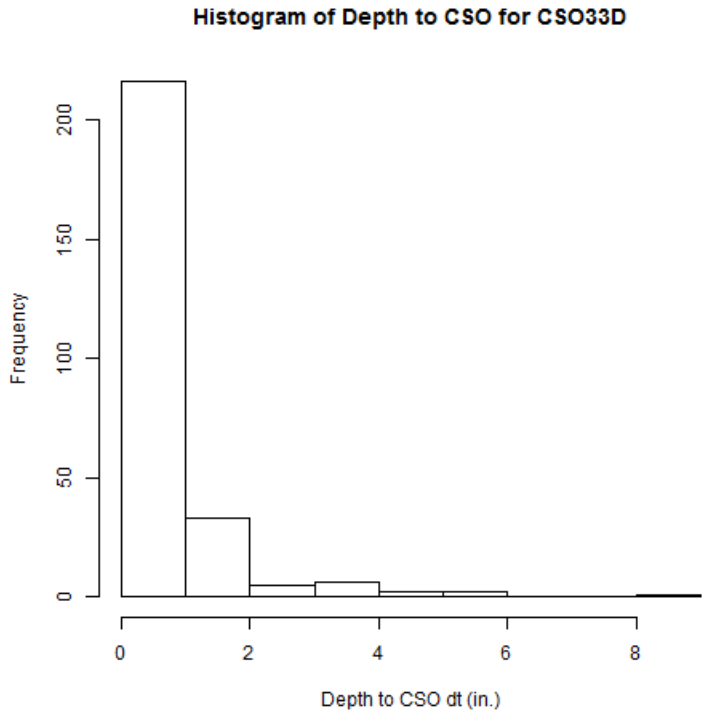


Figure 22: Histogram of dt for CSO33D

It is likely that this is because, as shown in Figure 9, CSO33D has a very large number of CSOs – the second most out of all CSO sheds under inspection, next to CSO12. There appears, by inspection, to be nothing immediately remarkable about the distribution of its values for the depth to CSO d_t as shown in Figure 10, its mean d_t, or the value of its initial minimum depth to CSO $d_{t,Cex}$.

5.4 Robust Analysis with Monte Carlo Methods

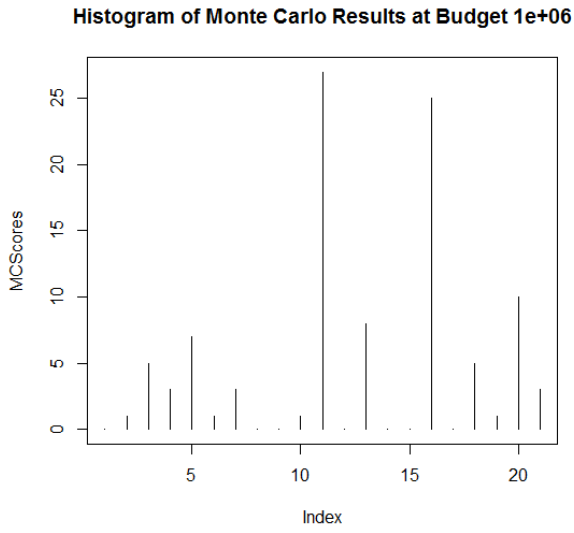
An analysis was performed with Monte Carlo methods to determine the behavior of the system with uncertainty. In this case, the system was run for 100 repetitions at different budget levels and resolution 100. Values of the decision variables at optimality are the means of the values optimal shed over the runs where they were optimal.

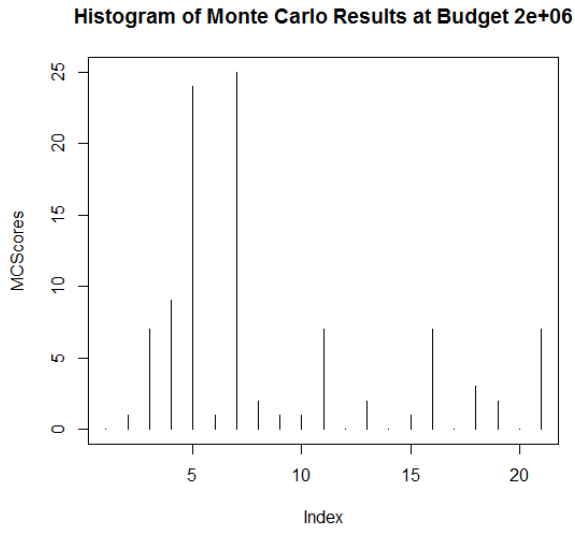
Table 8: Optimality with Randomness

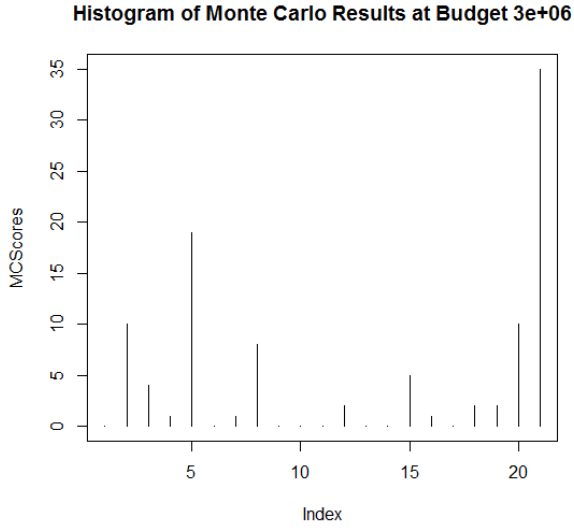
Budget	$1M	$2M	$3M	$4M	$5M	$7.5M	$10M
Shed	CSO38	CSO33D	CSO26	CSO06	CSO12	CSO24A	CSO23
BCR	93.23	147.04	79.92	141 910.8	178.51	1334.06	76.44
C_p	0.01296	0.0036	0.0036	0.00409	0.0124	0.01842	0.00405
$V_{storage}\ (m^3)$	16 425.72	19 914.74	50580.52	56 722.03	58 286.63	114 018.3	80 603.14
$\Delta V_{CSO}(m^3)$	1 397 144	4 145 338	5 309 578	13 509 820	13 377 400	16 599 056	8 441 993
$A_{garden}(m^2)$	5006.56	6070.01	15 416.94	17 288.87	17 766.68	34752.78	24 567.84
Cost ($)	808 108	979 760	2 488 449	2 790 597	2 867 720	5 609 446	3 965 494

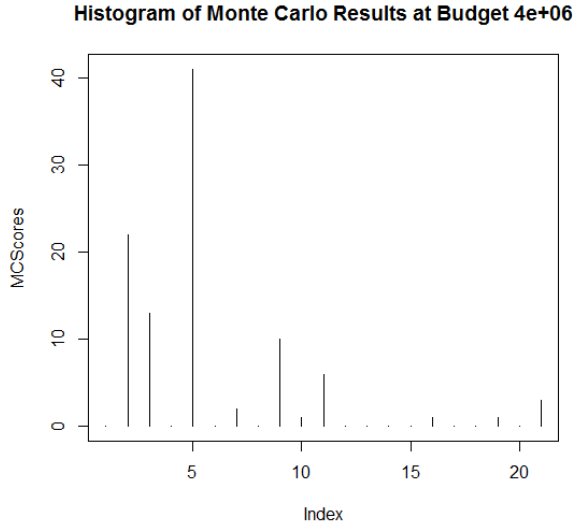
Looking at the histograms in Figure 23 and the data in Table 7, a few things are apparent. First, the result obtained in the deterministic case was never the same as the result obtained when randomness was introduced. This means that at each value of the budget – and quite possibly, everywhere between – the signal in the data that indicated the optimal solution was not strong enough to withstand the noise brought in by the randomness.

Even CSO33D, which was selected for LID intervention in the three deterministic run conditions where budget was highest, only appeared once in the case where randomness was introduced – and even then, it was in a lower-budget case.


Histogram of Monte Carlo Results at Budget 1e+06
MCScores
Index
0
5
10
15
20
25
5
10
15
20


Histogram of Monte Carlo Results at Budget 2e+06
MCScores
Index
0
5
10
15
20
25
5
10
15
20


Histogram of Monte Carlo Results at Budget 3e+06
MCScores
Index
0
5
10
15
20
25
30
35
5
10
15
20


Histogram of Monte Carlo Results at Budget 4e+06
MCScores
Index
0
10
20
30
40
5
10
15
20

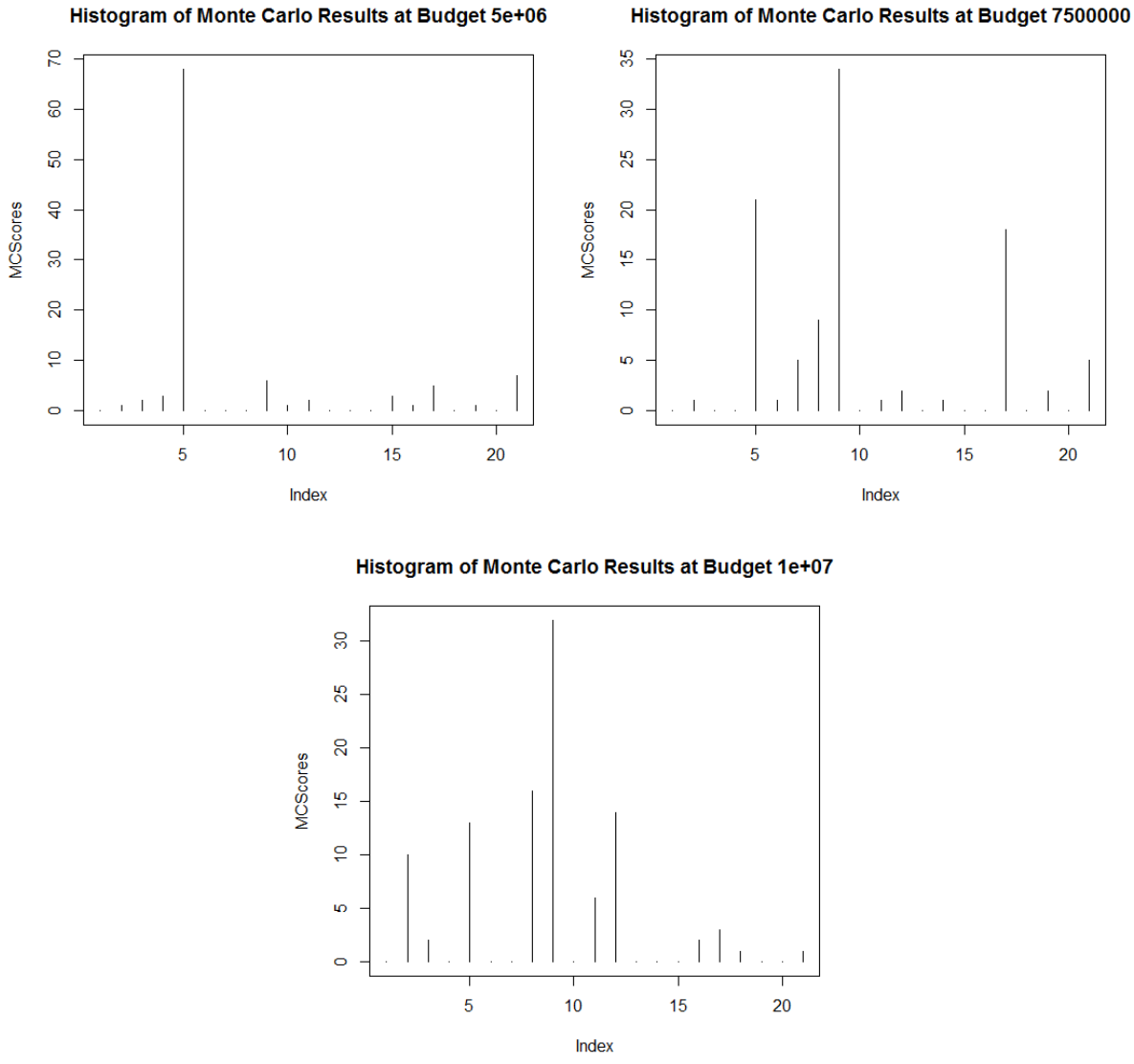


Figure 23: Histograms of Monte Carlo Results

These results indicate that it is perilous to accept without skepticism the outputs of decision models based on data where there is some level of uncertainty.

6. DISCUSSION

There are several components to this assessment method. The availability of data, the use of an algorithm for determining $d_{t,Cex}$ for each CSO in the record, the MIP formulation of LIDRA and the algorithm used to solve it, and the Monte Carlo simulation are all factors that determine the priority of CSO-sheds for LID infrastructure investment, and the extent of that investment. Therefore, it is important to consider how manipulating the components of the system are likely to affect the model. Each will be discussed in turn, in terms of possibilities for future research.

6.1 The Availability of Data

The assessment approach described in this manuscript is demonstrated with the help of a large data set. This data set spans over 13 years, and consists of 763 CSO-causing storms and a total of 3239 CSO events. This work could not have been compiled without this data, but what is to be done in municipalities which have not acquired a similar record? Without data, municipalities are forced to base their decisions on the outputs from urban stormwater numerical modelling packages, which may be costly to implement and hard to verify. Municipalities curious about urban stormwater infrastructure decisions should install CSO measurement systems as soon as possible in order to increase the quality of the decisions they make with public funds!

Basing infrastructure spending decisions on data will almost always be preferable to estimates, but if only a modest amount of data is available, then statistical methods can still potentially be used to supplement the recorded values with random variates. If a municipality has a small amount of data, say, two years' worth of CSO and storm recordings, then urban infrastructure planning professionals could model the available data with an appropriate distribution curve, in a manner similar to that which was demonstrated in Section 4.7. Then,

once the curve has been fit, random variates could be generated to supplement the existing data. Infrastructure planning professionals have the choice of either simulating storm and CSO records, or simulating $d_{t,Cex}$ values directly, as demonstrated.

