CRWR Online Report 06-09

Stage-Monitoring Network Optimization Using GIS

by

Sergio Ignacio Martinez Martinez, B.E.; M.S.E.

David R. Maidment, Ph.D.

August 2006

CENTER FOR RESEARCH IN WATER RESOURCES

Bureau of Engineering Research

The University of Texas at Austin

J.J. Pickle Research Campus

Austin, TX 78712-4497

This document is available online via World Wide Web at

http://www.crwr.utexas.edu/online.shtml

Acknowledgements

My first thank you is to Dr. David Maidment, my advisor, for his support, guidance and patience. I would also like to thank Dr. Daene McKinney, Dr. Randall Charbeneau, Dr. Ben Hodges, and Dr. Melba Crawford for serving on my dissertation committee, Dr. Venkatesh Merwade for all his kind advice and help, Dr. J. Wesley Barnes for his essential help on tabu search. Thanks go to Dr. Chandra Pathak of the South Florida Water Management District (SFWMD) for helping focus my PhD research, Dr. Alexandra Carvalho of Taylor Engineering Inc, for her guidance when writing reports for SFWMD, and Dr. Timothy Whiteaker and Rebecca Teasley, for their proofreading of this dissertation. Thanks also go to Sharon Bernard and the entire staff at the Center for Research in Water Resources (CRWR) for taking care of many things.

I would also like to thank Prof. Adriana Cafaggi-Felix of the National Autonomous University of Mexico (UNAM), my bachelor's advisor, Dr. Daniel F. Campos-Aranda, Prof. Arturo Dufour-Candelaria, and Prof. Armando Vazquez-Alfaro, for all their support for many years since I studied for my master's at the Autonomous University of San Luis Potosi (UASLP). Thanks also go to my friends and coworkers at CRWR, for all their support and friendship, to many friends I met at UT, from Mexico, United States and many other countries; they remind me that humans are like a big interconnected continent, if someone suffers the others suffer, if someone enjoys the others enjoy, and so on.

I am grateful to the Ministry of Education of Mexico (SEP), who via the Programa de Mejoramiento del Profesorado (PROMEP) granted me a PhD scholarship to study at UT-Austin. I am also grateful to the Autonomous University of Aguascalientes (UAA), my employer, who gave me permission to come here to advance my academic development. My especial thanks to my dean, Prof. Jorge Pio Monsivais-Santoyo, without his support, probably, I still would be dreaming about studying for a doctorate, to my fellow professors and former heads of department at UAA, Prof. Humberto Castañeda-Molina, Prof. Andrei Murillo-Mendez, and Prof. Alejandro Collazo-Tiscareño, for all their help and guidance. I cannot forget my students, they have been a continuous fountain of inspiration; I owe them a lot.

Finally, I need to express my immense gratitude to my parents, brothers and sisters, nephews and niece, and other members of my extended family who have always supported me in everything I do. Thank you all.

Abstract

Stage-Monitoring Network Optimization Using GIS

Sergio Ignacio Martinez Martinez, Ph.D.

The University of Texas at Austin, 2006

Supervisor: David R. Maidment

The South Florida Water Management District (SFWMD) has a large and expanding stage-monitoring network in need of optimization. The basic optimization problem is to keep in operation the least number of stage-monitoring stations (sites where the surface water elevation is constantly measured) in the best possible locations without having a serious loss of information. Two different methodologies, one for lakes, where water levels tend to be smooth and planar, and one for streams, where water levels can have significant changes produced by the varying hydraulic properties along the courses, are needed for optimizing the networks. Several optimization methods were tested including simulated annealing, genetic algorithm and tabu search. Tabu search was used as the definitive optimization method for both lakes and streams. Both methodologies

also need a spatial interpolation method because both need to estimate water elevations in specific points. The methodology for lakes uses inverse square distance weighting as the spatial interpolation method. The methodology for streams requires the use of HEC-RAS software developed by the US Army Corps of Engineers as the spatial interpolation method. The error of estimation of surface water elevations is performed via the root mean square error in both methodologies. Several cases located in lakes and streams in the Lake Kissimmee River Basin of the SFWMD are used to develop and test the methodologies. Additionally to the definitive methodology for stations in streams, two different but closely related earlier methodologies are developed, one for steady flow and the other for unsteady flow. Daily data are considered in the methodologies for stations in lakes and stations in streams with steady flow. In the case of stations with unsteady flow, it was determined that the data resolution should be at least one hour and flow and stage values should be instantaneous. Good results are obtained for the cases of stations in lakes and stations in streams with steady flow. For the case of stations in streams with unsteady flow, results are inconclusive. In addition to the optimization methodologies, a set of stage-network optimization guidelines are proposed.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Motivation

The management of water is a primary concern in the world. In populated coastal regions with a complex interaction between water, environment, and people, that concern is exacerbated. People should have water in good quantity and quality to support beneficial uses. In addition, humans and their land and buildings should not be subject to extreme floods or droughts. At the same time, local fauna and flora should not suffer deteriorated conditions because of the use of water by humans.

An ever-expanding need for water always adds new layers of complexity to the problem of water management. Somebody needs to take charge of all these issues. In the case of Florida, the water management districts are the organizations that deal with these problems. One important goal of a water management district is to know how much water is stored in each hydrologic/hydraulic component of the district, how much water is moving among these components, and how much water is interchanged between the district and its exterior. This knowledge may be required on a daily basis or even more frequently. The measurement of water elevations is a requisite to compute the storage of water in the components of the system. Flow may be obtained from the headwater elevation, tailwater elevation, and other parameters of a flow monitoring station at a weir or structure. Storage and flow depend on water surface elevation, that is, the computation of water balances based on the continuity equation, is a function of water surface elevation.

This computation of the water balance in the landscape is a very complex problem that is beyond the reach of this work; here, the emphasis is on the study of the measurement of water surface elevation. Water surface elevation, also called water elevation or stage, is measured in 1400 stations located in important points throughout the district, such as in lakes, wetlands, floodplains, rivers, creeks and canals. It is not necessary to establish a stage monitoring station in every corner of the district. It is better to employ the fewest number of stations positioned in strategic locations and at the same time to have a fair idea of the stage in the waterbodies of the district; that is, to optimize the stage monitoring network. Currently, there is no methodology to optimize stage-monitoring networks, yet the need for such a methodology in organizations like the South Florida Water Management District (SFWMD) certainly exists. The development of a methodology for optimizing the SFWMD stage-monitoring network is addressed in this work. The main research question that is addressed here is: given an admissible error, which existing stage monitoring stations should be left in place in the studied area of the district?

1.2 STUDY AREA

Because of the Everglades Restoration Project, the expansion of an already large water management system is pressing the SFWMD to optimize the current and future flow and stage-monitoring network. This district, one of the five water management districts of the state, is located on the southern part of the territory of Florida (Figure 1.1).

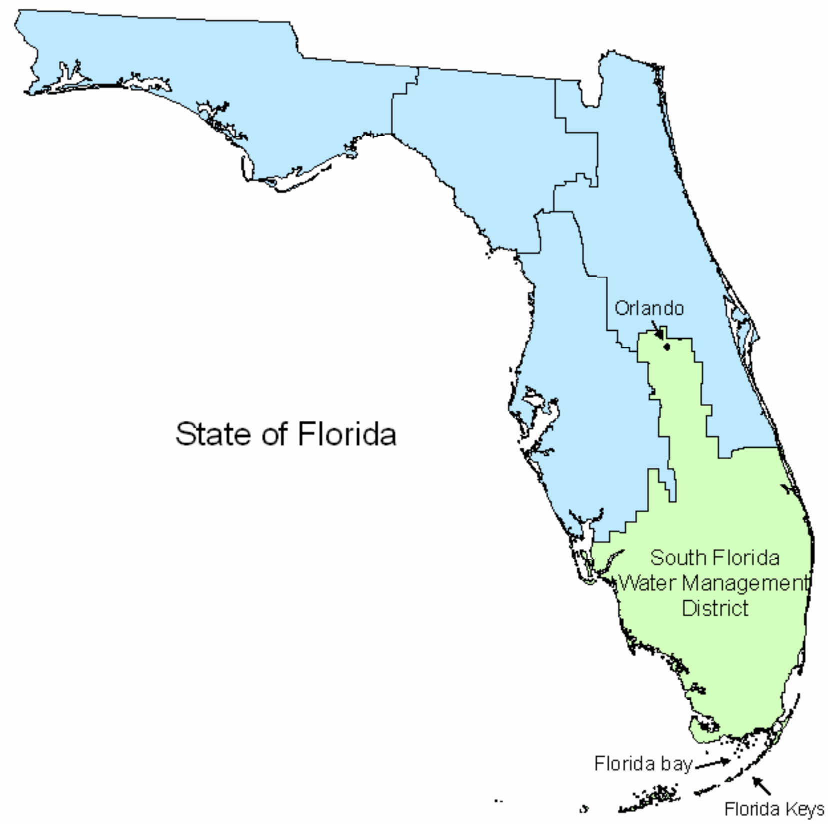


Figure 1.1: South Florida Water Management District (SFWMD).

The SFWMD contains one of the most diverse ecosystems in the world. From Orlando to the Florida Keys, it has a length of 240 miles. Naturally, most of Southern Florida was a marsh fed by rainfall. To make this area habitable, it was dredged and drained from 1850 to 1950. The central and southern Florida Flood Control Project, located between Orlando and Florida bay, consists of 1,800 miles of canals and levees and 200 water control structures. Sixteen pump stations send water south to Florida Bay and through waterways eastward and westward to both coasts. Canals distribute water cross the peninsula. An adverse effect produced by the development of land in the second half of the 20th century was the decline of wildlife and the environment. Since the 1970s, the SFWMD has been responsible for restoring the environment. Now, major restoration efforts are ongoing in the Everglades, Kissimmee River, Lake Okeechobee, and coastal estuaries (www.sfwmd.gov). The SFWMD provides flood protection and water supply to almost six million inhabitants. The district operates a surface-water distribution system, consisting of approximately 1800 miles of canals, over 450 managed structures, and over 1000 flow and stage gages (SFWMD 2004).

The Center for Research in Water Resources of the University of Texas at Austin participated with Taylor Engineering and the SFWMD during 2004 and 2005 in a pilot study (SFWMD 2005) that had as its objective the creation of a methodology to optimize the entire SFWMD stage-monitoring network. In order to develop that methodology, the Kissimmee River Basin, a smaller section of the SFWMD with 44 subbasins, was selected because it has lakes and streams that can be considered representative of those of the district and contains a network of 160 stage-monitoring stations and 58 flow structures. The Kissimmee River Basin includes three of Florida's major lakes, Lake Okeechobee, Lake Istokpoga and Lake Kissimmee, and the Kissimmee River, which connects Lake Kissimmee to Lake Okeechobee. The Kissimmee River Basin is located roughly between latitude 27° 00‘ N and 28° 30’ N and longitude 80° 37’ 30” W and 81° 37’ 30” W.

