

Optimization Tool User Manual

Prepared by

SAN FRANCISCO ESTUARY INSTITUTE

4911 Central Avenue, Richmond, CA 94804

Phone: 510-746-7334 (SFEI)

Fax: 510-746-7300

www.sfei.org



Table of Contents

1. OPTIMIZATION TOOL OVERVIEW.....	1
1.1 What Is It.....	1
1.2 Optimization Algorithm.....	1
1.3 Formulation of Optimization Problem.....	3
1.4 Structure of Optimization Tool	5
2. INSTALL OPTIMIZATION TOOL	7
3. RUN OPTIMIZATION TOOL	7
3.1 Define Management Questions.....	7
3.2 Create Input Files.....	8
3.3 Run the Tool.....	8
3.4 Consideration of Optimization Run-time.....	9
4. REVIEW AND INTERPERATE OUTPUT	9
5. FUTURE UPGRADES.....	10
6. REFERENCES.....	12

OPTIMIZATION TOOL OVERVIEW

1.1 What Is It

The optimization tool of the GreenPlan-IT toolkit is used to identify the most cost-effective Green Infrastructure (GI) implementation scenario among the many options available in urban and developing areas to meet desired programmatic goals. The tool uses an evolutionary optimization technique, Non-dominated Sorting Genetic Algorithm II (NSGA-II), to systematically evaluate the benefits (runoff and pollutant load reductions) and costs associated with various GI implementation scenarios (location, number, type, and size of GIs) and identifies the most cost-effective options for achieving desired flow mitigation and pollutant reduction. Within the GreenPlan-IT toolkit, the optimization tool is structurally designed as a standalone module to provide flexibility for the user community, but running the tool requires the support from both the GIS Site Locator Tool (SFEI, 2015) and Modeling Tool (SFEI, 2015), because it needs the site information generated from the GIS Site Locator Tool as boundary conditions data and also interacts with the modeling tool during the search process to estimate the flow and load reduction associated with each of the iteratively varying GI scenarios.

The optimization tool is designed to help local watershed planning agencies to develop stormwater management plans and coordinate watershed-scale investments to meet their programmatic needs. It is intended for knowledgeable users familiar with GI and the technical aspects of watershed modeling. The tool outputs, combined with other site specific information and management requirements, can be used to develop watershed-scale Green Infrastructure master plans.

1.2 Optimization Algorithm

Belonging to the family of evolutionary optimization techniques, NSGA-II is one of the most efficient and widely used multi-objective optimization algorithms that is capable of producing optimal or near-optimal tradeoff solutions among competing objectives (Deb, et al, 2002). NSGA-II incorporates a non-dominating sorting approach that makes it faster than any other multi-objective algorithm. In NSGA-II, solutions are sorted on the basis of the degree of dominance within the population and a solution that is not dominated by any other solution has the highest ranking. In addition, the algorithm employs a crowded-comparison operator to preserve diversity along the Pareto optimal front so that the entire Pareto-optimal region is found (Srinivas and Deb, 1994). Successful applications of the NSGA-II in addressing environmental problems have been reported in EPA SUSTAIN applications (USEPA 2009) as well as many case studies (Bekele and Nicklow, 2007; Bekele, et al, 2011; Maringanti, et al, 2008; Rodriguez, et al, 2011). The major operational steps of NSGA-II are described below.

- **Creation of first generation**

The algorithm begins with random generation of an initial population of potential solutions. The population is sorted based on the concept of Pareto dominance (non-domination) into each front. A solution is non-dominant to another solution when it performs no worse than the other solution in all objectives, and better than the other solution in at least one objective. The first front is the complete non-dominant set in the current population, the second front is dominated by the solutions in the first front only, and so on. At the end of the sorting, each solution is assigned a fitness (or rank) equal to its non-dominant level, with a smaller value indicating that the solution is dominated by fewer other solutions (1 is the best level, 2 is the next-best level, and so on). In addition, a crowding distance, defined as the size of the largest cuboid enclosing a solution without including any other solution in the population, is calculated for each individual solution as a measure to maintain solution diversity. Large average crowding distance results in better diversity in the population.