6.2 Determining the Minimum Depth to CSO

The assessment approach described in this manuscript makes use of a GLSA to generate a matching on the bipartite graph of CSOs and storms. From this matching, values for the minimum depth to CSO, d_t, are computed as LIDRA inputs. It is important to note here that this matching constitutes a correlation, and therefore, no causal inferences can be made about the relationship between CSOs and storm. That said, correlationary data retains a great deal of predictive power, and so the question arises as to how to improve the quality of these predictions.

GLSA algorithms are simple to implement, but they are imperfect. What, then, are the features of the data set that could potentially degrade the quality of the d_t values generated in the algorithm, and how can the algorithm for generating the d_t values be improved?

Note that the GLSA works by finding the storm immediately prior to each CSO, and records the time between storm onset and CSO onset as the edge weighting t. Consider, however, that it is possible for multiple discrete storm events to occur with an inter-arrival time less than this value t. Consider also that there exists a lag time between the arrival of the storm and the measurement of the CSO. This is an issue, because then the GLSA would then select the most recent storm despite the possibility that the storm which prompted the CSO is the one immediately prior to the one selected – the actual CSO surge would still be on its way. The

likelihood of this failure case of the GLSA could be avoided by taking the initial value t, and checking to see if another storm occurred within the window:

$$W = [t_{CSO} - 2t, \qquad t_{CSO} - t]$$

Equation 18: Window for CSO-storm pairing metaheuristic condition

In this case, the value d_t could be improved by using a metaheuristic which drops the CSO event entirely. In this case, the algorithm would avoid using CSO-storm data in situations where many discrete storms occurred. A consequence of this, however, would be a reduction of the number of d_t values as LIDRA inputs. This would result in a reduction of decision confidence, which could be exacerbated in the case of a small dataset.

Another possibility for GLSA failure is the case of CSO events not prompted by storms. This case can occur in the event of the spring melt, which is a significant issue in Spokane. Thankfully, the Spokane storm data used in this demonstration indicated these Dry Overflow (DO) events in the storm record, which allowed the GLSA to generate d_t values for these CSOs.

A third possibility for GLSA failure is CSO events caused by the combined effect of storms and spring melt, where the hydraulic loading on the sewer system is some combination of rainfall and meltwater. In this case, the computation of d_t by interpolation fails, as the value of the storm depth value D and the rainfall intensity:

$$i = D/T$$

Equation 19: Average Rainfall Intensity

no longer represent the hydraulic load on the sewer. How, then, is d_t to be accurately produced for these CSO cases? Decoupling the two sources of load may be impossible, it is hypothesized that a large standard deviation in d_t could indicate the confounding variable of spring melt. If these cases were removed from the ranking analysis on the basis of a standard deviation cutoff value, and if a between-groups comparison of the with and without spring melt cases were conducted, it would be interesting to see if the ranking outputs were to change or not – and, if so, to reveal the sensitivity of the ranking with respect to the standard deviation cutoff value.

6.3 LIDRA Algorithm Resolution

The LIDRA algorithm described in this manuscript is based on several user-defined values, one of which is the resolution of the search R. In this demonstration, C_{ex} gets the value 0.36 given in the Spokane Regional Stormwater Manual (City of Spokane, City of Spokane Valley, Spokane County, 2008) and R was arbitrarily selected to be 100.

R can be varied to give different results, and a brief discussion of the sensitivity of the results to a change in resolution follows. The hypothesis is that an increased resolution improves the estimate of the area of land to be converted rain garden, and therefore improves the estimate of the cost, up to a point. A few examples of the variation of the results with R are presented, with the budget set to $5M:

Table 9: Sensitivity of the LIDRA Algorithm to Resolution

Resolution	10	20	50	75	100	250	500
Shed	CSO38	CSO33D	CSO33C	CSO39	CSO33D	CSO33A	CSO26
BCR	0.304	44.55	12.37	17.52	47.90	6.297	12.72
C_p	0.0036	0.054	0.0072	0.0047	0.0504	0.00144	0.292
$V_{storage}\ (m^3)$	2326	95 863	30 256	52 099	95 863	76 628	86 009
$\Delta V_{CSO}(m^3)$	5657	580 775	21 192	299 957	624 507	352 237	101 453
$A_{garden}(m^2)$	708.96	29 219	9 222	15 879	29 219.04	23 356	26 215
Cost ($)	11 434	4 716 246	1 488 528	2 563 154	4 716 245	3 769 926	4 231 450

Clearly the system is sensitive to changes in resolution, as the algorithm selects different sheds at nearly every resolution investigated! The choice of resolution itself is not so simple. The total reduction in CSO is highest for the case where $R = 100$, and so is the benefit-cost ratio. There appears to be no obvious relationship between the resolution and the quality of results – which is a surprise, as one would expect that a more fine-grained search would always give better results. It appears that it is possible for better solutions to "slip between the cracks" at higher resolutions. There is no difference, however, between the second and third cases – so it appears that the algorithm has found the optimum with maximum precision. It is conceivable that in certain cases, the algorithm may pick an entirely different shed for intervention – although that is certainly not happening with the inputs from Spokane.

Finally, increasing R has an effect on the amount of time it takes to compute these results: $R = 20$ results in 4.71s of run time, $R = 200$ results in 10.17s of run time, and $R = 200$ results in 46.27s of run time. With very large datasets at high resolution, such as in the case of simulated rainfall and CSO models, or a bootstrapped dataset, this run time may increase a great deal.

6.4 Selection of Decision Variables

It is important to remember, after all, that the reduction in CSOs is a means to an end: the reduction of negative environmental impact from stormwater loading in the urban environment. Montalto et al base their LIDRA model on a reduction in CSO hours (Montalto, et al., 2007), but it should be noted that it's the dose that makes the poison: the environmental impact of CSO events is determined by the species and rate of contaminant loading (Mulligan, 2002). If the contaminant concentration is assumed to be constant for all time across all CSO outfalls, then the reduction of CSO volume is a more important goal than the reduction of CSO duration.

Therefore, the prioritization of CSO-sheds in this manuscript is dictated by a reduction in CSO discharge volume.

Future work could investigate the environmental impact of different prioritization criteria: total CSO duration, number of CSO events, or average CSO flow rate, and the sensitivity of the receiving body (Lau, Butler, & Schütze, 2002) (Eganhouse & Sherblom, 2001). Comparisons of these decision variables would have to take into account the transportation and fate of environmental contaminant loads. Other possibilities are measuring the concentration of environmental contaminants at the CSO outfalls to add new decision variables to the model. Including these variables would represent an improvement to the decision model.

As discussed, rain gardens are not the only LID infrastructure solution, and while space is not at a premium in larger, less-populated areas like CSO24A or CSO33B, there may be little area to devote to rain gardens in CSO26, the priority shed at selected by the Monte Carlo simulation given the $3M budgetary constraint. In CSO26, green roofs and permeable pavement may represent much more effective means of controlling CSOs. These approaches all have different costs associated with them, as well as different effects on C. Therefore, a model that attempts to solve the system for a mix of different LID solutions may produce better results in terms of the ratio of benefit to cost.

Finally, since this method is faster to implement than physically modelling of a municipal sewer system, it can be used as a feasibility study before hiring a consultant to perform the analysis.

6.6 Alternate Decision Models

Other decision models exist for comparing decisions on a cost-benefit basis. For example, Data Envelopment Analysis (DEA) compares efficiencies across a range of decision-making units (DMUs) in a set, ranking them in terms of relative efficiency (Charnes, Cooper, & Rhodes, 1978). In this case, CSO39 would be assigned the value 1, and other sheds would be given values somewhere between 0 and 1. An approach like this, if combined with the selection of multiple LID solutions, would be able to make further recommendations on improving the amenability of CSO sheds to LID solutions. With the benefit of LIDRA and abundant storm and CSO data, the AHP-based decision support model of Ahammed, Hewa, and Argue (Ahammed, Hewa, & Argue, 2012) could be improved beyond the level of depending simply on expert testimony.

The inclusion of other LID SWM solutions, as mentioned, may also be used in such a model. It would be very useful for infrastructure planning professionals to have a versatile tool which would be easy to deploy that would quickly give the best mix of LID solutions for each CSO-shed in a municipality, as well as simply suggesting priority regions for intervention.

6.7 Robustness of the Results

Comparing the results in the deterministic case with those outputs of the Monte Carlo simulation, it is easy to see that the inclusion of random variables makes for a more robust model. This is good advice for anyone attempting to make decisions based on data where there is an element of uncertainty inherent in the data. Most decisions, in fact, are made without perfect information, and so developing models that can be hardened against finicky data is an important priority when building decision support models.

Without the use of Monte Carlo methods, it would be very easy to select the wrong shed. Comparing the benefit-cost ratios of the sheds in the deterministic and Monte Carlo conditions, the benefit-cost ratios are markedly higher in all-but-one of the cases. This result indicates that, without the use of Monte Carlo methods, the signal in the data is not strong enough to withstand the noise!

6.8 Benefit-Cost Analysis

As a final note, an interesting pattern was noted in the results. For every optimal shed, the benefit-cost ratio was the highest of all the other sheds. This is interesting because the model was not designed to favour high cost-benefit ratios – the emergence of this result may be a consequence of the data, the model, or some combination of the two.

Further research is needed to determine why this is the case, but it is hypothesized that this behavior is governed by the CSO volumes, which are gamma-distributed. Higher volumes are rare, and therefore, the largest CSO volume captured in a given shed at a given C_{ex} is so much larger than the next-smallest volume that the larger volume determines the optimal shed completely. This would drive up the benefit-cost ratio rapidly for the winning shed.

7. CONCLUSION

This thesis has described a path from data to decision-making for urban stormwater infrastructure improvement. The described case of two large datasets, a simple algorithm, a well-known LID modelling tool, and the introduction of randomness gives a useful estimate of CSO reduction targeting for SWM professionals. This method can help planners avoid the costly error of targeting the wrong CSO regions, instead selecting those regions where taxpayer dollars will do the most good.

The method also represents a useful starting point for future research. Much work can be performed on the subjects of supplementing and improving the decision variable datasets, improving the algorithm which performs and describes the matchings on these datasets, selecting new LID and environmental impact assessment methods, and improving the robustness of the model with respect to variation in data. In an age where publicly-available data is all but abundant, it is important to use that data to inform decision making in the context of engineering for the public good.

It bears repeating that the results of data-driven prioritization of CSO-shed targeting for LID application described in this manuscript are intended to be predictive – and that therefore, experimentation is required to ensure that these predictions are accurate. It is thus recommended that LID is applied on a trial basis in one or two CSO-sheds. Then, once a sufficient amount of time has passed and an additional post-LID CSO event record is produced, the CSO records can be compared between the pre-LID and post-LID conditions to see whether the reduction in CSO corresponds to the prediction. If that is the case, then the approach described in this manuscript will be shown to be a useful tool for communities worldwide.

Endnotes

The author would like to acknowledge Lars Hendron and Chris Kuperstein from City of Spokane Wastewater Management, for curating and making available the data on the CSO and storm patterns in their city. Without their excellent data and patient participation, this thesis would surely not exist.

Works Cited

Ahammed, F., Hewa, G. A., & Argue, J. R. (2012). Applying multi-criteria decision analysis to select WSUD and LID technologies. *Water Science and Technology: Water Supply*, 844-853.

Appan, A. (1999). A dual-mode system for harnessing roofwater for non-potable uses. *Urban Water*, 317-321.

Ariratnam, S. T., & MacLeod, C. W. (2002). Financial Outlay Modeling for a Local Sewer Rehabilitation Strategy. *Journal of Construction Engineering and Management*, 486-495.

Battaglin, W. A., & Kolpin, D. W. (2009). Contaminants of emerging concern: introduction to a featured collection. *Journal of the American Water Resources Association, 45*(1), 1-3.

Bedient, P., Huber, W., & Vieux, B. (2008). *Hydrology and Floodplain Analysis (4th Edition).* Upper Saddle River, NJ: Prentice Hall.

Black, P. E. (2005, 2 2). *greedy algorithm*. Retrieved from Dictionary of Algorithms and Data Structures: https://xlinux.nist.gov/dads//HTML/greedyalgo.html

Brombach, H., Weiss, G., & Fuchs, S. (2005). A new database on urban runoff pollution: Comparison of separate and combined sewer systems. *Water Science & Technology, 51*(2), 119-131.

Cahill, T. H. (2012). *Low Impact Development and Sustainable Stormwater Management.* Hoboken, NJ.: Wiley.

CBC News. (2015, November 12). *Montreal's raw sewage dump makes international headlines*. Retrieved from CBC News Montreal: http://www.cbc.ca/news/canada/montreal/international-media-response-montreal-sewage-dump-st-lawrence-river-1.3316173

Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 429-444.

City of Spokane. (2016, 02 16). *Program Budget - CIty of Spokane, Washington*. Retrieved from spokanecity: https://my.spokanecity.org/budget/budget-tool/

City of Spokane, City of Spokane Valley, Spokane County. (2008). *Spokane Regional Stormwater Manual.* Spokane, WA.

City of Spokane; City of Spokane Valley; Spokane County. (2008). *Spokane Regional Stormwater Manual.* Spokane, WA.

Clemen, R. T., & Reilly, T. (2001). *Making Hard Decisions.* Pacific Grove, CA: Brooks/Cole.

Davis, M. L., & Cornwell, D. A. (2008). *Introduction to Environmental Engineering.* New York: McGraww-Hill.

Dietz, M., & Clausen, J. (2008). Stormwater runoff and export changes with development in a traditional and low impact subdivision. *Journal of Environmental Management, 87*(4), 560-566.

Eganhouse, R., & Sherblom, P. (2001). Anthropogenic organic contaminants in the effluent of a combined sewer overflow: impact on Boston Harbor. *Marine Environmental Research*, 51-74.

Elliott, A., & Trowsdale, S. (2007). A review of models for low impact urban stormwater drainage. *Environmental Modelling & Software, 22*(3), 394-405.

Fenner, R., & Sweeting, L. (1999). A decision support model for the rehablitation of "non-critical" sewers. *Water Science and Technology*, 193-200.