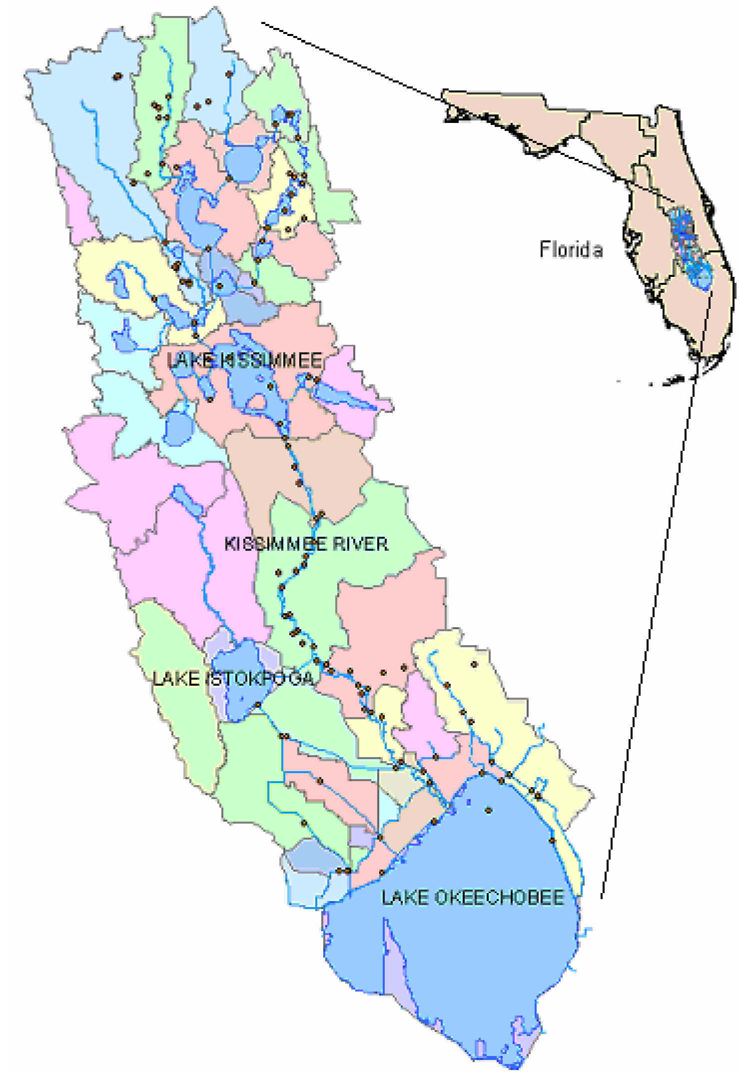


Figure 1.2: Kissimmee River Basin Study Area includes 44 subbasins, three major lakes, 160 stage-monitoring stations and 58 flow structures.

1.3 OBJECTIVES

A water management district can be composed of different hydrocomponents (hydrologic/hydraulic components) such as lakes, wetlands, rivers, and canals. They could be natural or artificial, and in general, all these hydrocomponents can be classified as lakes or streams; lakes if they have large cross sections, low water velocities and low spatial changes in water elevation; streams if they have small cross sections, significant water velocities and high changes in water elevation along their length. Examples of lakes are Lake Kissimmee or Lake Istokpoga. An example of a stream is the Kissimmee River (Figure 1.2).

In lakes and streams, it is necessary to monitor stages and to do it in an optimal manner; therefore, this research has the main objectives:

1. Given an existing stage-monitoring network of a lake or stream, identify which gages of the network are most important, or alternatively, which gages can be closed without serious loss of information.
2. Given the previous objective, study several methodologies to fulfill it, compare their results, and adopt the best methodology.
3. Given an unmonitored lake or stream, develop a series of rules to define where new gages should be located.

The three objectives are accomplished by using the datasets generated from data of the SFWMD. The methodologies developed to accomplish the first two objectives are tested by applying them to several case studies of the Kissimmee River Basin. The series of rules that fulfill the third objective is based on the insights gained at developing the methodologies for lakes and streams. Objective 1 is directly related with the main problem of this research. In very generic terms, the problem can be defined as: given a set of n stage-monitoring stations, identify a subset of r stations ($r < n$) that minimize a given objective function of measurement error defined from the water surface elevation estimation in the $n - r$ stations left outside the selected subset of r stations. If, for example, $n = 10$ and $r = 5$, the number of combinations of stations is ${}^{10}C_5 = 252$ but, if $n = 17$ and r is equal to 8, the number is ${}^{17}C_8 = 24310$ making the exhaustive evaluation of each subset an impractical solution, even for 252 combinations. An optimization method is needed to solve this problem.

1.4 ASSUMPTIONS AND THE SCOPE OF WORK

Based on the stated objectives, this research is trying to answer three questions:

1. Given the stage-monitoring network of a lake or stream, which are the most important gages of the network? Alternatively, which of the gages can be closed without serious loss of information?
2. Which is the best methodology to answer the previous question?
3. Given an unmonitored lake or stream, where should be the gages located?

The first two questions require the generation of procedures with results in quantitative form, while the third question requires the generation of a set of rules or guidelines that do not have strict quantitative results until the procedures that answer questions 1 and 2 are applied.

To answer these three questions the following assumptions are made:

- Retaining a subset of the measuring stations of a lake or stream, and using a spatial interpolation model, are adequate to estimate the water elevation in the remaining station locations.

- Historical data, measured in the same period in all the stations of a lake or stream can be used to optimize its measuring network.
- The application of a spatial interpolation method and of a gage optimization method provides the necessary and sufficient tools to optimize the stage-measuring network of a given lake or stream.
- In the case of a lake, a geointerpolation method can be used to estimate water surface elevations.
- In streams, the one-dimensional computation of the water profile constitutes an adequate interpolation method.

The work presented in this dissertation is applicable only to stage-monitoring networks that resemble those networks studied. This includes networks in regions with low topographic relief, which have interconnected waterbodies, and where the water supply is sufficient to have storage in the waterbodies all year round. The applicability of the procedures in other types of networks is beyond the scope of this dissertation.

The computational support to carry out this research is given by ArcGIS 9.1. ArcGIS is a GIS package published by the Environmental System Research Institute (ESRI) (2004). This package stores spatial and temporal data in geodatabases and offers customization capabilities using several programming languages and an extensive set of customizable components called ArcObjects (Burke 2003). In this work, Visual Basic for Applications and ArcObjects are used to develop computer routines to implement the developed procedures. In addition, for the case of the optimization of the stage-monitoring network in streams, HEC-RAS, the River Analysis System Software developed by the US Army Corps of Engineers (2004) is used as an interpolation routine. Figure 1.3 shows a flow profile, produced by HEC-RAS, from which it is possible to obtain water elevations at specific locations.

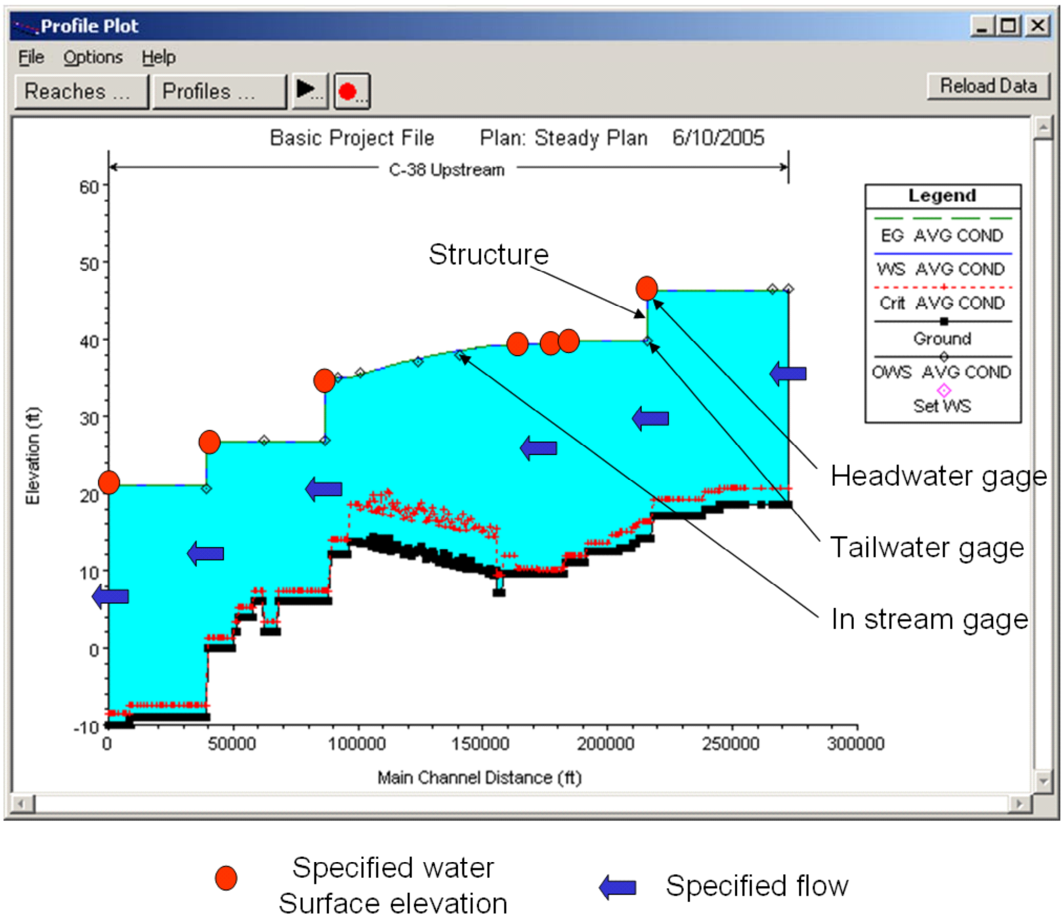


Figure 1.3: An example of a HEC-RAS computed flow profile (annotations are added).

1.5 Contributions from this Research and their Significance

The practical contribution of this research is a set of methodologies to optimize the current network of stage-monitoring stations of the SFWMD and the development of a set of rules that could be used in the design of new measuring networks or in the revision of current networks. The methodologies and the set of rules can be applied only to stage-monitoring networks that share similar characteristics to those of the SFWMD network. These characteristics are low topographic relief and water in waterbodies all year round. The gradient of the land surface along the channel of the Kissimmee River is of the order of 0.0001 and there is always water in the waterbodies of the study area.

An important theoretical contribution of this research is the identification of the need of two different approaches to optimize the measuring network: one for lakes, and one for streams. Water moves slowly in a lake, compared to how fast it moves in a stream. In a lake, water elevation has little change from one point to the other; while in a stream, water elevation can change dramatically from one section to the next. In addition, the optimization of the stage-monitoring network along streams can be achieved using steady flow computations. Unsteady flow computations can be used to optimize the station network along streams, but the process is more difficult and the results are less robust than those obtained using steady flow.

The results of this research provide the SFWMD with new tools for optimizing its stage-measuring network. In fact, these methods could be used in any district that has a similar network waiting to be optimized.

1.6 DISSERTATION OUTLINE

This dissertation is organized into nine chapters. This chapter, the first, introduces the problem. A literature and technology review is presented in the second chapter concerning global optimization methods, geointerpolation, and computational one-dimensional models. Chapter 3 describes the methodology followed when developing this work. The implementations of the customized routines that offer support to the network optimization methodologies are described in Chapter 4. Complete results of network optimizations in lakes, streams with steady flow, streams with unsteady flow, and an additional study about the application of tabu search to solve the problem are presented, respectively, in Chapters 5, 6, 7 and 8. Finally, Chapter 9 presents the conclusions and recommendations.

Chapter 2: Literature and Technology Review

2.1 INTRODUCTION

The literature presented in this chapter describes optimization methods and the application of these methods in water resources problems, including the optimization of monitoring station networks. Other work reviewed concerns spatial interpolation and open channel hydraulics. These subjects are important because they provide the background to estimate water elevations in lakes and in streams, respectively. Additionally, three open channel hydrodynamic models are reviewed because they could function as interpolation routines in the case of optimization of measuring station networks along streams. Finally, several papers that present the combined use of open channel hydraulics software and optimization methods are described.