- **Optimization Process**

In the first step of the optimization process, parent populations are selected from the population by using binary tournament selection based on the rank and crowding distance. An individual solution is selected if the rank is lesser than the other or if the crowding distance is greater than the other. The selected population generates an offspring population of the same size from the processes of crossover and mutation. The combined population of the current parent and offspring is then sorted again according to non-domination and only the best N individuals are selected to form a new parent population, where N is the population size. Elitism is ensured in this step because both the parent and the child populations are used in the sorting. Comparison of the current population with previously identified non-dominated solutions will be performed at each iteration. The new parent population is then used to create a new child population, and the process continues until the stopping criteria are met. At that point, the optimization is considered converged at an optimal front and the process can be stopped.

- **Stop Criteria**

The users can stop the NSGA-II using some user-defined stop criteria. The commonly used criteria include maximum number of iterations; no change in the new parent population for two consecutive loops; and no tangible improvement for the fitness function after a certain number of iterations. Usually, between 100 to 500 iterations is sufficient for the NSGA-II to converge to the Pareto-optimal front. Too large a number will take very long run time and very small number will stop the NSGA-II from reaching the global Pareto front to the problem.

1.3 Formulation of Optimization Problem

The objective of the optimization tool is to determine GI locations, types, and design configurations that minimize the total cost of management while satisfying water quality and quantity constraints. The development of the tool started with the formulation of a two-objective optimization problem that can be solved through the NSGA-II. In formulating the optimization problem, three key decisions need to be made: defining optimization objectives, determining decision variables, and identifying associated constraints.

- **Optimization Objectives**

For a GI project, the first optimization objective is usually defaulted to minimize the cost of GI implementation. Depending on management questions, the second optimization objective can be related to either water quality or quantity. The typical evaluation factor include peak flow, total runoff volume, average annual runoff, exceeding frequency, annual average loads, annual average concentration, and maximum days average concentration. At the current setting of the optimization tool, total runoff volume is selected as the evaluation factor, and therefore the second optimization objective is defined as maximizing runoff volume reduction.

- **Decision variables**

GI locations and design (size) affect the overall cost-effectiveness of a stormwater control system and represent important decision variables for optimization. Currently, three GI types are built into the optimization tool: Bioretention, Infiltration Trench, and Permeable Pavement at the recommendation of stakeholders. A typical design was employed for each GI type and these configurations remain unchanged during the optimization process. Table 1-1 summarizes key design parameters for each GI feature. Users can change these configurations during the creation of the SWMM template file (discussed in Section 3) to fit their specific needs.

Table 1-1 GI configurations used in optimization tool

Parameter	Bioretention	Infiltration Trench	Peamable Pavement
Physical Configuration			
Surface area (ft ²)	1000	500	5000
Surface depth (in)	12	12	N/A
Growing media depth (in)	18	N/A	N/A
Storage depth (in)	12	36	12
Underdrain	Yes	Yes	Yes
Infiltration			
Suction head (in)	24	N/A	N/A
Saturated conductivity (in/hr)	5	N/A	N/A
Effective porosity	0.437	N/A	N/A
Field capacity	0.105	N/A	N/A
Wilting point	0.047	N/A	N/A
Void ratio	0.5	0.5	0.18
Bottom filtration rate (in/hr)	0.1	0.1	0.1
Underdrain infiltration rate (in/hr)	0.14	0.14	0.14
Initial media saturation (%)	15	N/A	N/A
Peamability (in/hr)	N/A	N/A	100

Since the tool uses a typical design for each applicable GI type, the decision variables are therefore defined as the number of fixed-size units of the distributed GI types. The total number of decision variables = No. of basins * 3 of GI types.