Field, R., Sullivan, D., & Tafuri, A. N. (2004). *Management of Combined Sewer Overflows.* Boca Raton, FL.: CRC Press.

Gasperi, J., & al., e. (2008). Priority pollutants in wastewater and combined sewer overflow. *Science of the Total Environment*, 263-272.

Glover, F. (1967). Maximum matching in a convex bipartite graph. *Naval Research Logistics Quarterly*, 313-316.

Gupta, K., & Saul, A. J. (1996). Specific relationships for the first flush load in combined sewer flows. *Water Research*, 1244-1252.

Halfawy, M. R., Dridi, L., & Baker, S. (2008). Integration Decision Support System for Optimal Renewal Planning of Sewer Networks. *Journal of Computing in Civil Engineering*, 360-372.

Hood, M. J., Clausen, J. C., & Warner, G. S. (2007). Comparison of stormwater lag times for low impact and traditional residential development. *Journal of the American Water Resources Association, 43*(4), 1036-1046.

Houle, J. J., & al., e. (2013). Comparison of Maintenance Cost, Labor Demands, and System Performance for LID and Conventional Stormwater Management. *Journal of Environmental Engineering*, 932-938.

International Organization for Standardization. (2009, 09 15). *ISO/IEC/IEEE 9945:2009 - Information technology -- Portable Operating System Interface (POSIX®) Base Specifications, Issue 7*. (International Organization for Standardization) Retrieved 10 7, 2015, from http://www.iso.org/iso/catalogue_detail.htm?csnumber=50516

Jones, M. P., & Hunt, W. F. (2009). Performance of rainwater harvesting systems in the southeastern United States. *Resources, Conservation and Recycling*, 623-629.

Kammen, D. M., & Hassenzahl, D. M. (1999). *Should We Risk It? Exploring Tenvironmental, Health, and Technological Problem Solving.* Princeton, NJ: Princeton University Press.

Lau, J., Butler, D., & Schütze, M. (2002). Is combined sewer overflow spill frequency/volume a good indicator of receiving water quality impact? *Urban Water*, 181-189.

Lee, J. G., Heaney, J. P., & Lai, F.-h. (2016, 3 17). *Optimization of Decentralized BMP Controls in Urban Areas.* Retrieved from United States Environmental Protection Agency: https://cfpub.epa.gov/si/si_public_file_download.cfm?p_download_id=484278

Lee, J. H., & Bang, K. W. (2000). Characterization of Urban Stormwater Runoff. *Water Research, 34*(6), 1773-1780.

Maechler, M. (n.d.). *Date-time Conversion Functions*. (Seminar für Statistik, ETH Zurich) Retrieved 7 10, 2015, from https://stat.ethz.ch/R-manual/R-devel/library/base/html/as.POSIXlt.html

Maechler, M. (n.d.). *Date-time Conversion Functions to and from Character*. (Seminar für Statistik , ETH Zurich) Retrieved 07 10, 2015, from https://stat.ethz.ch/R-manual/R-devel/library/base/html/strptime.html

Matamoros, V., Garcia, J., & Bayona, J. M. (2008). Organic micropollutant removal in a full-scale surface flow constructed wetland fed with secondary effluent. *Water Research*, 653-660.

Montalto, F., Behr, C., Alfredo, K., Wolf, M., Arye, M., & Walsh, M. (2007). Rapid assessment of the cost-effectiveness of low impact development for CSO control. *Landscape and Urban Planning, 82*, 117-131.

Mulligan, C. N. (2002). *Environmental Biotreatment: Technologies for Water, Air, Soil, and Waste.* Rockville, MD: Government Institutes.

Nielson, B., & Turney, D. (2010). Green Infrastructure Optimization Analyses for Combined Sewer Overflow. *American Society of Civil Engineers Proceedings: Low Impact Development 2010*, 1533-1541.

Oregon State University. (2016, 3 17). *Final Step - LID 1.04 Lined filtration rain garden*. Retrieved from OSU Extension Service: Water and Watershed Education | Stormwater Solutions: http://extension.oregonstate.edu/stormwater/final-step-3

Quick-R. (2016, 02 06). *Graphical Parameters*. Retrieved from Quick-R: http://www.statmethods.net/advgraphs/parameters.html

Revelle, C. S., Whitlach, E. E., & Wright, J. R. (2004). *Civil and Environmental Systems Engineering.* Upper Saddle River, NJ.: Pearson Prentice Hall.

Richardson, S., & Ternes, T. (2005). Water Analysis: Emerging Contaminants and Current Issues. *Analytical Chemistry, 77*(12), 3807-3838.

Ripley, B. (2016, 03 21). *MASS: Support Functions and Datasets for Venables and Ripley's MASS*. Retrieved from CRAN - Package MASS: https://cran.r-project.org/web/packages/MASS/index.html

Saaty, T. L. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research*, 9-26.

Sample, D. J., & al., e. (2001). Geographic Information SYstems, Decision Support Systems, and Urban Storm-Water Management. *Journal of Water Resources Planning and Management*, 155-161.

Selman, B., & Kautz, H. A. (1993). An Empirical Study of Greedy Local Search for Satisfiability Testing. *AAAI* (pp. 46-51). Murray Hill, NJ: AT&T Bell Laboratories.

Servais, P., Seidl, M., & Mouchel, J.-M. (1999). Comparison of Parameters Characterizing Organic Matter in a Combined Sewer During Rainfall Events and Dry Weather. *Water Environment Research, 71*(4), 408-417.

Taebi, A., & Droste, R. L. (2004). Pollution loads in urban runoff and sanitary wastewater. *Science of the Total Environment, 327*(1-3), 175-184.

Thorolfsson. (1999). New strategies in stormwater-meltwater management in the city of Bergen, Norway. *Water Science and Technology, 39*(2), 169-176.

Toffol, S. D., Engelhard, C., & Rauch, W. (2007). Combined sewer system versus separate system – a comparison of ecological and economical performance indicators. *Water Science and Technolgoy, 55*(4), 255-264.

USEPA. (2015, 12 1). *EPANET*. Retrieved from US Environmental Protection Agency: http://www2.epa.gov/water-research/epanet

Western Regional Climate Center. (2005, 12 31). *SPOKANE WSO AIRPORT, WASHINGTON Period of Record Monthly Climate Summary*. (Western Regional Climate Center) Retrieved 01 28, 2015, from http://www.wrcc.dri.edu/cgi-bin/cliRECtM.pl?waspok

Wirahadikusumah, R., & al., e. (1998). Assessment technologies for sewer system rehabilitation. *Automation in Construction*, 259-270.

Wirahadikusumah, R., Abraham, D. M., & Castello, J. (1999). Markov decision process for sewer rehabilitation. *Engineering, Construction and Architectural Management*, 358-370.

Wolsey, L. A. (1998). *Integer Programming.* New Jersey: Wiley-Interscience.

Yazdanfar, Z., & Sharma, A. (2015). Urban drainage system planning and design – challenges with climate change and urbanization: a review. *Water Science & Technology, 72*(2), 165-179.