2.2 OPTIMIZATION

Water resources problems are solved using a design and analysis process. The analysis phase determines the behavior of a current or proposed system; while the design phase determines the size and required properties of the system components. A graphical explanation of this process is shown in Figure 2.1.

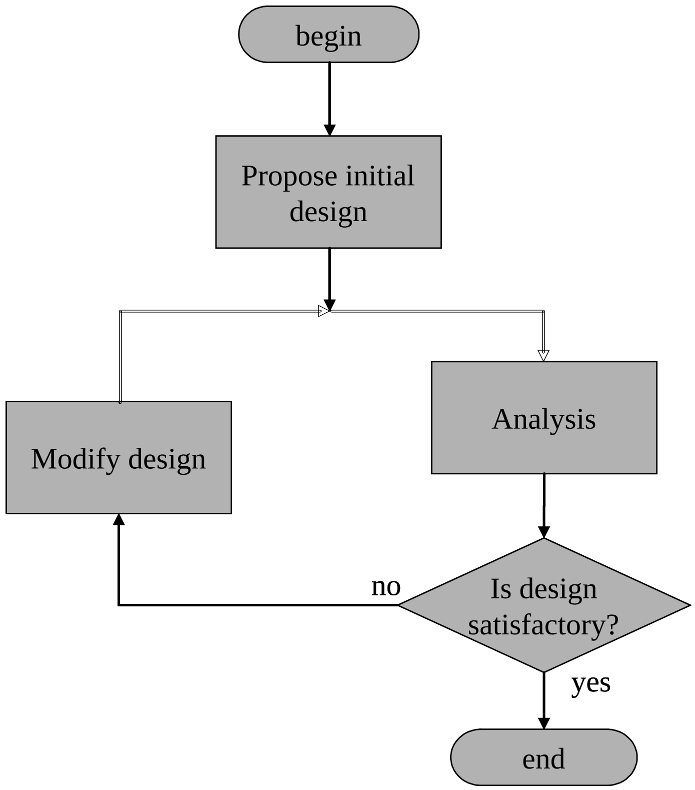


Figure 2.1: Design-analysis process.

Frequently, analysis and design are applied in a trial and error process heavily dependent on human decisions, which could lead to inefficient systems. Optimization procedures eliminate that trial and error process by searching automatically for the optimum system design (Mays and Tung 2005).

Typically, an optimization method or procedure includes a mathematical representation of the system and its relationships with its surroundings and a set of rules to find the best design. The representation includes an objective function that evaluates system performance and a set of constraints that limit the values of the decision variables. In water resource problems, constraints can be of two types. The first type of constraints embody inviolable physical limitations such as conservation of mass or limits on available resources; the second type is an implicit goal that could be violated, as budgetary limitations or schedules of water deliveries (Loucks et al. 1981). In general, an optimization problem can be formulated as: optimize the objective function $f(x)$ of a vector x of n decision variables; subject to a set of m constraints $g(x)=0$ and to the inequalities $\underline{x} < x < \overline{x}$ of bound constraints of decision variables, where $\underline{x}$ represents the lower bound and $\overline{x}$ represents the upper bound (Mays and Tung 2005). If a set of values of the decision variables simultaneously satisfies all constraints, it is said to be a feasible solution. The set of all feasible solutions constitute the feasible region. An optimal solution is a feasible solution that provides an optimal value of the objective function. The method of optimization used in a particular problem depends on the type of objective function, the type of constraints, and the number of decision variables. A problem of optimization can be of maximization or minimization; both are closely related because maximizing a function is equivalent to minimizing the negative of that function, and vice versa.

Depending on the nature of the objective function and the constraints, an optimization problem can be classified, for example (Mays and Tung 2005), as:

- Linear or nonlinear. It is linear if the objective function is linear and all constraints are linear. A nonlinear problem has a nonlinear objective function and/or one/several nonlinear constraints (See Figures 2.2.a and 2.2.b).

- Deterministic or probabilistic. It is deterministic if all variables take fixed values, but if variables take random values then it is probabilistic.
- Static or dynamic. A problem is static if it does not take into account time, but if it does, it is dynamic.
- Continuous or discrete: A problem where variables take continuous values is continuous and if its values only take discrete values, it is discrete.

The optimal point of a function can be either global or local. If the optimal point is the extreme value, that is, a maximum or a minimum, of the region of interest, that point is global; if it is the extreme value of a small neighborhood, it is local (see Figure 2.2.c)

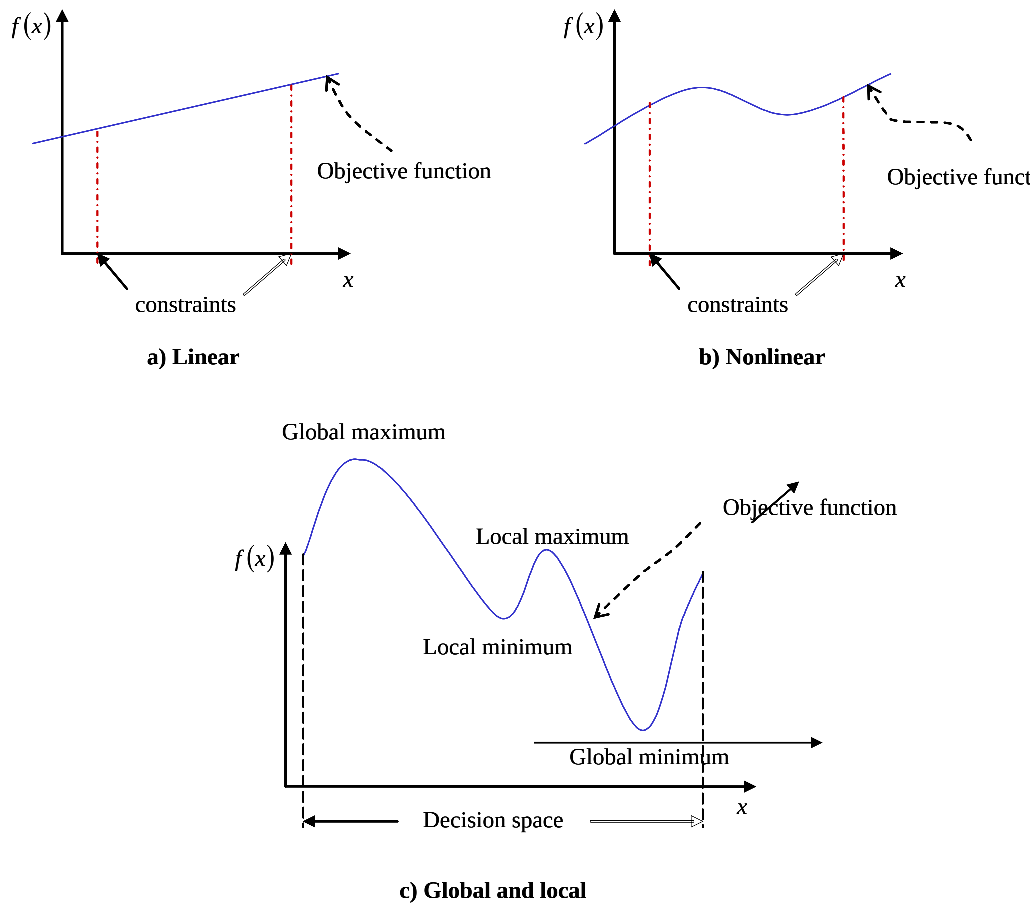


Figure 2.2: Simplified linear-nonlinear optimization problems. Distinction between global and local optima.

Often there is more interest in finding the global optimum than a local optimum, but this can be very difficult; frequently, problems have many local optima. The most suitable methods to locate optimal values depend upon the nature of the function been studied. There are two broad classes of methods: local and global. Local methods locate the optimal value within a neighborhood; global methods find the optimal point of the complete region. A local method is local because it only uses information about the objective function from the neighborhood of the current approximation when it is updating the approximation, and because it is expected that such a method converges to whatever local optimum is closest to the starting approximation. As a result, the global structure of an objective function is unknown to a local method. Some of these techniques, such as downhill simplex (Nelder and Mead 1965) and Powell's method (Powell 1964) do not require the derivatives of the objective function. Others, such as steepest descent, conjugate gradient methods, quasi-Newton methods and Newton methods require at least the gradient (Mays and Tung 2005). The choice of which local optimization technique to use depends upon whether there are one or more decision variables, whether the objective function can be differentiated, and how smooth the objective function is. If a good estimate of the position of the optimum value exists, a local method can be useful to improve it and find the optimum choice of values of the decision variables. If no such estimate exists, a global method must be used. The solution of a problem often requires the application of both local and global methods. The next sections are devoted to global optimization methods and sample applications in water resources problems.

2.2.1 Global Optimization Methods

Global optimization methods are useful when the search space is likely to have many extreme values, making it hard to locate the true global optimum. In the optimization process of a water resources system that depends on many decision variables, it is difficult to know the time needed to obtain the actual optimal solution. A decision between two paths has to be made, to obtain the optimal solution with a large amount of time consumed or to find a rapid, perhaps suboptimal, solution using an approximate or heuristic algorithm. Simulated annealing algorithms and genetic algorithms are two types of high quality general heuristic algorithms that are independent of the nature of the problem (Mays and Tung 2005). Both are randomization algorithms that explore the neighborhood of the current solution to generate a step by step improvement of the objective function. These algorithms use several techniques to avoid getting trapped in local optima. First, they can try a large number of initial solutions. Second, they introduce complex neighborhood structures to search large regions of the solution space. And, third, they accept a limited exploration of the solution space where the solution is inferior. When correctly applied, simulated annealing and genetic algorithm approaches can explore the search space better than a systematic search for a given number of function evaluations, and are more likely to find the true global optimum. Note that both of these approaches involve the use of probability in the generation of possible solutions and in the adoption of suboptimal solutions, and so may fail to find the true optimum. Since they are better at global searching than local optimization, it is useful to refine any final solution using a local search method such as steepest descent or a conjugate gradient method. Currently, simulated annealing and genetic algorithms are gaining more and more applications for water resource systems optimization (Mays and Tung 2005). Tabu Search algorithms are another type of high quality heuristic algorithms that can also be applied to water resource problems, especially when combinatorial optimization problems are present. Combinatorial optimization problems are characterized by a finite number of feasible solutions (Gen and Cheng 2000). The modern formulation of tabu search was originated by Glover in 1986 (Glover and Laguna 1993). The main characteristics of tabu search is the prohibition of visiting previously considered solutions, at least for several iterations, and to guide the search of solutions along the best direction of the current neighborhood. This method has been applied to a wide range of engineering problems and is still been refined.