- **Constraints**

For each GI type, the number of possible sites is constrained by the maximum number of potential sites identified by the GIS locator tool. The decision variables are also constrained by the total area that can be treated by GIs within each sub-basin. Through the discussion with the project Technical Advisory Committee (TAC), a 4% “rule of thumb” - defined as GI design at four-percent of the project area to capture 100% stormwater volume of a design storm, was used to size GI for this study. During the optimization process, the combined numbers of GIs are forced to be less or equal to the maximum numbers calculated by applying this 4% rule.

Mathematically, the optimization problem formulated in this tool can be expressed as:

$$\text{Minimize } \sum_{i=1}^n \text{cost}(LID_i)$$

Maximize runoff volume reduction

Subject to: $n \leq N_{max}$

where GI_i = a set of GI configuration decision variables associated with location i

N_{max} = maximum number of feasible sites

1.4 Structure of Optimization Tool

Structurally, the optimization tool is built around three functionalities each supported by a group of subroutines: Optimizer, GI performance evaluator, and cost calculator. The optimizer serves as the engine that generates GI scenarios and drives the search process through NSGA-II algorithm. The GI performance evaluator creates input files that incorporate the generated GI scenarios for the modeling tool, runs the model with new input files, and passes GI performance (flow and/or pollutant load reduction) for each of the generated GI scenarios back to the optimizer. The cost calculator estimates cost of each GI scenario and also passes that information to the optimizer.

The three function groups work hand-in-hand in an iterative and evolutionary fashion to search for the best GI solutions that meet the user's specific conditions and programmatic goals. During the search process, new GI scenarios are generated through the optimizer within the search space defined by maximum feasible locations that are identified by the GIS Site Locator tool, the GI performance are evaluated, and costs are estimated. The optimizer then systematically compares the cost and performance data and modifies the search path to generate a new set of viable GI options and repeats the process until the set criteria to end the iteration are reached. Figure 1-1 presents a conceptual overview of the tool.

The optimization tool was written in FORTRAN 90/95. The coding of NSGA-II were verified with example problems provided at Deb, et al (2002) to make sure the optimization algorithm was implemented correctly.