APPENDIX

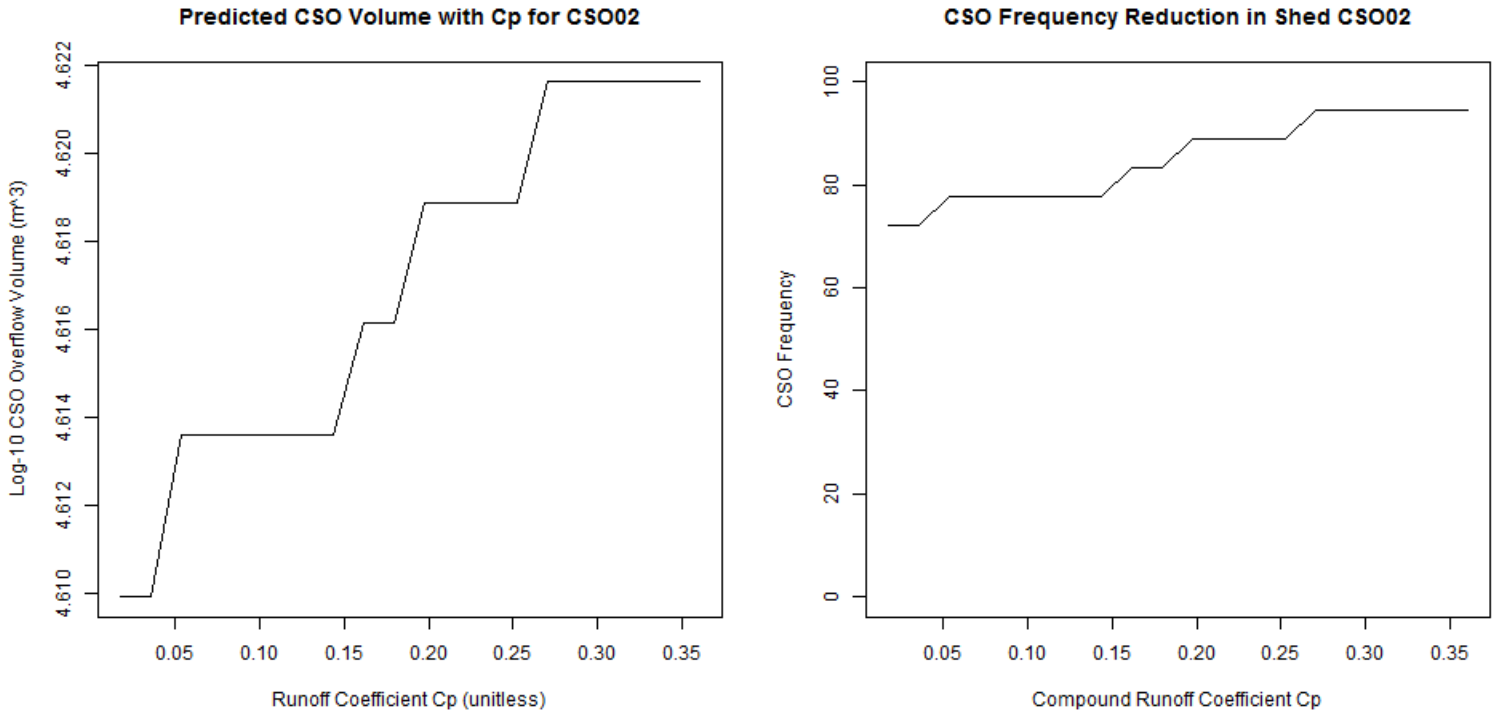


Figure 24: Predicted Change to CSO Parameters in CSO02

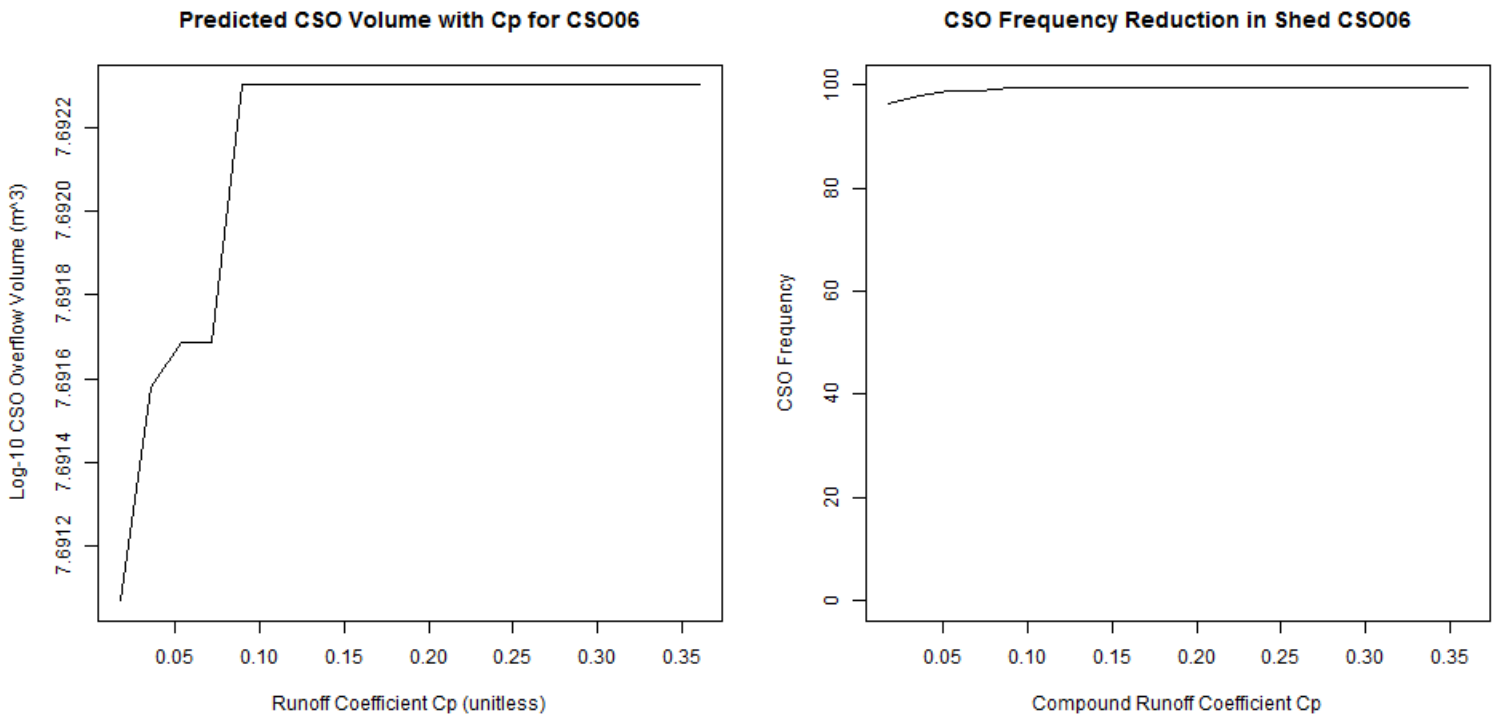


Figure 25: Predicted Change to CSO Parameters in CSO06

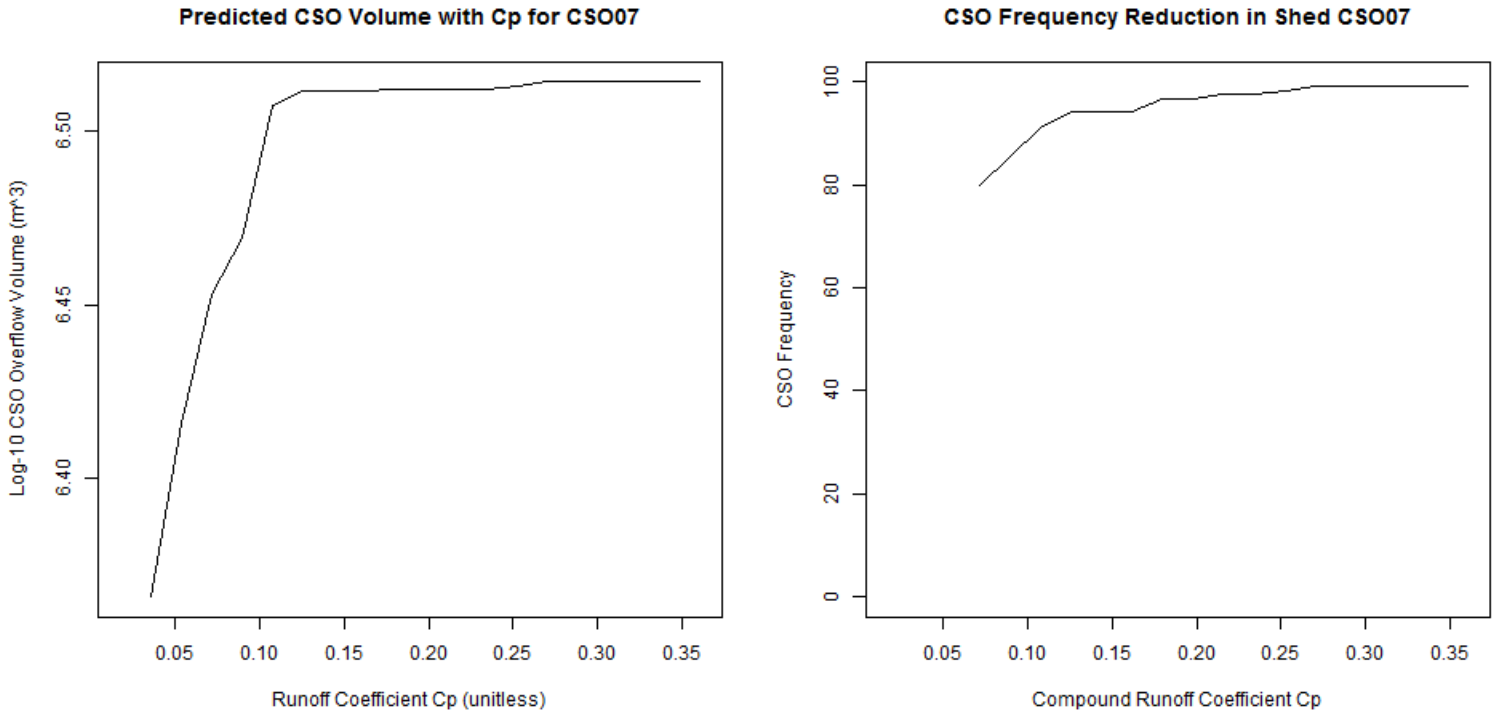


Figure 26: Predicted Change to CSO Parameters in CSO07

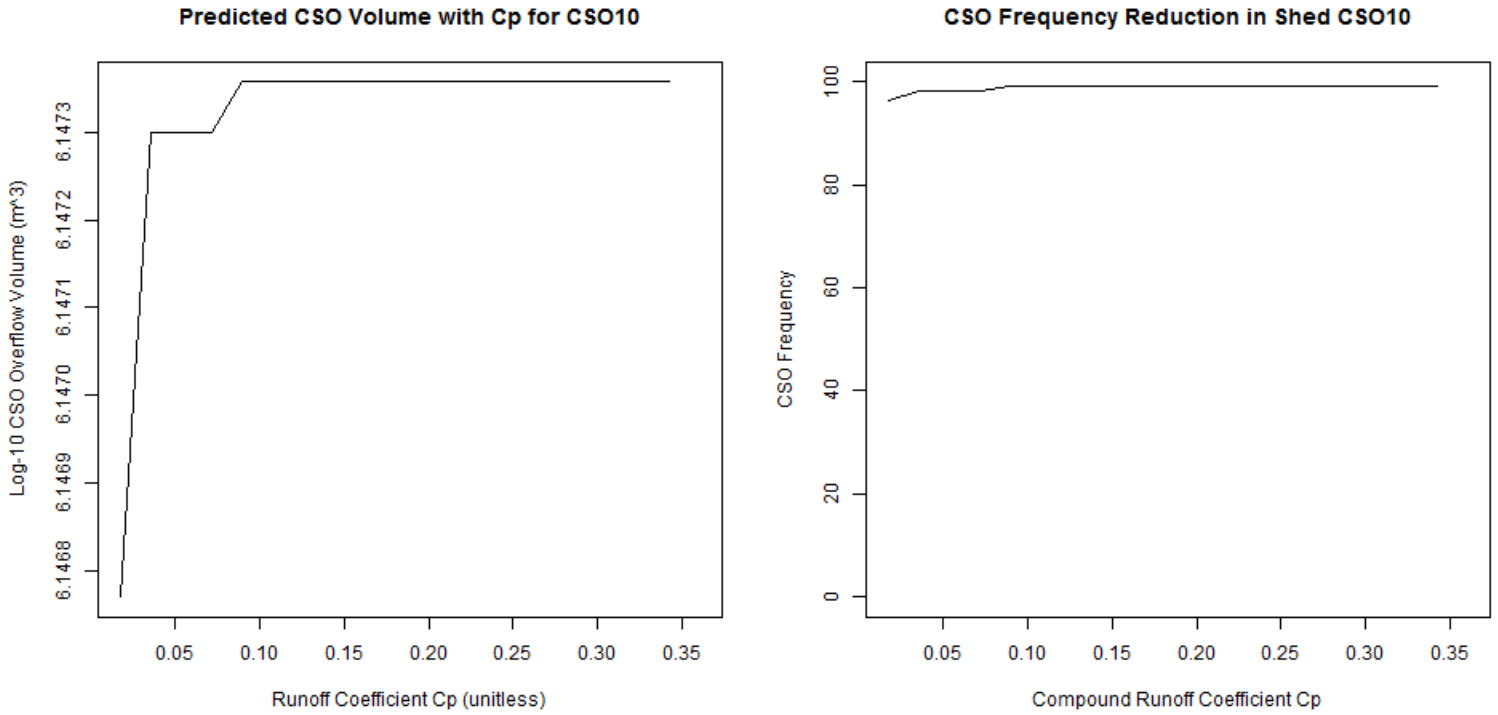


Figure 27: Predicted Change to CSO Parameters in CSO10

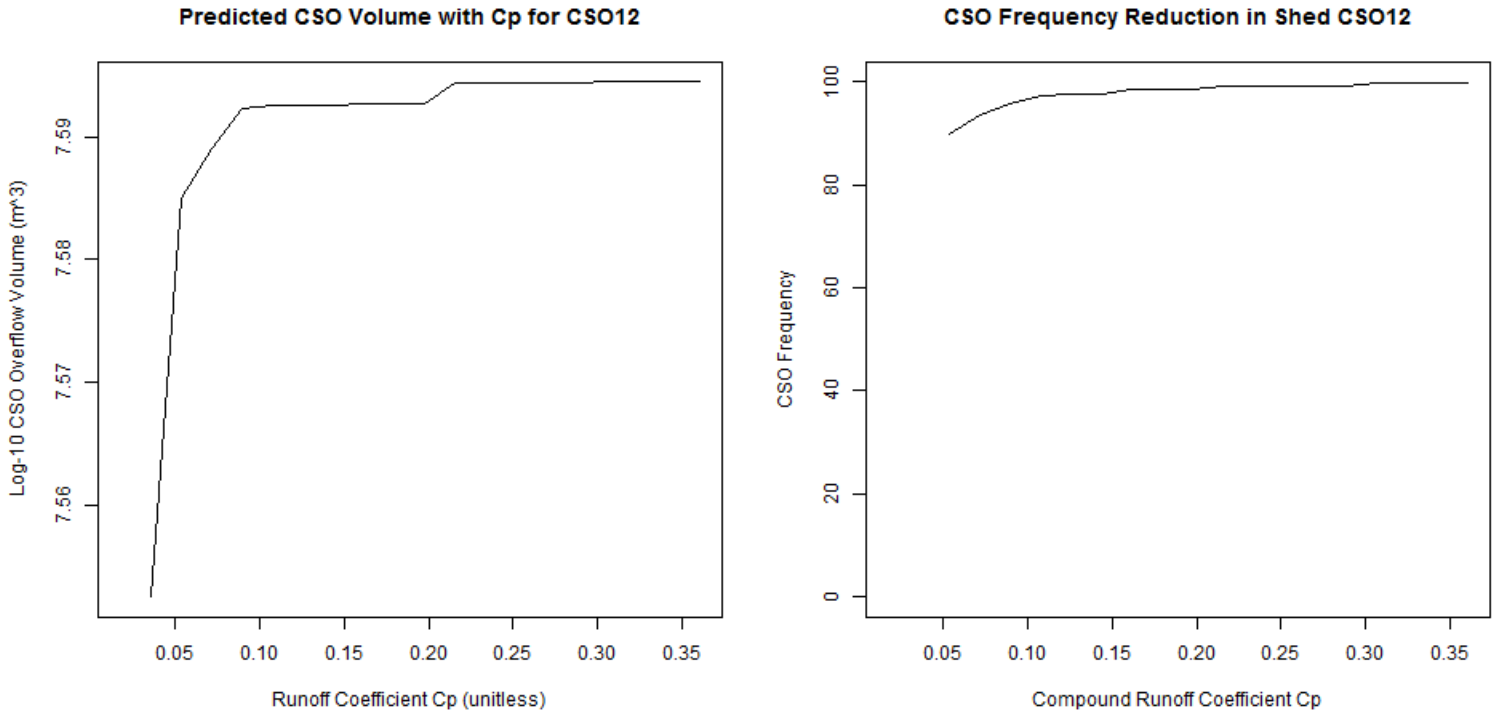