2.2.2 Simulated Annealing

The simulated annealing method formalized by Kirkpatrick et al (1983) is based on an analogy with the physical process of annealing. Annealing is the process that a material follows when it is subjected to two extended periods of change of temperature, one of heating and then one of cooling. At high temperatures, the molecules of a molten metal move about freely. If the metal is cooled slowly, its molecules gradually lose mobility and form a pure crystal that is completely ordered. This state constitutes the global minimum energy state for the system. On the contrary, if the material is cooled too rapidly it falls into an unordered state with multiple crystals or even with no discernible crystalline structure. This amorphous state is a local minimum energy state with higher energy than the global minimum. The important thing is the appropriate rate of cooling which allows molecules to settle in an orderly fashion. Simulated annealing allows occasional uphill jumps, permitting the escape from local minima and giving a better chance of finding the global minimum. The probability of accepting an inferior solution x_{new} is computed according to

$$\Pr(\text{accept } x_{new}) = \exp\left[-\frac{f(x_{new}) - f(x_{old})}{T}\right] \tag{2.1}$$

where x_{old} is the previous optimal solution; x_{new} is the new trial solution; $f(x)$ is the objective function value analogous to the energy level of the material; and T is the control parameter analogous to the temperature of the material. Two other requirements are needed for the algorithm: a method of generating new points x_{new} and an annealing schedule dictating how to reduce T as the execution of the algorithm progresses. The effect of the reduction of T as execution progresses is to make less and less probable the acceptance of inferior solutions, i.e. "crystallizing the material". Unfortunately both of these requirements are problem specific; a new set of decision variable values can be obtained probabilistically for each trial, while the annealing schedule depends on the expected range of function values and the shape of the function surface. Most of the time, some experimentation is required to develop an algorithm useful for a particular problem.

2.2.3 Applications of Simulated Annealing in Monitoring Networks

Simulated annealing has been applied in the solution of problems closely related to the problem of this research. Among them are the following:

Lee and Ellis (1996) applied eight algorithms, among them simulated annealing, tabu search and genetic algorithms, to the problem of groundwater pollutant concentration monitoring network design and then compared the algorithms' results. The objective function that the authors minimized was the variance of the interpolation estimation. Considering the frequency of finding the true solution, the rank of its solutions compared with the best and the worst solutions of all algorithms, the time of computation, the ease of implementation and the number of required parameters, the authors found that simulated annealing was the best algorithm for their problem. The second best, a close second, was tabu search. A possible limitation of this study is that the authors used two simplistic synthetic networks that consider monitoring points in the vertices of square cells in a rectangular and quasi-rectangular regions, so further research with a more realistic data set could be needed.

Pardo-Igúzquiza (1998) presented a method for establishing an optimal network design for the estimation of areal averages of rainfall events. This author used a geostatistical variance-reduction method in conjunction with simulated annealing. Several synthetic examples were used. Two problems were attacked: the optimal identification of a subset of stations of an existing station network and the expansion of a current network. The author concluded that simulated annealing is a good method to optimize these types of networks.

A method for optimizing the selection of river sampling sites was presented by Dixon et al. (1999). They presented procedures using a geographic information system (GIS), graph theory, and a simulated annealing algorithm. In their paper, three case studies are presented showing how to use the methodology in three different situations. These authors concluded that the method produces a sampling program consistent with monitoring program goals.

Three optimization models are proposed in Nunes et al. (2004a) to select the best subset of stations from a groundwater-monitoring network. The monitored variable could be water level, pollutant concentration or temperature. One model finds the stations that minimize the variance of the estimation error, which results in the better spatial distribution of stations, producing the maximum spatial accuracy. Another model minimizes temporal redundancy. Temporal redundancy is a measure of the similarity between time series. This model retains those stations that show larger differences between them over time. The third one simultaneously maximizes spatial accuracy and minimizes temporal redundancy. The models are solved using simulated annealing and an annealing schedule based on statistical entropy. A 32-station synthetic case study is used to compare results of the proposed models when a 17 station-subset is to be chosen. Comparing the results of the three models, the authors concluded that the consideration of time and space data in the third model contributes to the selection of the most relevant subset of stations; neither with stations as evenly distributed in space as in the first model, nor as grouped in high temporal variability zones as in the second model.

A method for designing groundwater-monitoring networks to define the extent of contamination with nitrates from agricultural origin is proposed by Nunes et al (2004b). This method is good for reducing existing networks with stations with incomplete time series. Simulated annealing is used to minimize the variance of the estimation error obtained by kriging defined with spatial data. Optimization is performed for several measurement times, obtaining an equal number of optimized small-dimension networks; stations that are repeated more often in these networks are selected to form part of the final network. A nitrate-monitoring network in Portugal is used as a case study. The original 89-station network was reduced to 16 stations by this approach.

2.2.4 Genetic Algorithms

Genetic algorithms are search algorithms inspired by the way natural populations evolve. John Holland and collaborators at the University of Michigan were the first developers of genetic algorithms. Holland and his team worked in the field of artificial intelligence and their main preoccupation was to develop robust methods, that is, methods that in many different optimization problems should be able to find the global optimum (Goldberg, 1989). In genetic algorithms, as in nature, the best-fitted individuals have better chances of passing their genes to the next generation. Population evolution is driven by crossover and mutation. Crossover occurs when a pair of parents give their children a mixed share of their genes. However, there is no perfect reproduction of genes; sometimes nature makes mistakes, called mutations. Mutations may lead to better-fitted or unfitted individuals. When a new generation matures, the process is repeated again.

The key features of genetic algorithms are:

- A point (or set of decision variable values) in the search space is encoded as a chromosome.
- A population of many chromosomes is developed and maintained.
- New sets of points are generated by combining existing solutions.

- Optimal solutions are evolved by iteratively producing new generations of chromosomes in which the better solutions are combined through crossover while the worst ones are discarded.
- In addition, a small amount of *mutation* is introduced, randomly changing bits in the chromosomes.

Usually a chromosome is a string of bits formed by concatenation of the bit strings representing each of the *n* decision variables. The number of bits used to encode each parameter depends on the desired tolerance. Eventually, applying a genetic algorithm, all the current generation of individuals should end up close to the global optimum. A typical graphic representation of the general process of genetic algorithms is shown in Figure 2.3.

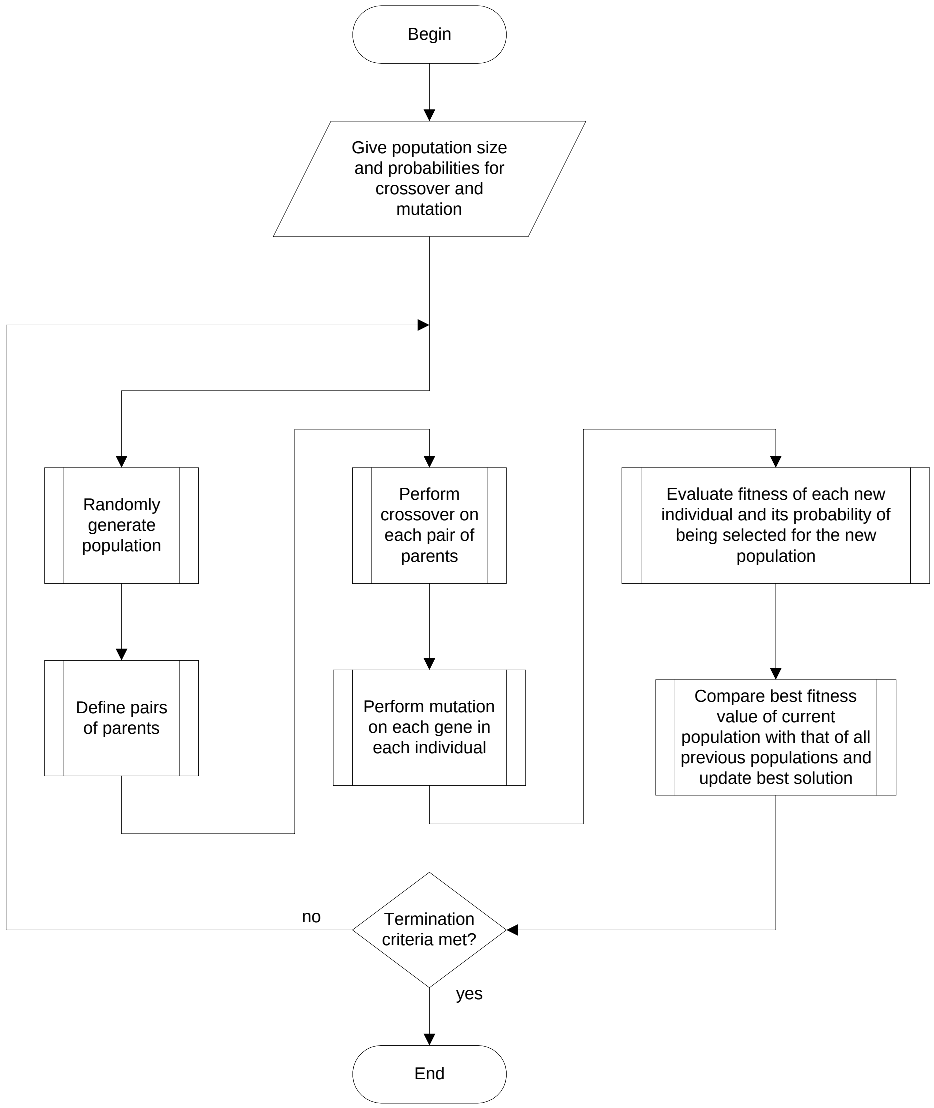


Figure 2.3: Genetic algorithm flow chart.

2.2.5 Applications of Genetic Algorithms in Monitoring Networks

Many water resource problems have been solved applying genetic algorithms. Among those problems, closely related to the problem of this research, are the following:

Cieniawski et al. (1995) presents an optimization problem solved with genetic algorithms that find the optimal location of monitoring wells to maximize the probability of leak detection and minimize contaminated area, given potential contaminant sources and an unknown hydraulic conductivity field. The authors use four implementations of genetic algorithms and one of simulated annealing. They studied the best location of five wells using a dataset of 200 simulated plume realizations and 135 potential well locations. The results of simulated annealing are duplicated using a weighted objective genetic algorithm. In their case, the multi-objective genetic algorithms show promising results, but the authors recommend further study.

A methodology for sampling plan design to reduce the costs of long-term groundwater contamination monitoring sites is presented in Reed et al. (2000). The method combines a fate-and-transport model, plume interpolation, and a genetic algorithm. The interpolation methods considered are inverse-distance weighting, ordinary kriging and a combined method of both approaches. The application of the methodology using the hybrid interpolation approach to a case study indicated that the sampling costs could be reduced by as much as 60% with minimal loss in accuracy of the global mass estimates.

Reed and Minsker (2004) present a problem consisting of the optimization of a long-term contaminant-in-groundwater monitoring system that has more than two objectives. The solution combines kriging and a genetic algorithm to balance four objectives, minimizing sampling costs, maximizing the accuracy of interpolated plume maps, maximizing the relative accuracy of contaminant mass estimates, and minimizing estimation uncertainty. Optimizing the problem with respect to these objectives reduced the decision space from 500 million designs to a set of 1,156 designs. Then, visualization of eight designs aided stakeholders in the understanding and balancing of objectives to obtain a single compromise solution.

2.2.6 Tabu Search

Tabu search derives and exploits a collection of principles of intelligent problem solving (Glover and Laguna 1993). A fundamental element of tabu search is the use of flexible memory. Flexible memory implies existence of the processes of creating and exploiting structures that take advantage of history. In summary, by using memory structures, tabu search imposes restrictions to guide the search process. A general mathematical optimization problem can be used to explain the method. That is, minimize $f(x)$ subject to $x \in X$. The objective function $f(x)$ may be linear or nonlinear, and the condition $x \in X$ denotes constraints on the vector x. These constraints may include linear or nonlinear inequalities, and some of the components of x may have discrete values. In combinatorial optimization problems, this general formulation can be a codified formulation of a problem of interest. For example, the constraints may specify logical conditions or variable relationships that would be difficult to write down mathematically and better left as verbal provisions, such as sets of rules. Frequently, in these cases a more complex structure must be simplified using variables as codes (Glover and Laguna 1993). For example, in a gage monitoring network problem an element of x may be a binary variable that receives a value of

1 to code for assigning a gage to a subset of the original set, and receives a value 0 to indicate the assignment does not occur.