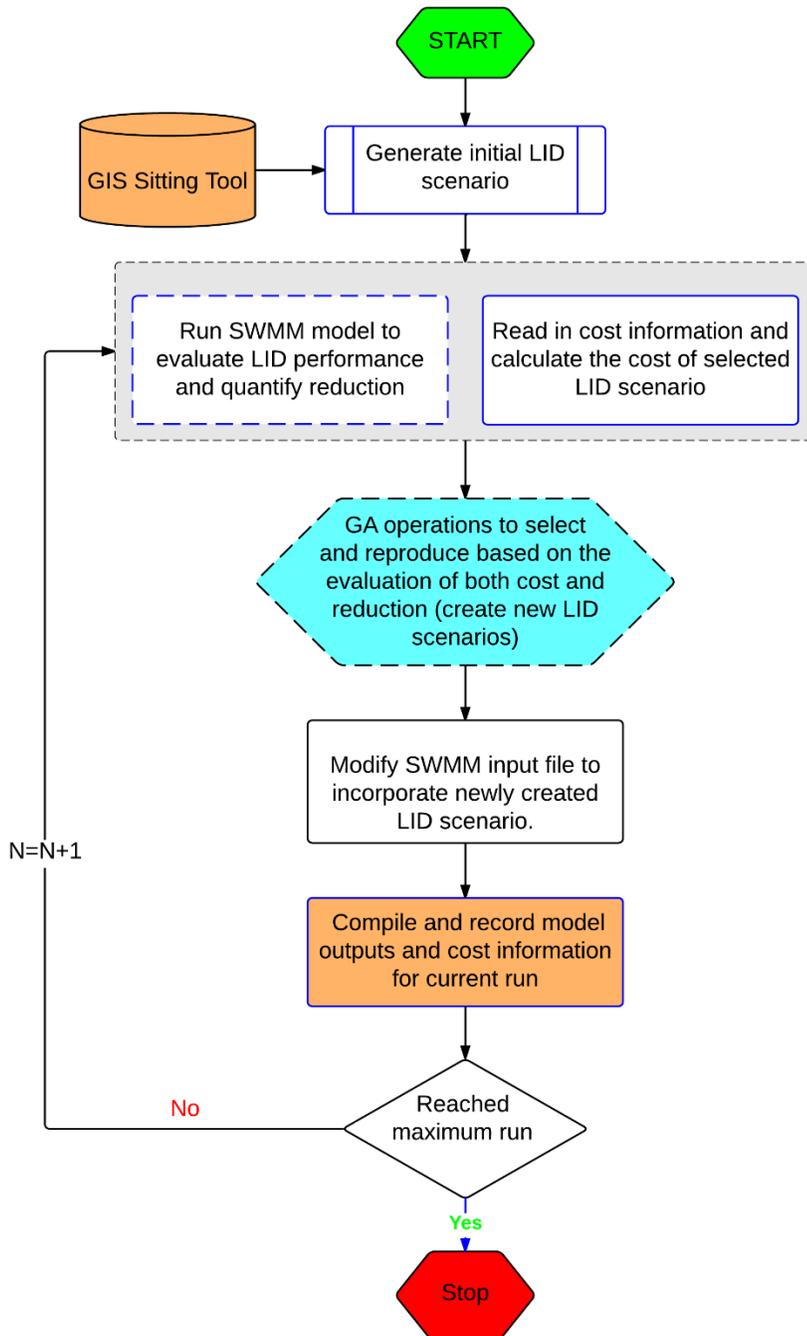


Figure 1-1 Flowchart of Optimization Tool

2 INSTALL OPTIMIZATION TOOL

The optimization tool is designed to run under the Windows 2003/2013/NT/XP/Vista/7 operating system of an IBM/Intel-compatible personal computer. The minimum requirements for hardware are a 400 MHz Pentium processor, 128 MB Ram, and 10 GB of available space. At this time it has not been compiled to run on Apple computers or in other operating systems.

To install the optimization tool and to get ready to use it on your PC, follow these steps:

1. Check that your Windows PC meets the system requirements.
2. Create project directory/folders to store the tool and input/output files. The main project folder can be under any root directory with any user-defined name. An example might be: “*c:\My Folders\Optimization*”. Within the main folder, create three sub-folders as:
 - .\Binary*
 - .\Setup Files*
 - .\Outputs*
3. Download the tool package from the GreenPlan website: <http://greenplanit.sfei.org/books/toolkit-downloads>. Save the executable in the **Binary** folder and example input files in the **Setup Files** folder.
4. Save a DOS version SWMM5 executable (swmm5.exe) in the **Outputs** folder.

3 RUN OPTIMIZATION TOOL

The optimization tool, as it currently stands, can be run as a console application from the command line within a DOS window. The steps of running the tool are as follow.

3.1 Define Management Questions

The application of optimization tool began with the definition of management problem, which involves selecting assessment point and determining programmatic management goals.

- **Assessment point**

An assessment point is the location in the study basin where runoff or pollutant loading reduction will be evaluated relative to optimization goals. The assessment point can be at the watershed outlet, key tributary outlets, or the downstream node of a stream segment.

- **Management targets**

Management targets can be based on flow and/or water quality and are specified by users for each assessment point. For example, a management target for flow can be a desired reduction in average annual volume, peak discharge, or exceedance frequency. Different management targets will results in different sets of optimal solutions.

3.2 Create Input Files

Three input files are required to run the optimization tool. Each file needs to be created according to the specific format provided by the template files. All of them should be stored in the folder `./Setup Files`.

- **GI_Maxnumber.csv.** This file contains both acreage of each sub-basin and maximum number of feasible sites for each GI type within each sub-basin. The acreage of each sub-basin is usually obtained from watershed delineation. The maximum number of possible GI sites are identified through the GIS Site Locator Tool given the typical size of each GI type and used to constrain the optimization space. Depending on the size of the study area, the resolution and ranking of GIS data in the Site locator tool, and size of each GI type, these numbers could display a wide range.
- **GI_setup.csv.** This file contains three GI configuration parameters and unit cost of each GI type, all can be user-defined. The GI parameters are surface area for each GI unit, top width of overland flow surface of each unit, and % of initial media saturation. The rest of GI parameters are specified in the baseline SWMM input file. At current setting, each GI type is assigned with typical configuration parameters. Users can change any of these parameters to reflect their design criteria or fit their specific needs.