Figure 28: Predicted Change to CSO Parameters in CSO12

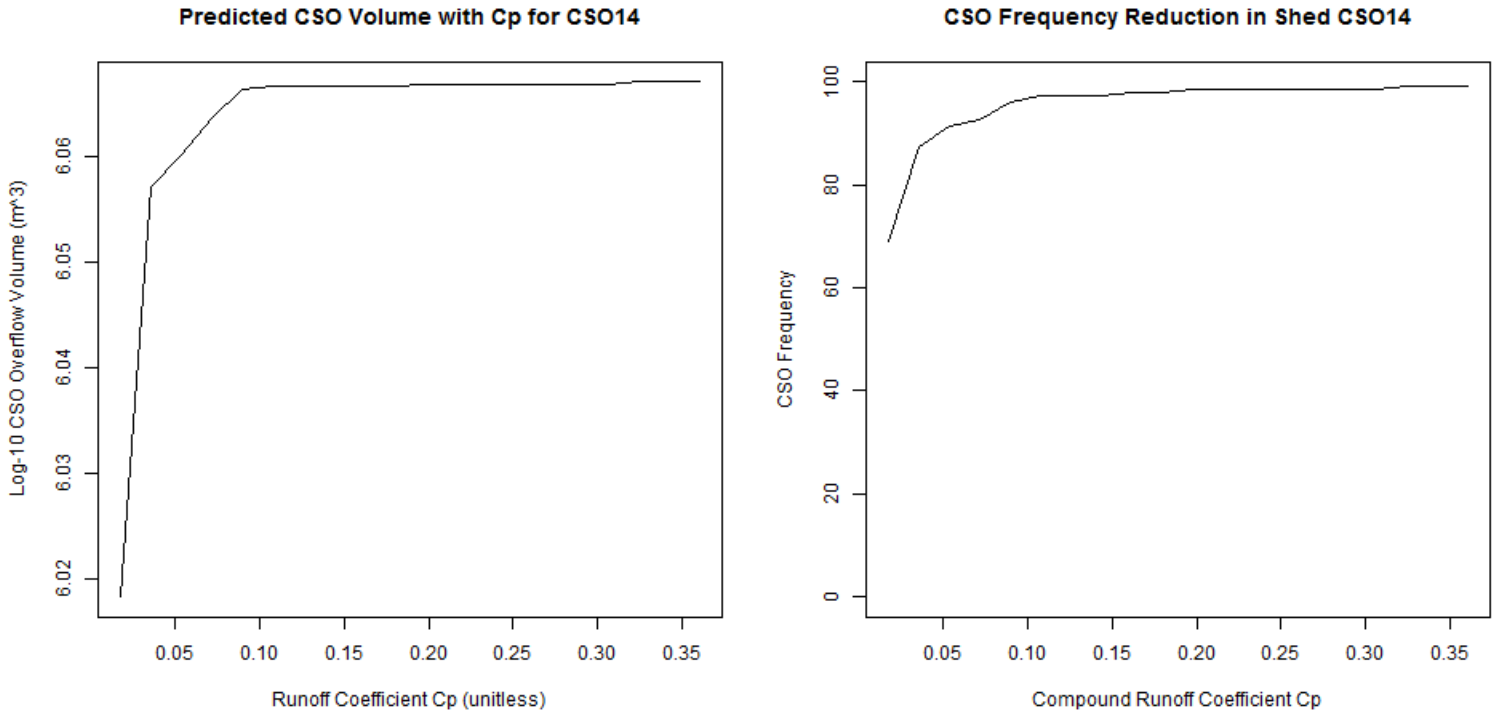


Figure 29: Predicted Change to CSO Parameters in CSO14

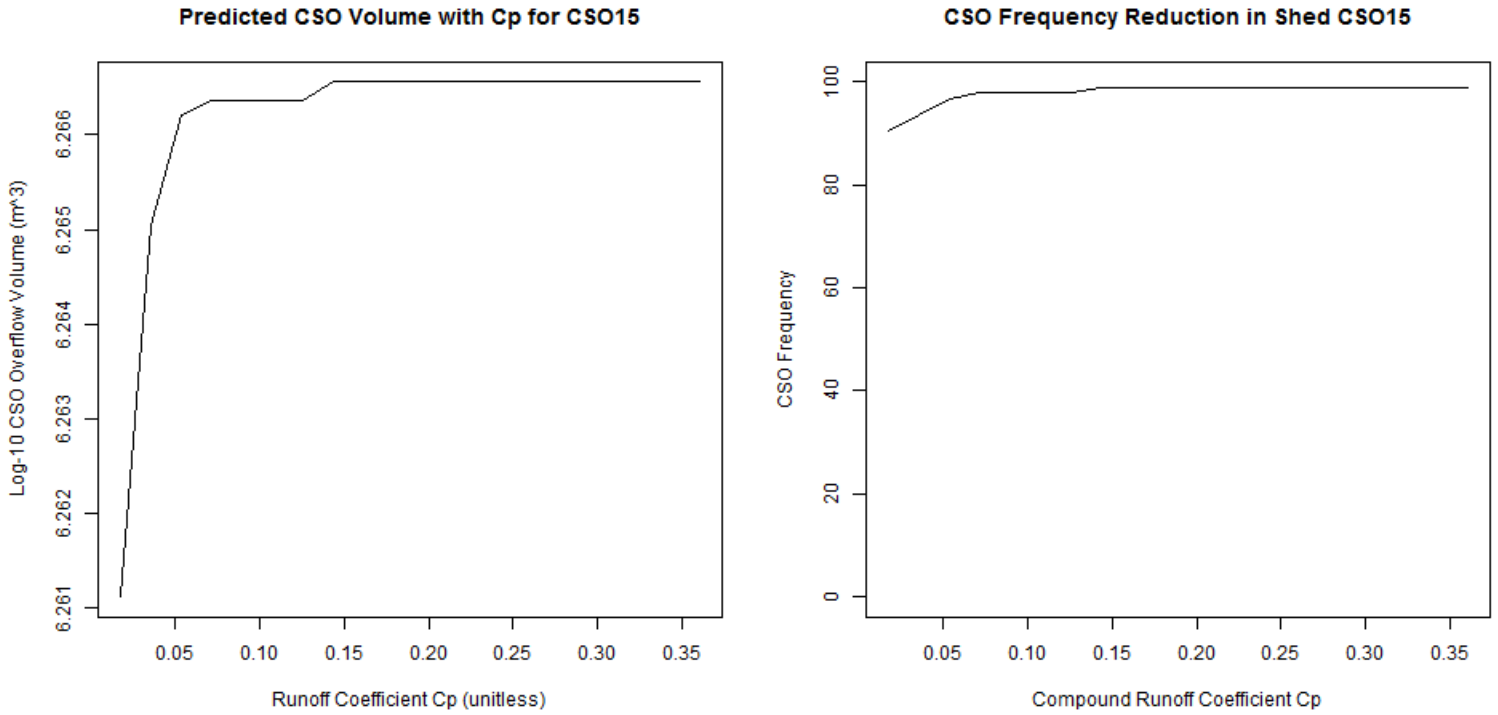


Figure 30: Predicted Change to CSO Parameters in CSO15

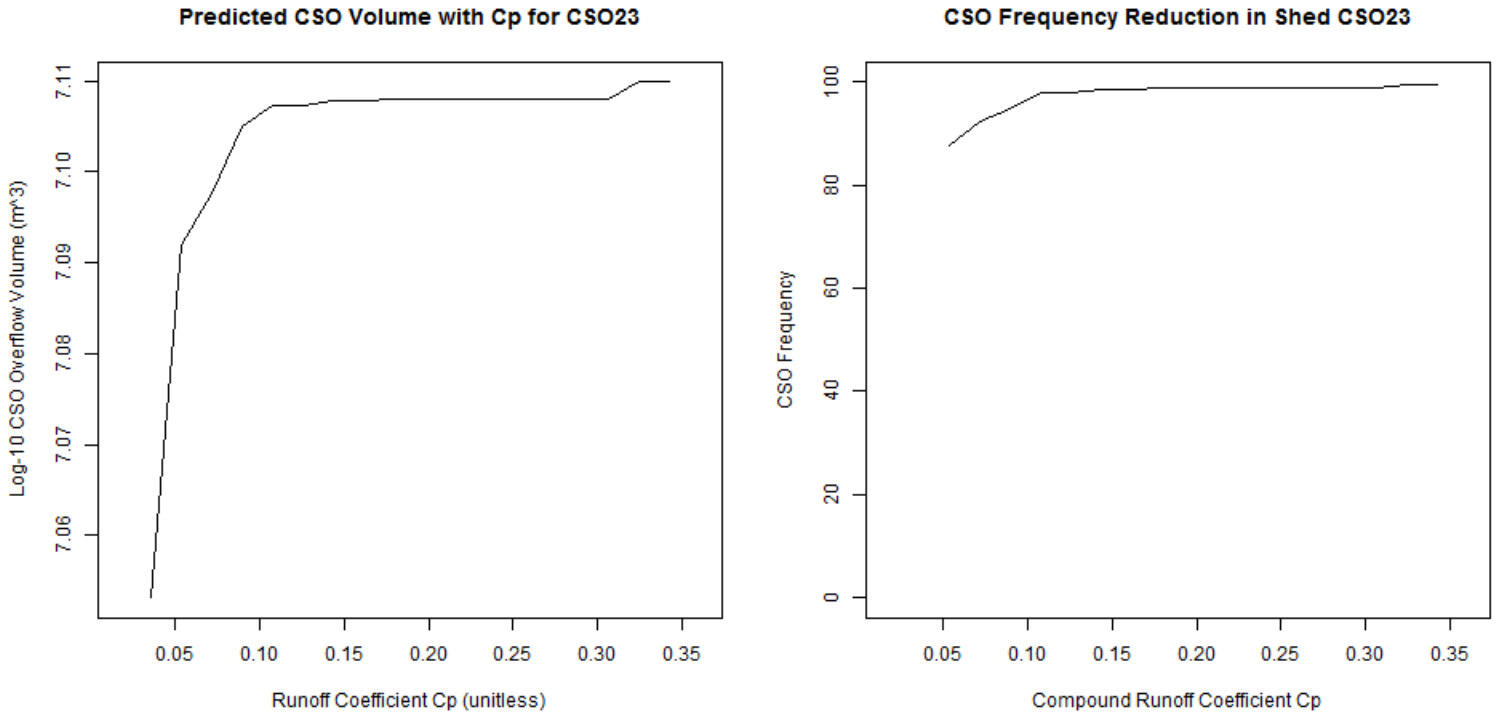


Figure 31: Predicted Change to CSO Parameters in CSO23

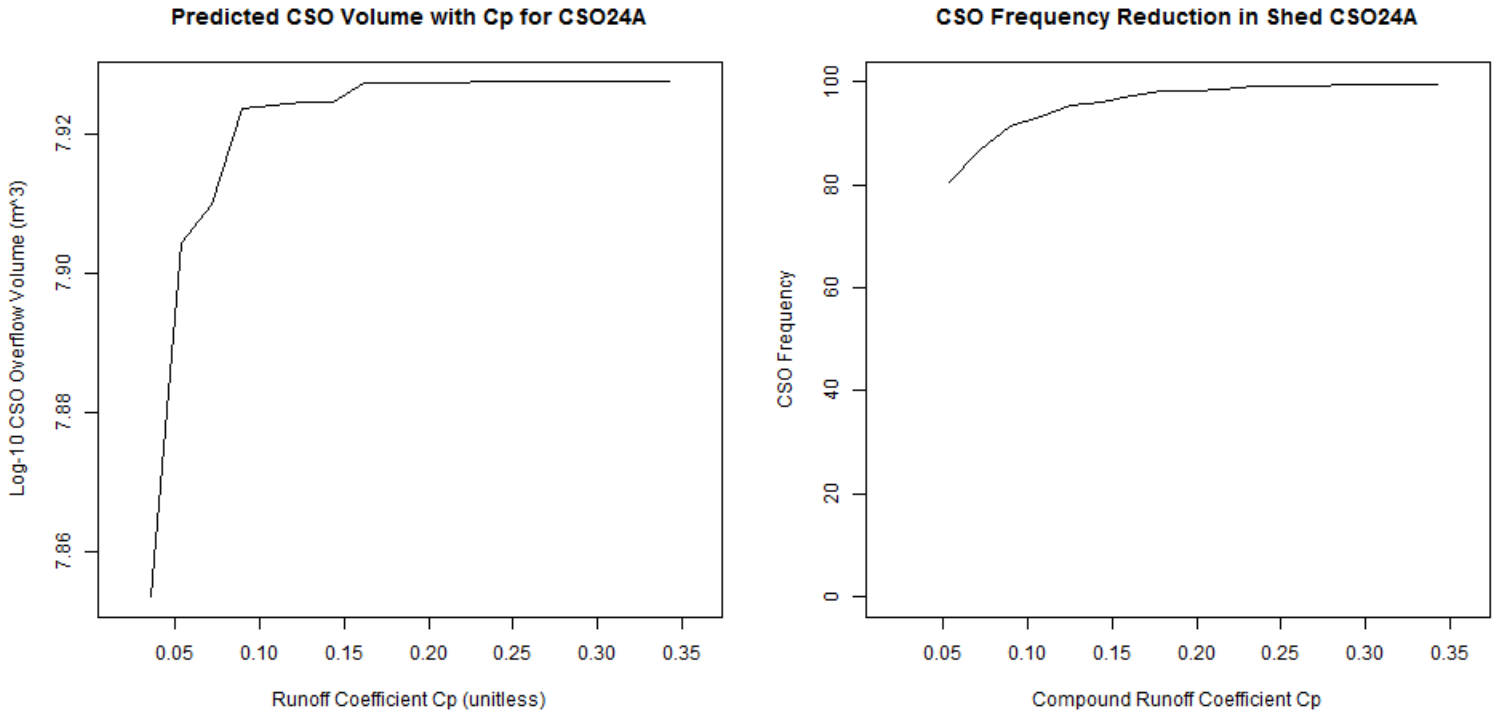


Figure 32: Predicted Change to CSO Parameters in CSO24A