Tabu search may be characterized as a form of neighborhood search (Glover et al 1993). In neighborhood search, each solution $x \in X$ has an associate set of neighbors, $N(x) \subset X$, called the neighborhood of x. Each solution $x' \in N(x)$ can be reached directly from x by an operation called a *move*, and x is said to move to x' when such an operation is performed. Normally in tabu search, neighborhoods are assumed symmetric, i.e. x' is a neighbor of x if and only if x is a neighbor of x'.

A basic algorithm of the tabu search method consists of the following three steps (Glover and Laguna 1993):

- Step 1 (Initialization)

  Select a starting solution $xnow \in X$. Record the current best known solution by setting $xbest = xnow$ and define $best\_func = f(xbest)$. Initialize the history record H; H implements the memory structures used in the given problem to classify a subset of the moves in a neighborhood as forbidden (or tabu). These structures are based on the knowledge of the problem.

- Step 2 (Choice and termination)

  Determine $Candidate\_N(xnow)$ as a subset of $N(H, xnow)$. $N(H, xnow)$ is also a subset of neighbors of $xnow$, it takes into account the restrictions created by H. Select $xnext$ from $Candidate\_N(xnow)$ to minimize $f(H, x)$ over this set. Terminate by a chosen iteration cut-off rule.

- Step 3 (Update)

  Re-set *xnow* = *xnext*. If *f*(*xnow*) < *best_func*, then *xbest* = *xnow* and *best_func* = *f*(*xbest*). Update the history record *H*. Return to Step 2.

Figure 2.4 shows a flowchart that represents the previous algorithm. Grey colored shapes correspond to Step 1 (Initialization), blue shapes to Step 2 (Choice and termination), and green shapes to Step 3 (Update).

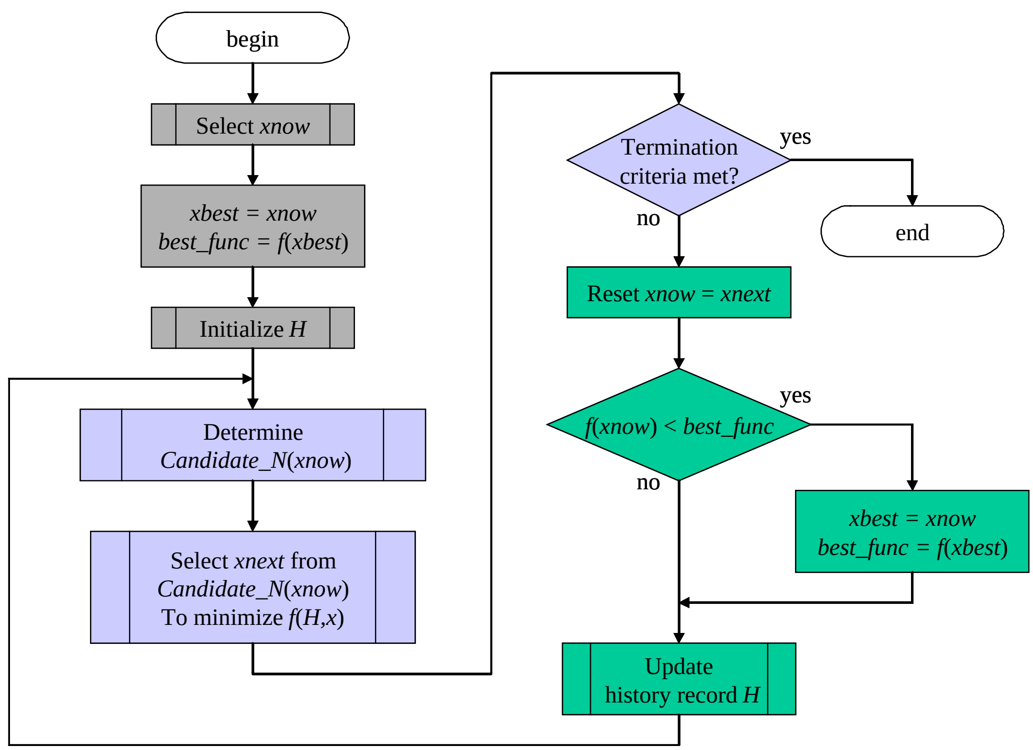


Figure 2.4: Flowchart of a basic algorithm of tabu search.

The behavior of the method depends on how the history *H* is defined and utilized, and on how the candidate neighborhood *Candidate_N*(*xnow*) and the evaluation function $f(H, x)$ are determined. In the simplest cases, *Candidate_N*(*xnow*) can be equal to $N(xnow)$ and $f(H, x) = f(x)$. One such a simple case is illustrated in Figure 2.5, where the current solution x has five neighbors $N_1(x)$, $N_2(x)$, $N_3(x)$, $N_4(x)$, and $N_5(x)$. The move to the fifth neighbor is a tabu move and the move to the third neighbor produces the best next solution.

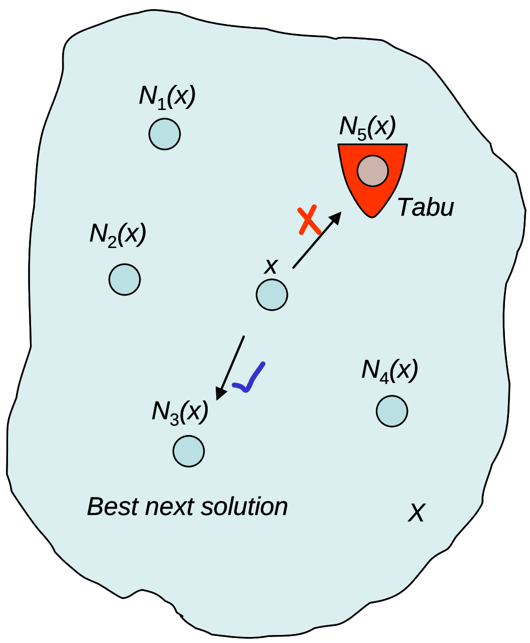


Figure 2.5: Tabu search example of finding the best next solution.

A *move* in the context of the stage-monitoring network optimization problem can be illustrated with the case of a subset with $r = 3$ stations taken from a set of $n = 7$ stations. The subset of stations is $x = \{2, 4, 7\}$. How many subsets of $r = 3$ are neighbors of subset x? What is the move associated with each neighbor? These questions can be answered considering Figure 2.6.

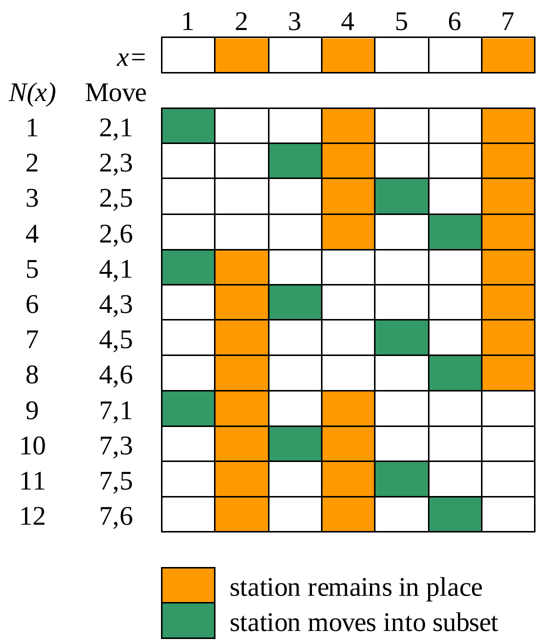


Figure 2.6: Subset of three stations. Neighbors of three stations and associated moves.

If the second station is removed from subset x there are four stations that can go into the subset, stations 1, 3, 5, and 6. The swap of station 2 for any of the four stations creates the moves (2, 1), (2, 3), (2, 5), and (2, 6). These moves produce the neighbors $N_1(x) = \{1, 4, 7\}$, $N_2(x) = \{3, 4, 7\}$, $N_3(x) = \{4, 5, 7\}$, and $N_4(x) = \{3, 6, 7\}$. These neighbors and moves are close to the top of Figure 2.6. The orange colored cells represent stations that remain in subset x, green cells represent stations that move into subset x. The remaining eight neighbors and eight moves are below the first four. The total number of swap moves/neighbors is $r\,(n - r)$. In this example is $3 \times (7 - 3) = 12$. There are also insertion moves and deletion moves. These moves generate their corresponding neighbors. The total number of insertion moves/neighbors of subset x is $n - r$, while the total number of deletion moves/neighbors is r. The total number of moves/neighbors of subset x is $(n - r)(r + 1) + r$. In this example is $(7 - 3) \times (3 + 1) + 3 = 19$.

The memory structures of tabu search operate by reference to four principal dimensions, consisting of recency, frequency, quality, and influence. These dimensions are set against a background of logical structure and connectivity (Glover and Laguna 1993). *Recency* and *frequency* refer to how recently or frequently certain moves or solution components, called *attributes*, have participated in generating past solutions. *Quality* refers to how good a move or solution is. Associated with a move there is a move value, which represents the change on the objective function value resulting of the proposed exchange. Move values provide a basis for evaluating the quality of a move, although other criteria can also be important. *Influence* measures the degree of change induced by a move in the solution structure or feasibility. A move of high influence induces a significant change in the structure of the current solution. High influence moves are important, especially during intervals of breaking away from local optimality (Glover and Laguna 1993). Moves of lower influence are tolerated until the gain from them appears to be negligible. At such a point, and in the absence of improving moves, *aspiration criteria* should shift to give influential moves a higher rank. An *aspiration criterion* is a condition that allows a tabu move to override its tabu classification in the case that the tabu move would result in a solution better than any visited so far. The literature about tabu search is extensive, a sample is constituted by Glover (1989), Laguna and Barnes (1991), Glover and Laguna (1993), Battiti and Tecchiolli (1994), Barnes and Chambers (1995), Carlton and Barnes (1996), and Barnes et al (2003).

2.2.7 Applications of Tabu Search in Related Water Resource Problems

Tabu search has been widely used, however, the application of tabu search to water resource problems has been limited. Three examples are presented in this section.

Zheng and Wang (1996) apply tabu search and simulated annealing to the problem of the identification of the optimal parameter structure of one-dimensional groundwater flow models. Parameter structure refers to how the values of parameters spatially vary in a distributed-parameter system. The parameter structure is taken into account in the parameter identification process through zonation or interpolation. In the zonation approach, the model grid is subdivided into zones, and the aquifer properties are treated as constant within each zone, subsequent adjustment is made using calibration runs. In the interpolation approach, the parameter structure is defined by a set of the model points where unique parameter values are initially specified and are subsequently adjusted. The problem of the paper is treated as a combinatorial optimization problem considering the existence of several zones with constant aquifer properties and with adjustable boundaries defined by a grid. The performance of tabu search and simulated annealing is compared with simple grid search and descent search techniques, using preliminary results from one-dimensional examples. Only simulated annealing and tabu search reached the optimality in the two test problems considered. In the first test problem, the number of function evaluations made using simulated annealing was 905 while tabu search made 276; in the second test problem, simulated annealing made 74334 and tabu search made 2646. That is, tabu search performs more efficiently than simulated annealing.

As is mentioned in section 2.2.3, Lee and Ellis (1996) applied eight algorithms, including simulated annealing, tabu search and genetic algorithms, to the problem of groundwater pollutant concentration monitoring network design. According to the criteria previously mentioned, they found that simulated annealing was the best algorithm. The second best method was tabu search. In order of preference, considering the ease of implementation, coding complexity, number of required parameters and computational effort, they recommend to apply either simulated annealing or tabu search.