The unit cost should be specified as \$/square feet. Implementing GI at the landscape scale would incur many costs ranging from traffic control, construction, to maintenance and operation. At minimum, the costs considered include construction, design and engineering, and maintenance and operation. The reliable cost information for each GI feature is critical for identifying optimal solutions, because the optimization process and solutions are highly sensitive to the cost function. Therefore, wherever possible, local cost information should be collected and used to construct this file.

- **SWMM_base.inp.** Stormwater runoff from the modeling tool (SWMM) is the forcing function that drives GI simulation. A baseline SWMM input file needs to be created to serve as the reference point from which generated GI scenarios will be measured. Under normal circumstances, the baseline file is created from model calibration to local watersheds. The majority of GI configuration parameters as listed in Table 1-1 are also specified in this file (GI_CONTROLS section). By default, Windows version of SWMM5 should be used for both model calibration and GI specifications, because of its easy-to use interfaces and many features such as an integrated environment for editing study area input data; running hydrologic, hydraulic and water quality simulations; and viewing the results in summary reports, time series graphs and tables, and statistical frequency analyses.

3.3 Run the Tool

Currently, the tool is set up to run in two ways, both relatively straightforward:

- Open the *Binary* folder in Windows Explorer, and click the *OPTIM.exe* OR
- Open a DOS window, change the working directory to *./My Folders/Optimization/Binary*, and type *OPTIM*.

At current setup, input/output files are hardcoded in the programs to expedite the tool development, so no inputs are required for the user. The key NSGA-II parameters are also hardcoded, with number of generation = 200, population size =100, crossover probability=0.9 and mutation probability =0.1. This setup is not ideal because it doesn't provide users with enough flexibility to use the tool. During the next phase of tool development we will strive to make the tool more flexible to allow users to 1) pass the paths and names of input/output files into the tool as an argument; and 2) determine key NSGA-II parameters such as the number of iterations and population size.

3.4 Consideration of Optimization Run-time

The total number of iterative runs needed for the optimization process to converge to the optimal solution is dependent on the number of decision variables, model simulation period, and the complexity of the model (number of sub-basins and stream network). More model runs usually leads to longer computation time. If the optimization tool is applied to a large watershed with many feasible GI sites and complex stream network and the optimization process is based on long-term continuous simulation, a large computation time will be needed to reach the optimal solution. In general, the computational efficiency can be achieved through reducing the number of decision variables, simulation time, and complexity of the problem.

4 REVIEW AND INTERPERATE OUTPUTS

The optimization tool produces a text file that contains reduction and cost information for all solutions and also saves SWMM input/report files for all intermediate solutions. All these files are stored in the folder *./Outputs*. The number of intermediate SWMM files can be in the order of tens of thousands, and thus, large hardware space is needed to store these files.

- **cost_reduction.txt.** This is the main output file that contains generation#, population# within each generation, cost, and runoff volume or load reduction in percentage. Users can grab the information here to make the cost-reduction curve using Excel or other graphing software. The reduction information in this file is needed to help you identify the optimal solution associated with a specific reduction goal. For example, if a 30% reduction is desired, users can go to the file and find the generation # and population# with reduction closest to 30%. The optimal solution can then be extracted from the SWMM input file with the identified generation # and population#.

- **SWMM input/report files.** The SWMM input file (.inp) and report file (.rpt) for each generated GI scenario was saved to keep a record of the optimization process, and most importantly to track the optimal combinations of GI number and types. The files are named as *SWMM_base_g#_p#.inp* and *SWMM_base_g#_p#.rpt*, respectively, where g represents generation, p as population and # as the number of generation and population. For example, *SWMM_base_g002_p080.inp* represents SWMM input file for generation 2 and population 80. Once users identify the targeted .inp file, they can click open the file and mine the optimal GI solutions from the [GI_USAGE] section. The runoff, pollutants, and GI performance information for the entire basin as well as each individual sub-basins can be viewed and extracted from corresponding .rpt file.