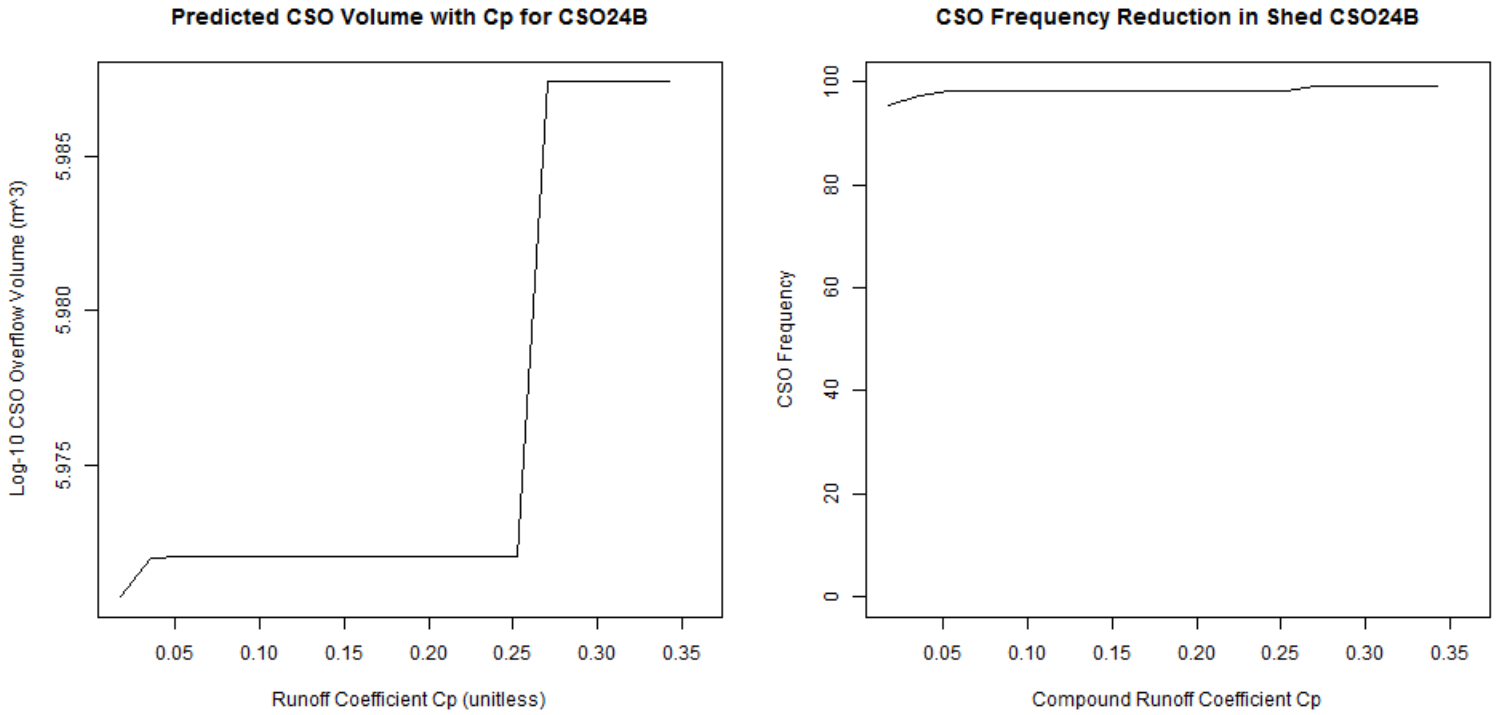


Figure 33: Predicted Change to CSO Parameters in CSO24B

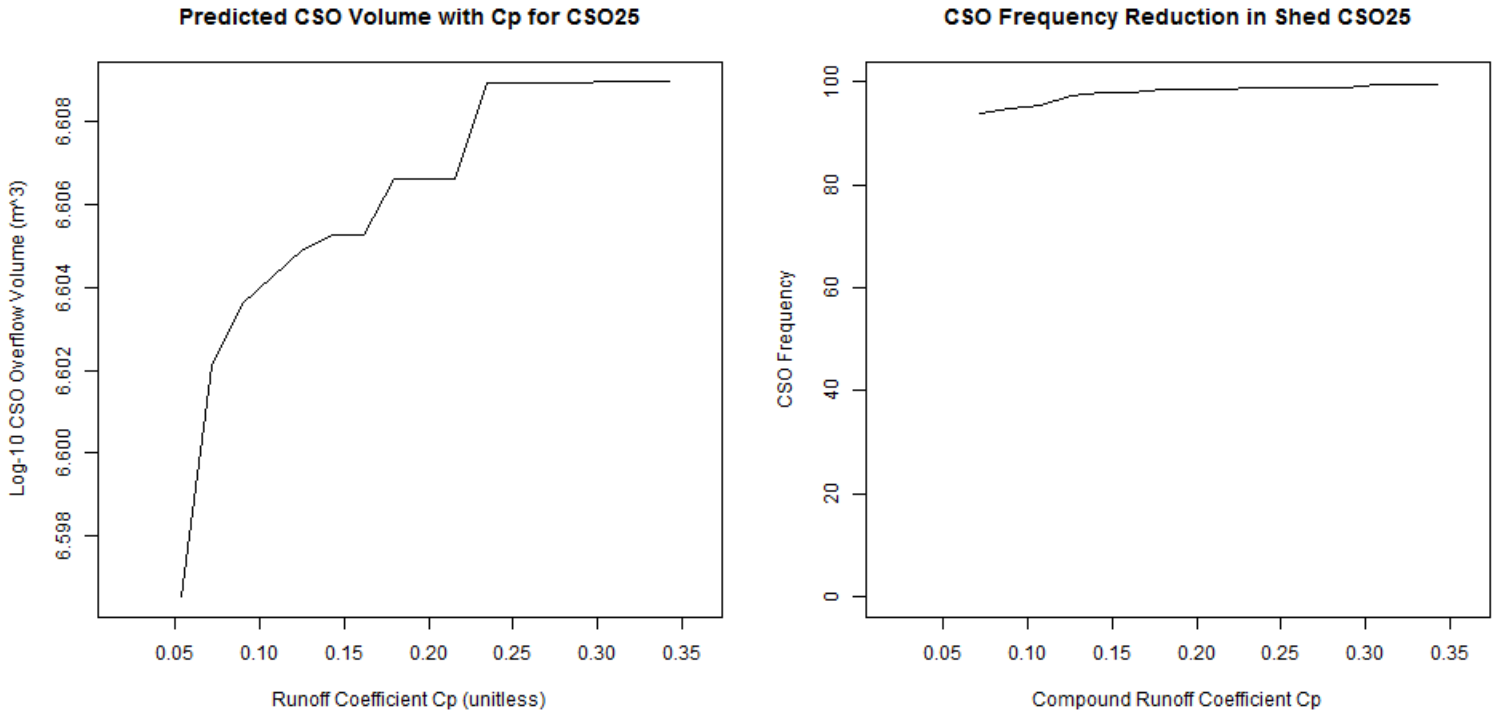


Figure 34: Predicted Change to CSO Parameters in CSO25

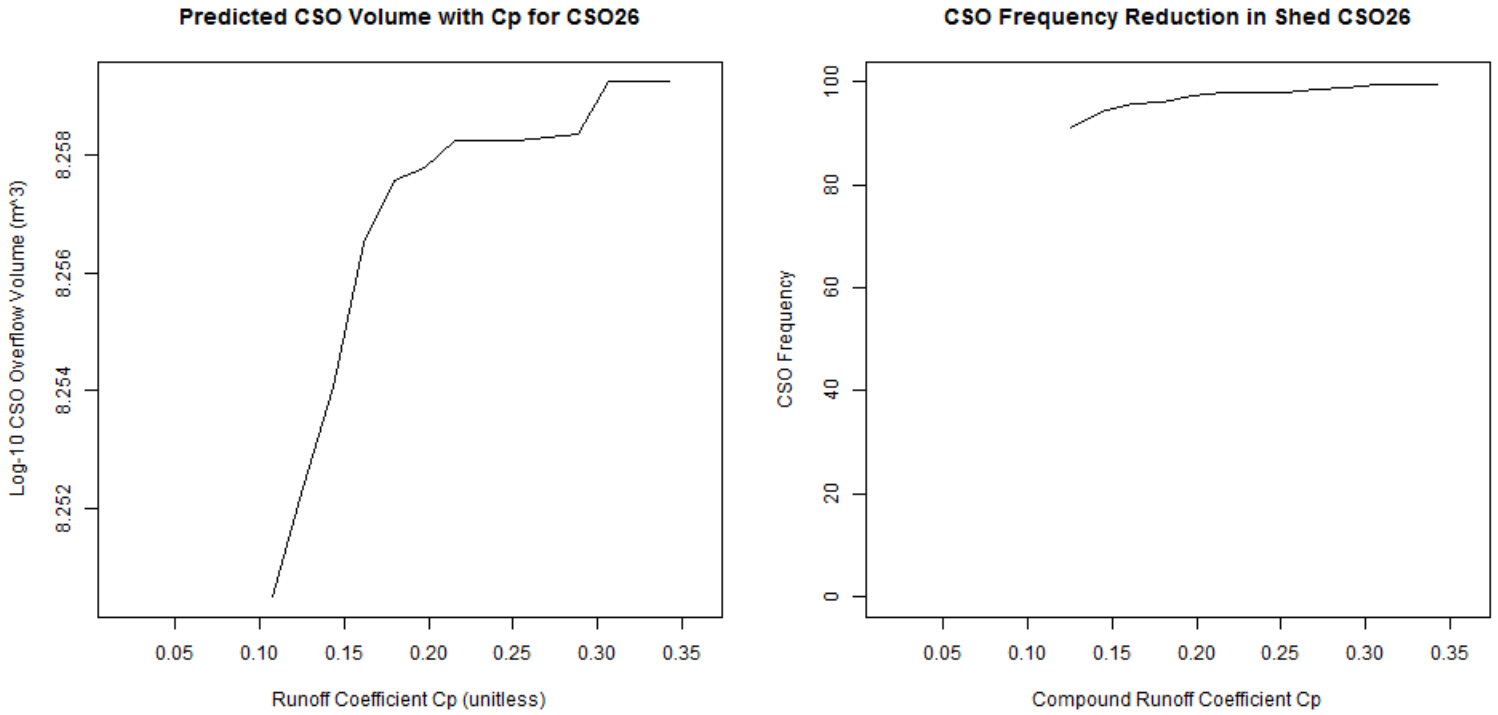


Figure 35: Predicted Change to CSO Parameters in CSO26

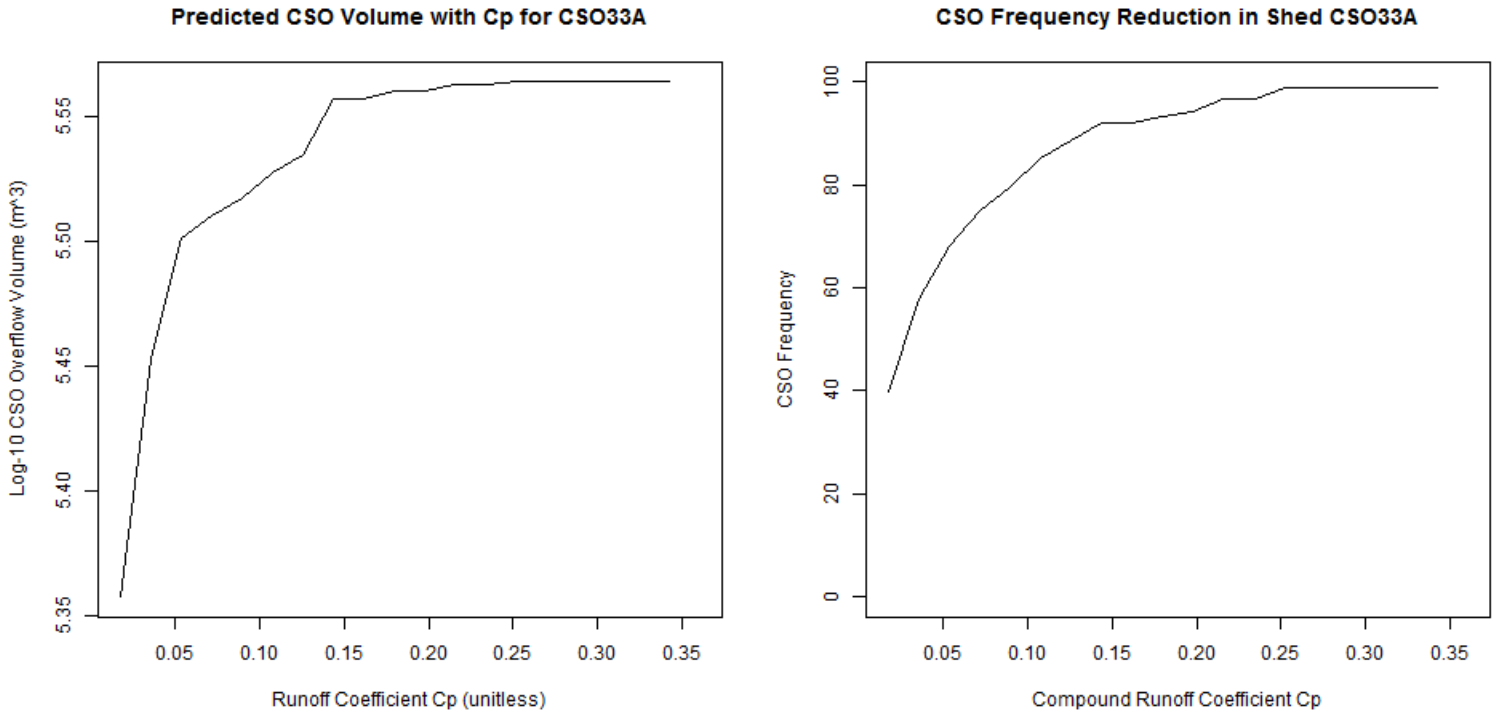


Figure 36: Predicted Change to CSO Parameters in CSO33A