Cunha and Ribeiro (2004) present a tabu search algorithm to find the least-cost design of looped water distribution networks. Discrete diameters of pipes are chosen in this combinatorial optimization problem. The results obtained for five published classical water-distribution network case studies favorably compare to those obtained using other methods. Additionally, the computation times spent in the optimization process, using a Pentium PC at 433 MHz, were from 16 to 445 seconds. This study demonstrates the usefulness of tabu search in solving this optimization problem.

2.2.8 Combinatorial Optimization Problems

The problem addressed in this research is a combinatorial optimization problem because it has a finite number of feasible solutions built from unordered subsets of r stations taken from n stations. Combinatorial optimization problems abound in everyday life and in engineering design (Gen and Cheng 2000). Although the optimal solution to such finite problems can be found by enumeration, in usual practice it is impossible, because in practical problems the number of feasible solutions can be extremely high. A major trend in solving such hard problems is to utilize a heuristic search. Combinatorial optimizations problems have a high number of different features and properties, but in essence, they can be classified as one of four different types (Gen and Cheng 2000), here enumerated with examples of stage-monitoring stations, in which one has to:

- determine a permutation of several items associated with the problem;

  An ordered list of n stage-monitoring stations taken one at a time is an example. In that list, one station of the n stations is the first member of the list. The same first station and other of the remaining $n - 1$ stations are the second member of the list. The same first and second stations and another of the remaining $n - 2$ stations are the third member of the list, and so on.
- determine a combination of several items;

  The combination of r stations taken from n stage-monitoring stations in a lake provides one example.
- determine both permutation and combination of several items;

  One example, is a partial ordered list of r stations taken one at a time, complemented with $n - r$ subsets of stations taken from n stage-monitoring stations

- determine a permutation and/or a combination of several items but subject to a set of constraints;

  This type of problem can be exemplified with the combination of *r* stage-monitoring stations taken from *n* stations in a stream subject to a set of hydraulic constraints.

This research solves two problems. One in which one has to determine a combination of several stage-monitoring stations in a lake (second type), and other in which one has to determine a combination of several stage-monitoring stations in a stream subject to a set of hydraulic constraints (fourth type).

2.3 SPATIAL INTERPOLATION

One uses interpolation functions if one has to create a continuous estimation surface from measured point values. Interpolation functions assume that values measured at the points are spatially correlated; closer points tend to have similar values. According to the use of statistics, there are two types of interpolation methods, functional and geostatistical. The first type are directly based on nearby measured values and on mathematical functions to produce a smooth surface, two examples of functional methods are inverse distance weighting (IDW) and spline; geostatistical interpolation methods are based on statistical models that use autocorrelation, the statistical relationship among measured points. One example of a geostatistical method is kriging. In either type of method, values measured at known points are used to estimate values at unknown points. Another classification of point interpolation methods is based on whether or not they preserve the original point values; if a method preserves those points, it is called exact, if not, it is called approximate. An interpolation method can be for prediction or for characterization. A method is for prediction if it is good at minimizing prediction errors at unknown points; a method for characterization produces a surface that globally looks like the measured surface (Caruso and Quarta 1998). Another classification characterizes an interpolation method as global or local. The method is global if models long range variations, as can be the case of IDW. The method is local if it uses a specific number of neighboring points to estimate the value of an unsampled point, making the accommodation of anomalies feasible without affecting the value of the interpolation at other points. Spline is a local method (Hartkamp et al. 1999). If IDW uses only neighboring points it becomes local.

The geointerpolation methods implemented in the ArcGIS Spatial Analyst Extension of the ArcGIS Desktop software offer a point of departure to decide as to what interpolation to use. These methods are inverse distance weighting, kriging, spline, and natural neighbors.

No single spatial interpolation method is good for all applications. The applicability of an interpolation method depends on several factors including:

- Point density, many points or few points per given area;
- Spacing, points located irregularly or regularly in the studied region;
- Variable considered, such as water elevation, air temperature, snow water equivalent, and others;
- Statistical properties of the variable, such as range, variance, and correlation;

- The extent and characteristics of the geographic area where the interpolations are to be made;
- And, the criteria used to determine a good fit.

2.3.1 Inverse Distance Weighting

Inverse Distance Weighting (IDW) interpolation estimates point values using linearly weighted sample data points located around that point. Data points closer to the estimated point have more influence or weight, while those farther from the estimated point have less influence. In other words, influence on estimated point values decreases with distance. A functional interpolation method, IDW is an exact interpolator, one whose predicted values are equal to the measured values at the sample point locations. Inverse Distance Weighting is computed using the equation

$$I_i = \frac{\sum_{j \neq i} \frac{M_j}{d_{i,j}^p}}{\sum_{j \neq i} \frac{1}{d_{i,j}^p}} \tag{2.2}$$

or

$$I_i = \sum_{j \neq i} w_j M_j \tag{2.3}$$

where the weighting factor is

$$w_j = \frac{\frac{1}{d_{i,j}^p}}{\sum_{j \neq i} \frac{1}{d_{i,j}^p}} \tag{2.4}$$

I_i is the interpolated value at point i, M_j represents the measured value at point j; $d_{i,j} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$ is the distance between point i (x_i, y_i) and point j (x_j, y_j); p is the power to which the distance between points i and j is raised. Each point j is part of the set of points used to interpolate the value in point i. In the context of our problem, there are r points (that is, r measuring stations being evaluated) while the point i (or station i) is part of the remaining $n - r$ stations being considered for exclusion from the measurement set. If the distance $d_{i,j} < d_{min}$ (d_{min} is a minimum distance) then $d_{i,j} = d_{min}$ to prevent an infinite weight. If $d_{i,j} > d_{max}$ (d_{max} is a maximum distance) then $d_{i,j} = d_{max}$ to prevent the consideration of excessively distant points.

Three parameters can be used in the implementation of different behaviors of IDW interpolation: power, search radius, and barrier. A high power, p, gives more influence to points in closest proximity, producing rough surfaces. A lower power, p, gives more influence to points farther away, producing smooth surfaces. A search radius only permits the consideration of points located within the circle defined by the radius. A fixed or variable search radius can be defined and used when there are a specified minimum number of points; if there are not enough points the search radius can be increased. Otherwise, a variable search radius can be used defining the number of points to be employed in the interpolation. In a general case, known points are located at irregular locations, producing a varying search radius from point to point. Additionally, a maximum search radius could be specified, limiting the number of points included in the interpolation.

Commonly, a variable search radius is used when the modeled phenomenon has a great amount of variation. Finally, barriers can be implemented to define break lines to avoid the inclusion of unwanted data points located on the other side of the barrier.

Among the advantages of IDW are its simplicity, reasonable results for a wide variety of data, and, in general, meaningful interpolated results. Among its disadvantages are: its high sensitivity to w_i (the weighting function); its fragility to uneven point distributions; and its incapability to compute the global maximum and minimum outside the known points (Caruso et al. 1998). If there is extreme spatial variability, as in the case of snow cover distribution, IDW does not reliably estimate values (Erxleben et al. 2002). For climatology variables and in the mining industry, IDW is a common approach (Hartkamp et al. 1999).

2.3.2 Kriging, Spline and Natural Neighbor

Kriging interpolation fits a mathematical function to a specified number of known values to determine values for specified points or for an entire continuous surface. The kriging interpolation method estimates values, like IDW, by assigning a weight based on the distance between measured points and the estimation location. Kriging, apart from the distance between mentioned points, is dependent on the overall spatial arrangement among the measured points. The weight w_j is determined from a model fit to the measured points, the distance to the prediction location, and the spatial relationships among the measured values around the prediction location. In fact, there are different types of kriging interpolation, as described in Caruso et al. (1998), Hartkamp et al. (1999), Sárközy (1999), and Erxleben et al. (2002). Two types of kriging are ordinary and universal kriging; the difference between them is that ordinary kriging assumes that the data has an unknown constant trend and universal kriging assumes that the trend of the data varies and that the regression coefficients of the trend are unknown. Erxleben et al (2002) found ordinary kriging to be inadequate when predicting extreme spatially variable surfaces. Collins and Bolstad (1996) found that kriging was superior to IDW and spline interpolation for temperature estimation, and effectively described surfaces with regionalized variables. In an experimental comparison study of ordinary and universal kriging and IDW where the effects of surface type, sampling pattern, noise level, and strength of small-scale spatial correlation were considered, Zimmerman et al (1999) found that the kriging methods were superior to the IDW methods over all the factors. Three types of surfaces were used: a tilted plane, a strong peak based on a sin function (a "sombrero") and an undulating, increasing surface, which is the sum of sine and cosine functions (a Morrison's surface). For the plane, universal kriging was better than ordinary kriging but the difference was not statistically significant. The same occurred for the Morrison's surface. For the sombrero, ordinary kriging was better than universal kriging. According to the authors, this result is interesting because it indicates that for some types of surfaces it is better to completely ignore the modeling of trend, as ordinary kriging does, than to model trend inappropriately, as apparently universal kriging does, for the sombrero. Considering the method pattern-interaction, ordinary kriging performed better than universal kriging for the clustered pattern, whereas ordinary kriging and universal kriging were not significantly different for the other three patterns that include a simple random pattern. Unlike most other interpolation methods, kriging provides a measure of error or uncertainty of the interpolated surface, which may prove useful in identifying areas where more data would be helpful.

Spline interpolation, or thin plate interpolation, states that the predicted surface must pass exactly through known data points and the surface must have minimum curvature, which is achieved by minimizing the cumulative sums of the squares of the second derivative terms of the predicted surface. The minimum-curvature technique produces a smooth, continuous, and differentiable surface. Typically useful for dense, regularly spaced data, spline interpolation is generally not recommended for interpolation of irregularly spaced data (Collins and Bolstad 1996). In addition, spline interpolation provides no estimation of error.

Natural Neighbor, like IDW, is a weighted-average interpolation method. It determines the weights with the well-known Theissen technique used in the estimation of basin average precipitation. First, construct the Theissen polygons, also known as Voronoi cells, of the monitoring stations with data (Figure 2.7).

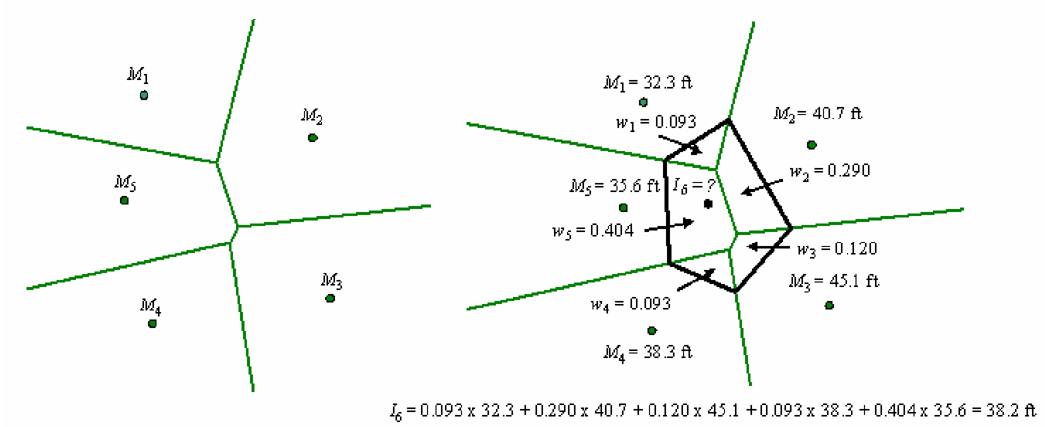


Figure 2.7: Natural neighbor interpolation.