It is important to emphasize that users must interpret the optimization results in the context of specific problem formulation, assumptions, constraints, and optimization goals unique to their study. If one or more assumptions are changed, the optimization might have resulted in a completely different set of solutions in terms of GI selection, distribution, and cost. It also should be noted that because of the large variation and uncertainty associated with unit GI cost information, the total cost associated with various reduction goals calculated from the unit cost do not necessarily represent the true cost of an optimum solution for the basin evaluated and are not transferable to other basins. Rather, these costs should be interpreted as a common basis to evaluate and compare the relative performance of different GI scenarios. The optimization tool provides a framework to identify optimal solutions for addressing stormwater management issues at a watershed scale. Its application must be preceded by an intimate understanding of the study area and the influential factors affecting stormwater of the study area.

5 FUTURE UPGRADES

To ensure the optimization tool is comprehensive and flexible enough to handle a variety of situations/questions, the tool needs to be continuously evolving. With further stakeholder consultation, possible future enhancements to the optimization tool include:

- **More flexibility**

At current setup, the input file names/paths and many important decision variables such as the total number of iterative runs and the size of population were predetermined and hardcoded to expedite the tool development. The next phase of the tool development could make key decision variables of an optimization problem as user-defined inputs to provide flexibility for broad applicability. Also, the tool is currently tailored toward stormwater volume reduction, future upgrades could enable the tool to handle a variety of management targets on both water quantity and quality.

- **More GI types**

Only three GI types are included in the optimization tool at present stage. During the next stage of development, more GI types as well as centralized regional facilities such as enlarged bioretention could be included in the mix to develop a diverse set of management options to a wide range of stormwater management problems.

- **New programming language**

The optimization tool is currently written in FORTRAN 90/95, which is primarily intended for numerical computation and remains one of the best computational languages ever developed. Although there is a vast body of existing FORTRAN code in the environmental field (much of which is publicly available and of high quality), FORTRAN certainly shows its age and its use is now limited to maintenance of and interfacing with legacy software. A future enhancement may include the migration of the current FORTRAN code to a modern programming language such as Python or C++ to make it more compatible with current trends and development in computer programming and easier to interface with other tools and software.

6 REFERENCES

- Bekele, E. G. and Nicklow, J. W. (2007). Multi-objective automatic calibration of SWAT using NSGA-II, *Journal of Hydrology*, 341, 165– 176.
- Bekele, E. G., Lian, Y. and Demissie, M. (2011). Development and Application of Coupled Optimization-Watershed Models for Selection and Placement of Best Management Practices in the Mackinaw River Watershed, Illinois State Water Survey, Institute of Natural Resource Sustainability, University of Illinois at Urbana-Champaign.
- Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Transactions on Evolutionary Computation*, 6(2) (2002) 182-197.
- Maringanti, C., Chaubey, I., Arabi, M. and Engel, B. (2008). A multi-objective optimization tool for the selection and placement of BMPs for pesticide control, *Hydrol. Earth Syst. Sci. Discuss.*, 5, 1821–1862.
- Rodriguez, H. G., Popp, J., Maringanti, C., and Chaubey, I. (2011). Selection and placement of best management practices used to reduce water quality degradation in Lincoln Lake watershed, *Water Resources Research*, Vol. 47, W01507, doi:10.1029/2009WR008549.
- Srinivas, N. and Deb, K. (1994). Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms. *Evolutionary Computation*, 2(3):221 - 248.
- USEPA (2009), SUSTAIN - A Framework for Placement of Best Management Practices in Urban Watersheds to Protect Water Quality, EPA/600/R-09/095.
- USEPA (2011), Report on Enhanced Framework (SUSTAIN) and Field Applications for Placement of BMPs in Urban Watersheds, EPA 600/R-11/144.