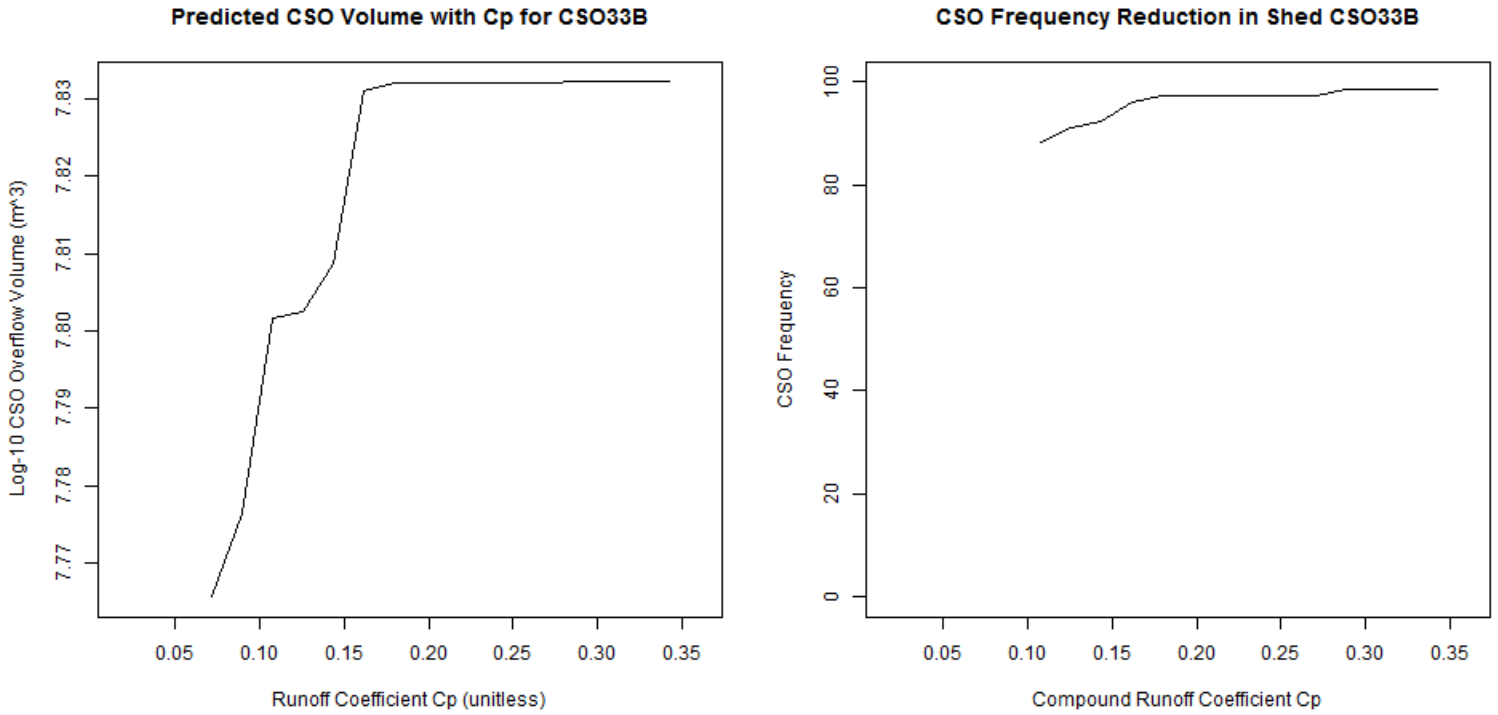


Figure 37: Predicted Change to CSO Parameters in CSO33B

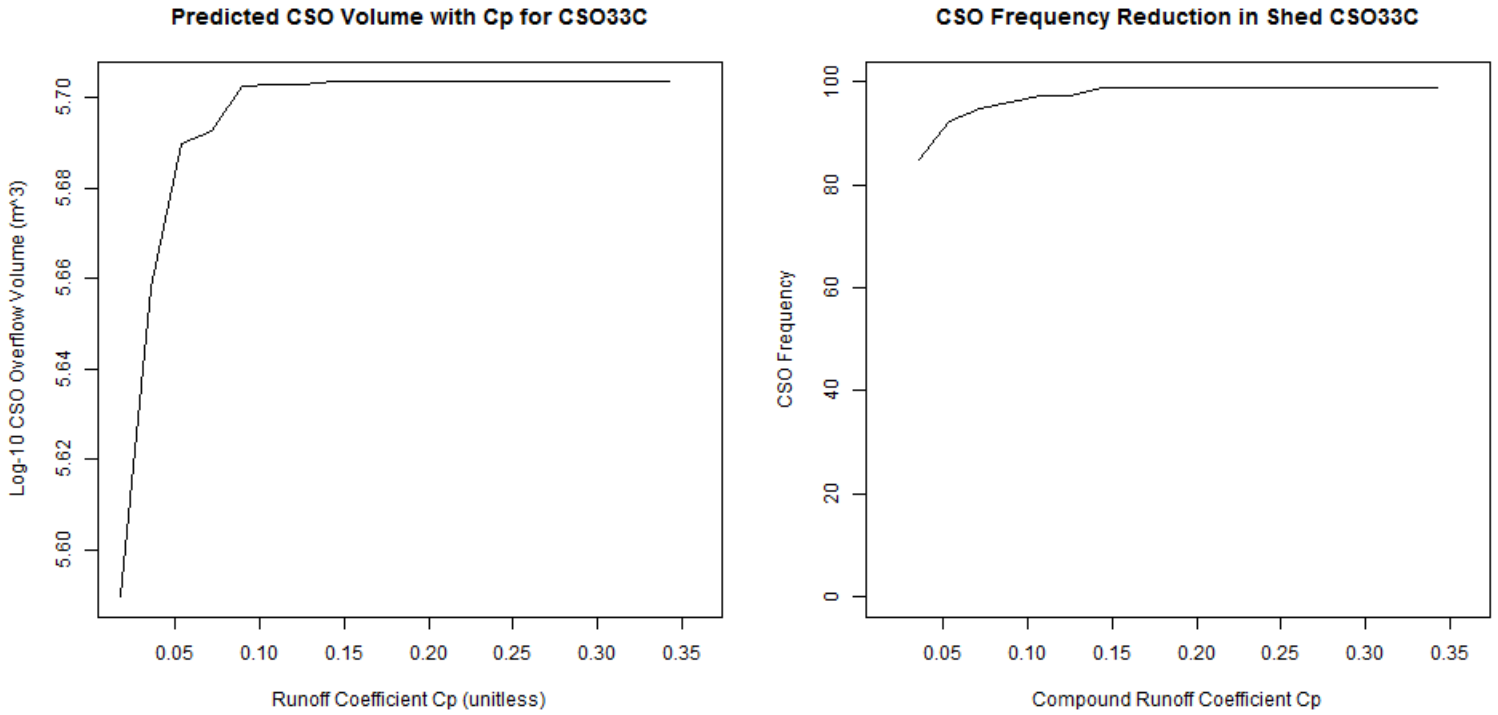


Figure 38: Predicted Change to CSO Parameters in CSO33C

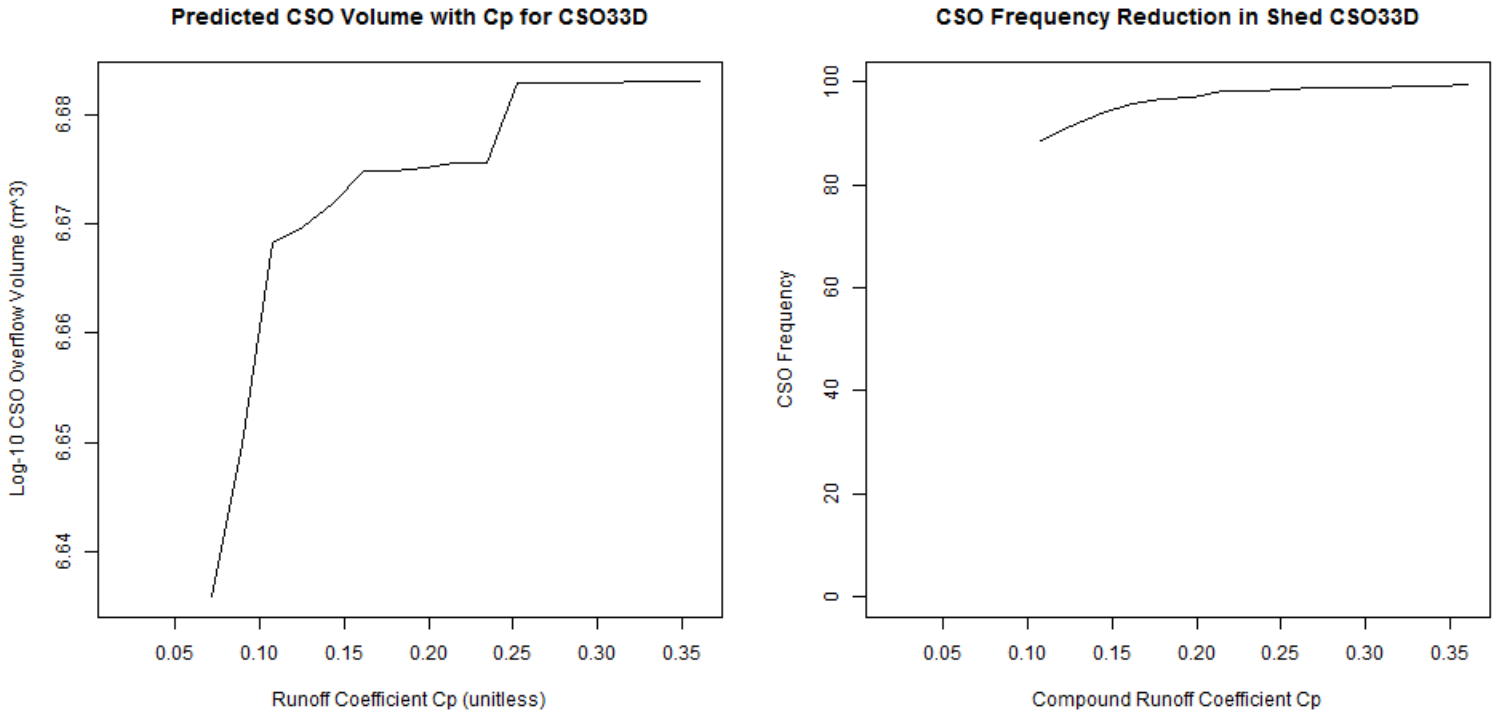


Figure 39: Predicted Change to CSO Parameters in CSO33D

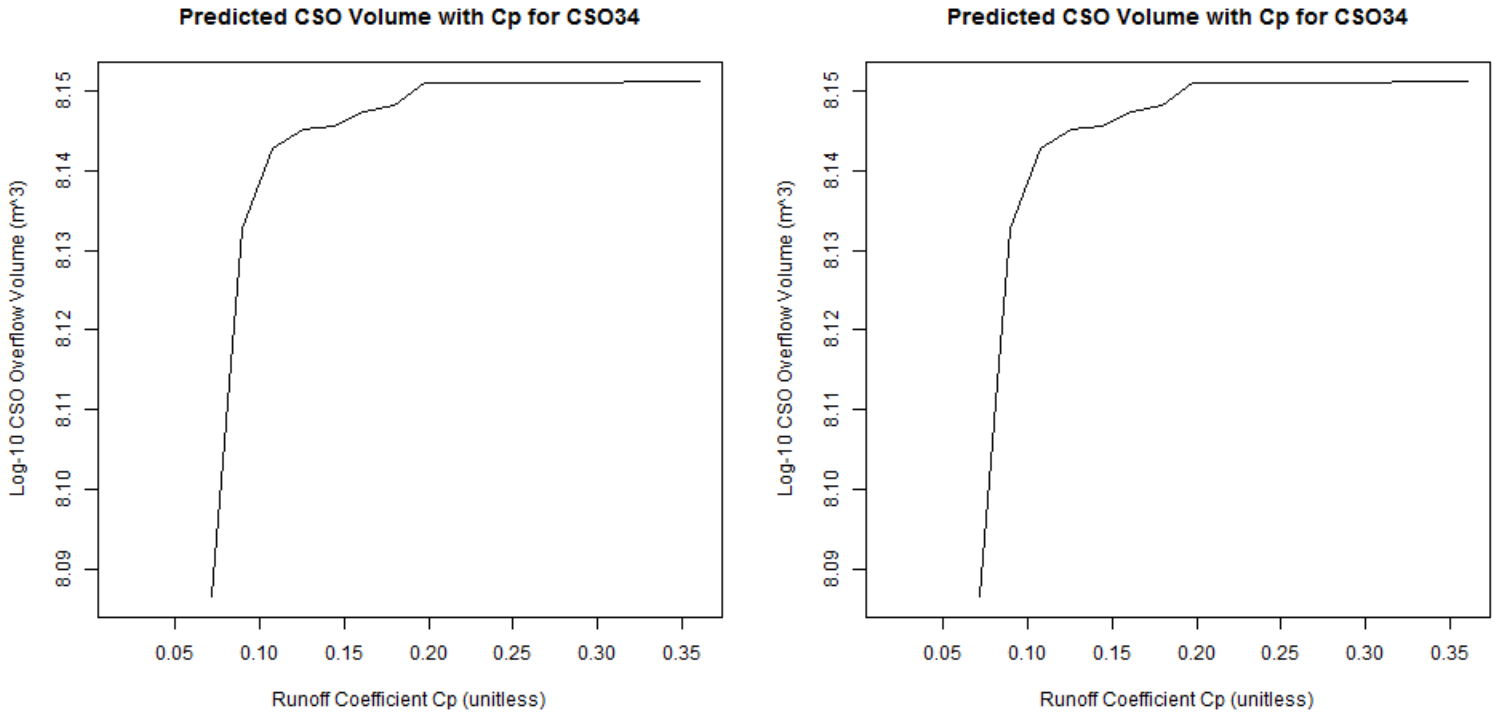


Figure 40: Predicted Change to CSO Parameters in CSO34

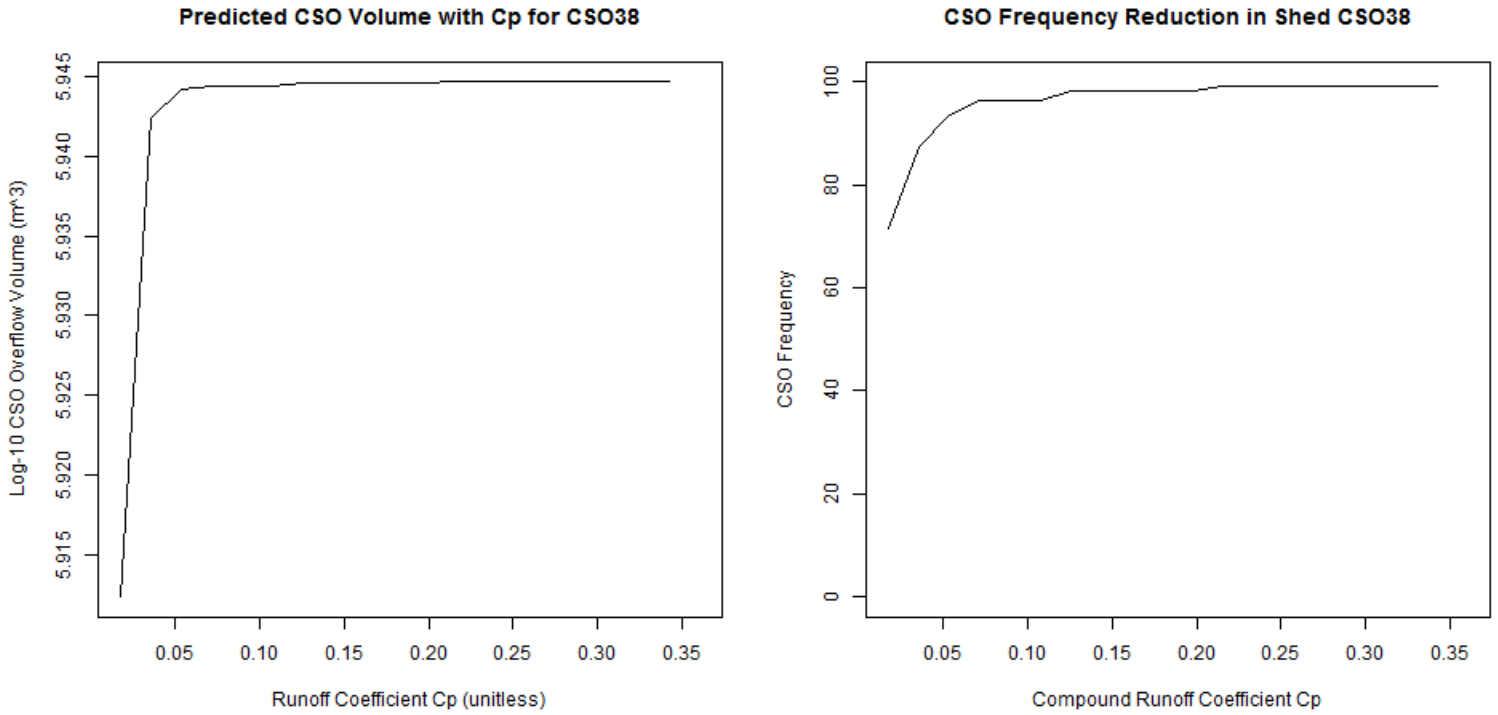