Second, to estimate the value in the station *i*, it is necessary to redefine the cells with the addition of the new station *i*. Third, two sets of cells are now available, the original one and the new cell system constructed after inserting the new point. The cell of the new point covers some parts of the cells originally owned by particular monitoring stations. These particular stations must be involved into the interpolation of the new point. These monitoring points are called natural neighbors (Sárközy 1999). The weights of the natural neighbors are the areas that the new cell cuts out from the original cell owned by a particular neighbor divided by the area of the new cell. Mathematically, the interpolated value in the new point *i* is

$$I_i = \frac{\sum_j a_j M_j}{\sum_j a_j} = \frac{\sum_j a_j M_j}{A_i} = \sum_j \frac{a_j}{A_i} M_j = \sum_j w_j M_j \tag{2.5}$$

where a_j represents the area that the new cell *i* cuts out from the original cell owned by neighbor *j*. The sum of cut out areas is A_i, the area of the new cell *i*, and the weight of each neighbor is $w_j = a_j/A_i$. Natural Neighbor interpolation does not require specification of parameters such as radius, number of neighbors, or weights, making it a very good general-purpose interpolation method.

Sorenson and Maidment (2004) used IDW, kriging, spline and natural neighbor methods for computing depth elevation surfaces in a study where there was a need to predict the spatial variation of depth and duration of inundation in the S-65BC sub-basin of the Kissimmee River basin (see Figure 1.2). They found that the differences in the results of all the tested interpolation methods were not important, when compared to the effect of estimation of water depth of two alternative land surface elevation grids. An additional interesting finding is the great variability of mean error spatial distribution between different land surface elevation grids; this finding suggests that something other than interpolation method influences water depth prediction. Sorenson and Maidment concluded that is possible that geospatial interpolation is not the best way to predict water depths, that hydraulic modeling may be required to compute water depths more accurately.

2.4 HYDRAULIC MODELING AND OPTIMIZING

Hydraulic modeling of water movement in a hydrologic system is based on the conservation of mass and momentum. This modeling is frequently done in one dimension in the direction of flow. One-dimensional models are based on the solution of the Saint-Venant equations for one-dimensional unsteady flow in open channels; these equations mathematically embody the principles of conservation of mass and momentum. Several available one-dimensional models may be adequate for the purposes of this research. Among them are HEC-RAS, MIKE 11 and FLDWAV. Any of these models coupled with an optimization method may be used to develop a methodology to optimize stage-monitoring networks located along streams.

2.4.1 One-Dimensional Hydraulic Models

One-dimensional hydraulic modeling is the simplest formulation that is capable of representing flows circulating within networks of channels. Advantages of one-dimensional models (Singh 2004) are: fast computations; relatively little field data is required to set up the models; and, traditionally one-dimensional models have more powerful capabilities to describe control structures than do higher-dimensional models. However, the use of one-dimensional models can lead to inaccurate results; most of the errors come from severe violations of one or more of the basic assumptions such as one-dimensional flow, uniform velocity and horizontal water level across the cross section, or small bottom slope (Kutija 2003). Building on the one-dimensional Saint-Venant equations, quasi two- dimensional hydrodynamic models can be defined. Flow in the main channel is described by the one-dimensional equations while the land surrounding the main channel is divided into compartments. Water level at each compartment is solved using the continuity equation while the flows across the compartment boundaries are calculated based on the water levels in the adjacent compartments and the boundary characteristics. This approach has considerable advantages over the one-dimensional models in modeling multidimensional flow situations.

However, compared with the one-dimensional hydraulic models an extensive increase in computation time can be expected (Kutija 2003).

Data requirements of one-dimensional models (Singh 2004) include the following:

- Basic data: data that describes the physical layout of the river or the catchments being modeled. This includes, for instance, cross-sections, weir crest levels and dimensions, catchments areas, etc.
- Parameters: parameters that are used in the equations being solved by the specific hydraulic model. Examples of such parameters are Manning's *n* roughness coefficients and contraction/expansion loss coefficients. In unsteady flow models, adequate computational time steps and intervals of sampling of discharge and water level time series could also considered very important parameters. However, most parameters have to be estimated through calibration and verification processes, literature review, or from experience.
- Boundary data: models require boundary conditions; typically, these are time series of measurements such as discharge or water level at the boundaries of the modeled system.
- Calibration and verification data: these data are not required to run the model, having the three previous data groups available are sufficient. However, if model parameters need to be estimated, then additional data for calibration and verification are required. These additional data are, for instance, measured time series of discharge and water level.

General descriptions of HEC-RAS, MIKE 11, and FLDWAV (three important models), are included in the next subsections.

2.4.2 HEC-RAS

HEC-RAS, the River Analysis System Software developed by the U.S. Corps of Engineers' Hydrologic Engineering Center (HEC), is an MS Windows application comprised of a graphical user interface, several hydraulic analysis components, and functionality to store, manage, graph and report data and results. In version 3.1.2 (USACE 2004), the HEC-RAS model contains two one-dimensional hydraulic analysis options, one for open-channel steady flow computations and other for unsteady flow simulations. In future versions, it will also include movable boundary-sediment transport computations. The two current components use a common geometric data representation and common geometric and hydraulic computation routines. HEC-RAS performs one-dimensional hydraulic calculations for a full network of natural and constructed channels. In addition to the two hydraulic analysis options, the system contains several hydraulic design features that can be invoked once the basic water surface profiles are computed.

The steady flow component is intended for calculating water surface profiles for steady gradually varied flow. The system can handle a looped network of channels, a dendritic system, or a single river reach. This component allows the modeling of subcritical, supercritical, and mixed flow profiles. Its computational procedure is based on the solution of the one-dimensional energy equation. Energy losses are evaluated by friction using Manning's equation and by contraction/expansion applying a coefficient to the change in velocity head. The momentum equation is used in situations where the water surface profile varies rapidly. These situations include mixed flow regime calculations, such as hydraulic jumps, hydraulics of bridges and stream junctions. The effects of various obstructions such as bridges, culverts, weirs, and structures in the flood plain may be considered in the computations. The steady flow component is designed to evaluate floodway encroachments for flood plain management and flood insurance studies. In addition, the change in water surface profiles due to channel and levee modifications can be assessed. Special features of the steady flow component include multiple plan analyses, multiple profile computations, multiple bridge and/or culvert opening analysis, and split flow optimizations.

The unsteady flow simulation component can simulate one-dimensional unsteady flow through a full network of open channels. The unsteady flow equation solver was adapted from Dr. Robert L. Barkau's UNET model (USACE 2004). This unsteady flow component was developed primarily for subcritical flow regime calculations. However, with the release of Version 3.1, the model can perform mixed flow regime (subcritical, supercritical, hydraulic jumps and drawdown) calculations. The hydraulic calculations for cross-sections, bridges, culverts, and other hydraulic structures developed for the steady flow component were incorporated into the unsteady flow module.

In HEC-RAS, data are stored in ASCII and binary files as well as in HEC-DSS files. The HEC Data Storage System (HEC-DSS) is a database format specifically developed for hydrological time-series data by the HEC. User input data are stored in ASCII files under separated categories of project, plan, geometry, steady flow and unsteady flow. Output data is predominantly stored in separate binary files. Data can be transferred between HEC-RAS and other programs utilizing HEC-DSS files. The user interface allows the user to manage his/her data. The user is requested to enter a single filename for the project being developed. Once the project filename is entered, all other files are automatically created and named by the interface as needed; the interface renames, moves, and deletes files on a project-by-project basis.

Graphics include *X*-*Y* plots of the river system schematic, cross-sections, profiles, hydrographs, and other hydraulic variables. A three-dimensional plot of multiple cross-sections is also provided. Tabular output is available. Users can select from predefined tables or develop customized tables. All graphical and tabular output can be displayed on the screen, sent to a printer or plotter, or, using the Windows clipboard, passed to a word processor or spreadsheet. Predefined or customized reports of data and results can be printed out.

2.4.3 MIKE 11

One important commercial software package is MIKE 11 (the current version is version 2005), published by DHI Water & Environment, an international consulting and research organization with headquarters in Denmark (DHI Water & Environment 2005). MIKE 11 is a versatile and modular engineering tool for modeling conditions in rivers, lakes/reservoirs, irrigation canals and other inland water systems. It is designed for flood risk analysis and mapping, design of flood alleviation systems, real-time flood forecasting, real-time water quality forecasting and pollutant tracking, hydraulic analysis/design of structures including bridges, drainage and irrigation studies, optimization of river and reservoir operations, dam break analysis, water quality issues, integrated groundwater and surface water analysis.

The hydrodynamic module, which is the core of MIKE 11, contains an implicit, finite difference computation of unsteady flows in rivers and estuaries. The formulations can be applied to branched and looped networks and quasi two-dimensional flow simulation on flood plains. The MIKE 11 computational scheme is applicable to vertically homogeneous flow conditions ranging from steep river flows to tidally influenced estuaries. Both subcritical and supercritical flow can be described by means of a numerical scheme, which adapts according to the local flow conditions.

The complete non-linear equations of open channel flow (Saint-Venant) can be solved numerically between all grid points at specified time intervals for given boundary conditions. In addition to this fully dynamic description, a choice of other flow descriptions is available: high-order, fully dynamic, diffusive wave, kinematic wave, quasi-steady state, and kinematic routing (Muskingum, Muskingum-Cunge). Within the standard hydrodynamic model, advanced computational formulations enable flow over a variety of structures: broad-crested weirs, culverts, bridges, pumps, regulating structures, control structures, dam-break structures, user-defined structures, and tabulated structures. The hydrodynamic module is available in a number of different sizes, and according to the manufacturer, can be upgraded at any time. A number of add-on modules are furthermore available so that the system can be tailored to suit specific requirements. One add-on model is the Quasi Steady State module, which provides a rapid, though simplified, method of solving the mass and flow equations. This module is often used in long-term morphological studies or in long-term hydraulic simulations. It may also be used as an aid in defining initial conditions when preparing a dynamic hydrodynamic simulation.

2.4.4 FLDWAV

The National Weather Service (NWS) FLDWAV model is a combination of the NWS DAMBRK and DWOPER models (NOAA 2005). FLDWAV includes the best capabilities of both models and a few enhancements that make it a better model. FLDWAV is a generalized flood routing model that can be used for real-time flood forecasting of dam-break and/or natural floods, dam-breach flood analysis or overtopping, floodplain inundation mapping for emergency planning, and design of waterway improvements. The model can compute the outflow flood wave hydrograph from a dam, or the flood wave hydrograph can be user specified. The flood wave is then routed through the downstream channel using a four-point implicit finite-difference numerical solution of the complete Saint-Venant equations of one-dimensional unsteady flow along with appropriate internal boundary equations representing downstream structures (NOAA 2005).

In FLDWAV, river systems that have a dendritic structure (*n*th-order tributaries) can be modeled as well as channel bifurcations (islands). The model also can route unsteady flows occurring simultaneously in a system of interconnected rivers. Any of the rivers may have one or more structures (dams, bridges, levees, etc.) which control the flow and which may breach if failure conditions are reached. FLDWAV can handle subcritical flow, supercritical flow, or a combination of each, varying in space and time from one to another. A new computational scheme (LPI) has been developed to model mixed flow (NOAA 2005).