Figure 41: Predicted Change to CSO Parameters in CSO38

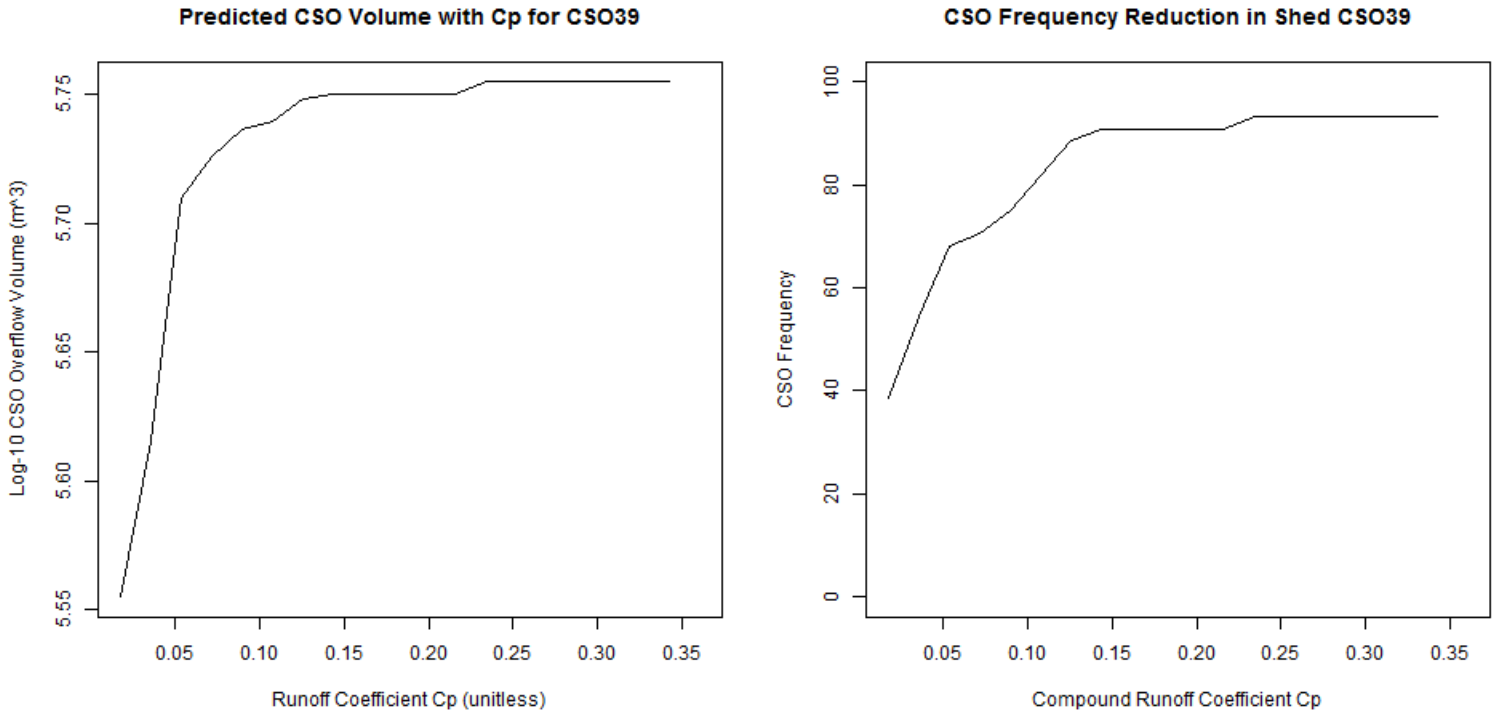


Figure 42: Predicted Change to CSO Parameters in CSO39

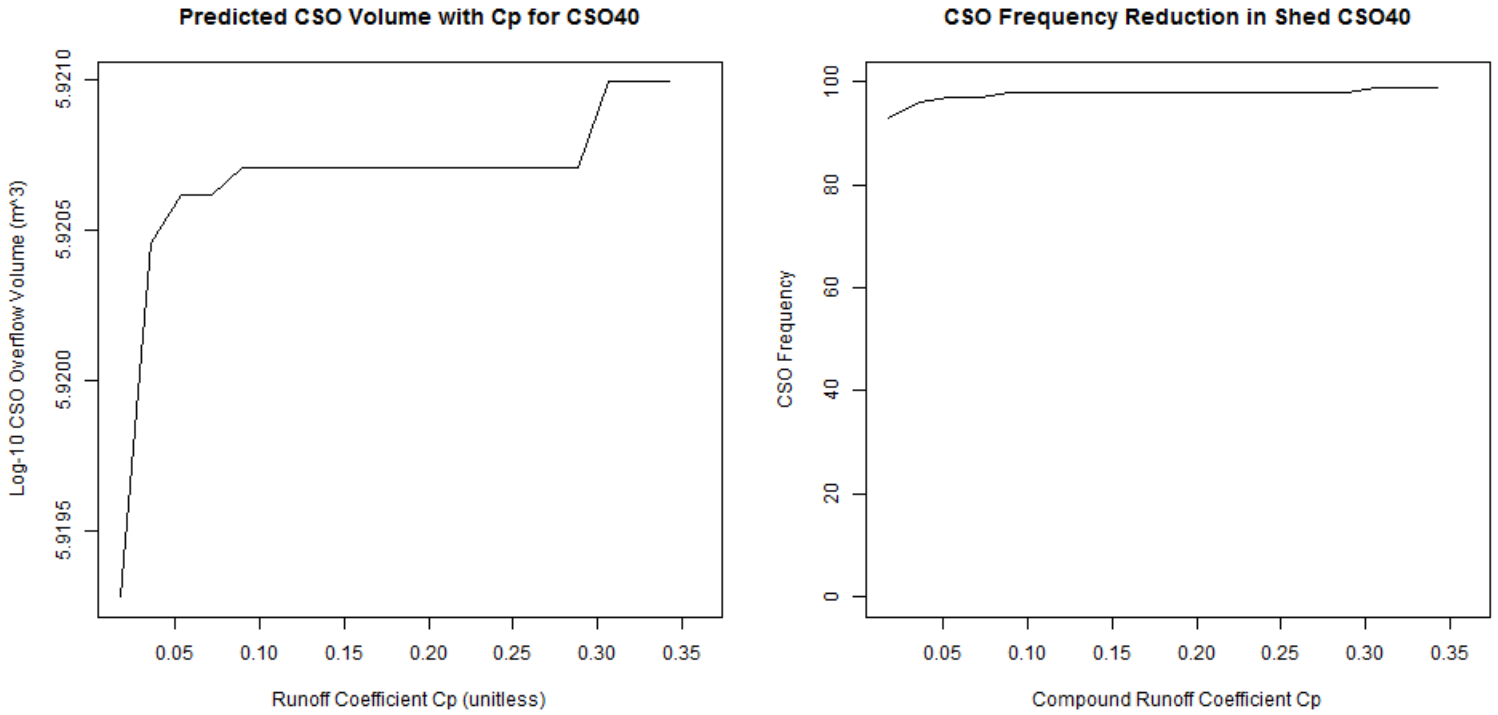


Figure 43: Predicted Change to CSO Parameters in CSO40

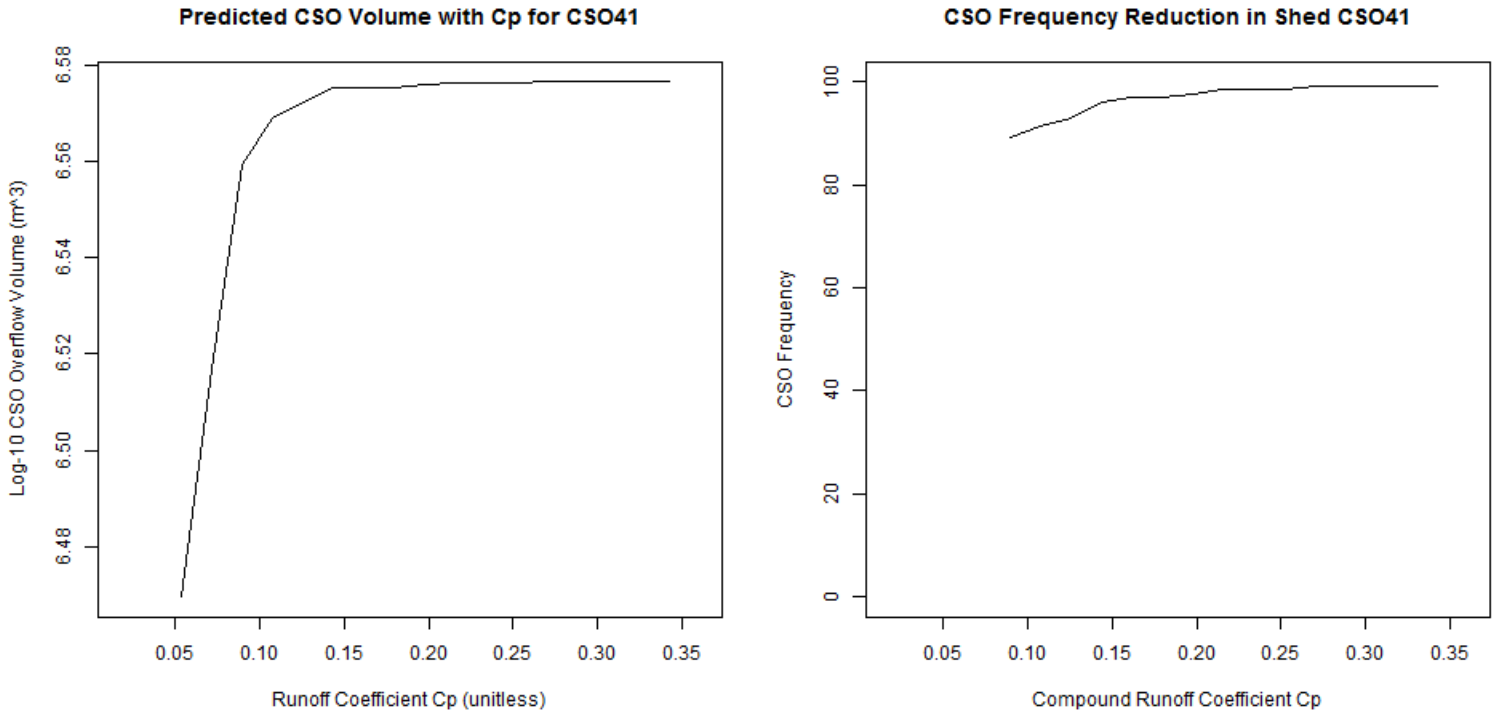


Figure 44: Predicted Change to CSO Parameters in CSO41

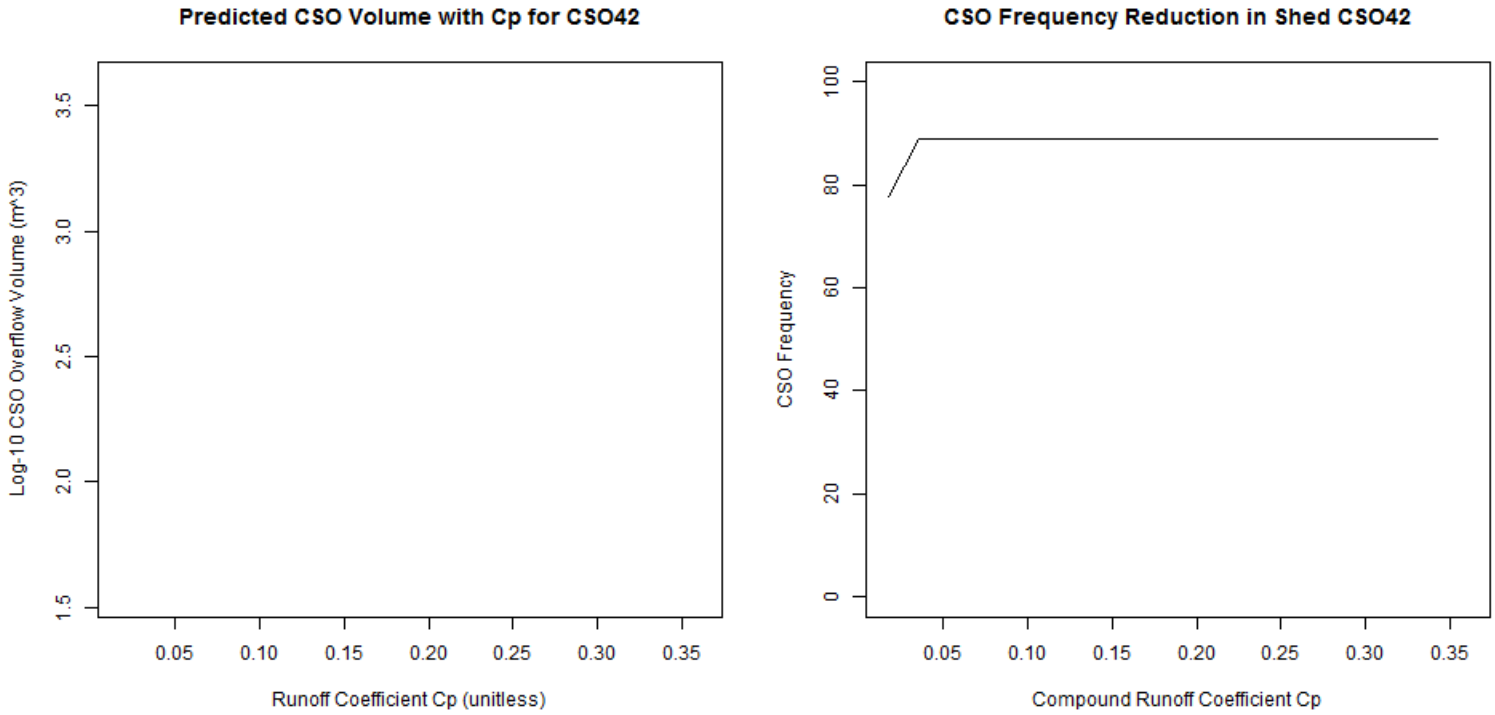


Figure 45: Predicted Change to CSO Parameters in CSO42