Initial conditions include the initial water surface elevations and discharges at each of the read-in cross section locations. FLDWAV can start up in either a steady state or an unsteady state condition. The initial computational time step may be read in or generated by the model. A Manning's *n* table is defined for each channel reach bounded by gauging stations and is specified as a function of either water surface elevation, *h*, or discharge, *Q*. FLWAV can automatically determine Manning's *n* so that the difference between computed water elevations hydrographs and corresponding observed hydrographs is minimized. In the Local Partial Inertial (LPI) mixed-flow algorithm, used in FLDWAV, portions of the routing reach with low Froude numbers are modeled with the inertial terms in the momentum equation included, while those portions with high Froude numbers are modeled with little or no inertial terms included, improving numerical stability (Fread and Lewis 1998).

A characteristics-based explicit scheme has been added to FLDWAV to model instantaneous dam failures and very rapidly occurring failures with a time of failure less than 3 minutes. This scheme is also applicable to the complicated flows in the mixed-flow regime (Jin and Fread 1997). FLDWAV has the capability of using multiple routing techniques in a river system. Currently, there are four routing techniques available: dynamic implicit, dynamic explicit, level pool (storage), and diffusion. In each reach between adjacent cross sections, the user assigns a routing technique. The LPI computational scheme may also be applied to specific rivers. The user can specify multiple computational time steps throughout the temporal range of the inflow hydrograph. The capability of modeling mud/debris unsteady flows by including an additional friction slope term in the momentum equation of the Saint-Venant equations has been added to FLDWAV.

The FLDWAV model is run in command line form. Input and output files are needed for it to run. It produces several output color graphics displays (hydrographs, peak profiles, cross sections, and stage-discharge relationships). Additionally, an interactive input program is under development to make FLDWAV user friendly (NOAA 2005). The current version of NWS FLDWAV is version 2-0-0 published on June 1, 2000. For a more detailed description of the FLDWAV model, see papers by Fread and Lewis (1988) and Fread (1993).

2.4.5 Water Resources Optimization Problems Solved using Open Channel Hydraulics Models

Nicklow and collaborators (2001, 2003 and 2004) have studied the optimization of dam operations using global search optimization methods. In all cases, they modeled the system with FLDWAV. This body of work demonstrates the feasibility of coupling a one-dimensional open channel hydraulic model to global search optimization methods. Details of these authors' work follows.

Minder and Nicklow (2001) present a methodology and model for the time dependent evaluation and optimization of dam operations. The optimal dam operation scheme chosen minimizes the environmental impact caused by water level fluctuations while maintaining navigable stages within the river. This problem is formulated as a discrete-time optimal control problem in which a hybrid genetic algorithm is interfaced with the FLDWAV model. FLDWAV solves the governing hydraulic constraints, while the genetic algorithm solves the overall control problem. The application of the model is demonstrated in the optimization of a hypothetical three reservoir river network. Authors conclude that the preliminary results of the methodology are satisfactory but the application to a real hydraulic network will be performed in future work.

Dessalegne and Nickow (2003) present a discrete-time optimal methodology to minimize water level fluctuations within multi-reservoir river networks while meeting required stages at user defined cross-sections, reservoir storage levels, bounds on dam releases and gate openings, and minimum pool levels behind locks and dams. The presented methodology integrates a statistical preprocessor tool, the FLDWAV model, a genetic algorithm and a simulated annealing algorithm, as optional optimizers. The preprocessor determines the historical range of water levels at desired cross-sections, while the simulator solves the governing hydraulic equations for open channel flow. The optimizer solves the overall control problem and identifies the optimal dam operation policies. Operational constraints such as water elevation and storage level bounds and minimum level for navigation are handled by additive penalty functions. The methodology is successfully applied to a hypothetical three-reservoir river system and a major portion of the Illinois River Waterway.

Dessalegne and Nicklow (2004) applied an optimization algorithm derived from the artificial life paradigm to determine the optimal dam operations in a multi-reservoir river system. These authors used the same optimal dam operation scheme of their anterior work. Again, they coupled the optimization method with FLDWAV. This model was found capable to solve the stated problem.

2.5 CONCLUSIONS OF CHAPTER

Two optimization procedures are needed to optimize the stage-monitoring network of the SFWMD, one for lakes, where water levels tend to be smooth and planar, and the other for streams, where water levels can have significant changes produced by the varying hydraulic properties along the courses. Both methods are intended for identifying the most important gages of an existing network; those operating gages that provide the minimum loss of information in the estimation of water surface elevation in the location of the non-operating gages. The two important subjects upon which to base the solution of the stated problem are optimization and interpolation. The more promising optimization methods, namely, simulated annealing, genetic algorithms and tabu search, were reviewed, finding several successful monitoring network applications of them. For example, Pardo-Igúzquiza (1998) proposes an optimal network design for the estimation of spatial averages of precipitation using simulated annealing. In this research, the optimization of the stage-monitoring network of lakes is approached using simulated annealing, a genetic algorithm and tabu search.

The existence of stage-monitoring stations in lakes and in streams requires that two types of interpolation methods should be studied and applied. The first interpolation method, applicable to lakes, is a more geometric interpretation than a physical interpretation because it is based on representation of surfaces. The methods studied for geometric interpolation were inverse distance weighting, kriging, spline, and natural neighbor. Pardo-Igúzquiza (1998) uses kriging as its interpolation method; in lakes, this research uses inverse distance weighting. The second type of interpolation methods are more physical than geometric because they are based on hydraulic principles. Three computational models, HEC-RAS, MIKE 11 and FLDWAV, can act as hydraulic interpolators. Both types of interpolators can be coupled with an optimization method to optimize stage-monitoring networks. Published evidence (Minder and Nicklow (2001), Dessalegne and Nicklow (2003), and Dessalegne and Nicklow (2004)), of the coupling of FLDWAV, a hydraulic model, with several optimization methods in the optimization of dam operations demonstrates the feasibility of the approach for optimizing the stage-monitoring network along streams. The optimization of the stage monitoring in a stream using HEC-RAS combined with a sequential optimization method and tabu search is studied here. In summary, the revised literature and computational models offer a good foundation to build the methodologies needed.

Chapter 3: Methodology

3.1 INTRODUCTION

The development of the methodologies for optimizing the stage-monitoring network of the SFWMD was based on the study of the network of the Kissimmee River Basin considering measured daily data from October 1, 2001 to September 30, 2004. Several case studies were selected. A set of customized computer programs that make use of the ArcGIS and HEC-RAS software were developed. The research methodology used to develop the optimization methodologies has four phases: one dedicated to the network optimization in lakes, another to network optimization in streams with steady flow, another to network optimization in streams with unsteady flow and another to study the use of a different optimization method in the cases of lakes and streams in steady flow.

3.2 OPTIMIZATION OF STAGE-MONITORING NETWORKS IN LAKES

The development of the methodology for optimizing the stage-monitoring network in lakes considered three steps: one to compare optimization methods, another to develop a preliminary methodology, and the third to develop the definitive methodology.

3.2.1 Procedure for comparing Simulated Annealing and Genetic Algorithm Methods Performance

To keep the analysis simple, not involving temporal change of stage, two case studies were selected, case study 1 is a synthetic daily dataset and case study

2 considers 17 stage-monitoring stations of the southern portion of subbasin S-65BC, also known as Pool BC, operating on August 8, 2002. The data of these case studies is shown in Table 3.1 and Table 3.2. Figure 3.1 also shows the stations of subbasin S-65BC. Both datasets are not properly data of a lake; the first one resembles a tilted plane, the second is a river segment including the flood plain. Both can be considered extreme cases useful to judge which of the optimization methods works better.

Station	*x* (mi)	*y* (mi)	*z* (ft)
1	16	27	37.7
2	34	9	32.8
3	99	4	26.1
4	57	74	44.2
5	22	53	43
6	98	58	39.2
7	99	29	32.3
8	80	3	27.3
9	32	81	47
10	24	67	46.5
11	23	37	39
12	40	54	42
13	58	55	41
14	72	40	36.8
15	75	23	32
16	53	34	36.3
17	47	22	34

Table 3.1: Dataset of case study 1. Synthetic values.

Station	*x* (ft)	*y* (ft)	*z* (ft)
PC31	614232.2394	1132814.735	38.62
PC32	608376.8805	1132305.672	37.58
PC33	610889.2972	1126335.113	36.7
PC42	603292.4501	1138107.402	38.43
PC51	597015.5841	1148152.539	39.39
PC52	595254.3778	1147195.726	39.5
KRBNS	600720.4775	1136932.592	38.5
KRDRS	592257.4124	1145918.466	39.57
PC12	612622.2984	1115113.433	35.68
PC21	608139.1486	1121503.439	36.5
PC22	603613.9042	1121857.563	36.7
PC34	604076.2238	1128513.657	37.49
PC35	602128.1908	1127617.1	37.37
PC44	599079.7837	1135695.141	38.39
PC45	597848.3582	1134844.598	38.34
PC53	593305.2225	1146612.242	39.57
PC11R	613172.465	1117307.001	35.59

Table 3.2: Dataset of case study 2. Seventeen stage-monitoring stations of subbasin S-65BC on August 08, 2002.

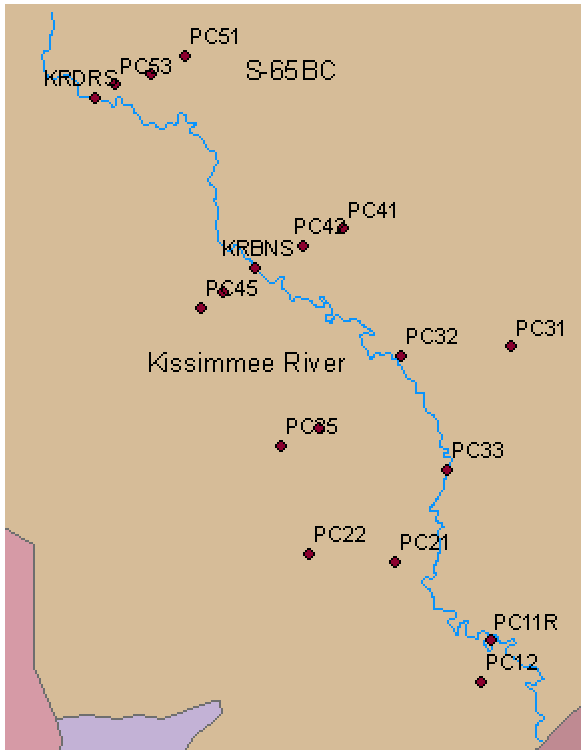


Figure 3.1: Seventeen stage-monitoring stations of basin S-65BC. On August 8, 2002, station PC41 has no data. (This is not a lake, but it can be considered as an extreme case useful to test optimization methods).

The solution of the problem of developing a methodology to obtain the optimal subset of stage-monitoring stations of a lake needs a measure of error. A plausible measure of error is the root mean square error of interpolated values of water surface elevation in the lake. Assume the existence of a set of *n* water level measuring stations from which is needed to choose a subset of *r* stations. The root mean square error defined over the $n - r$ omitted stations is:

$$RMSE = \sqrt{\frac{1}{n-r}\sum_{i=1}^{n-r}\left(I_i - M_i\right)^2} \tag{3.1}$$

where I_i is the interpolated value and M_i is the measured value, both for station i.

The number of subsets of r stations taken from a set of n stations can be computed with:

$$C_r^n = \frac{n!}{r!(n-r)!} \tag{3.2}$$

Figure 3.2 shows the number of combinations for several n and r values. It can be seen that the number of combinations grows with n. Furthermore, if n grows and r is close to n/2 the number of combinations grows rapidly according to:

$$C_{n/2}^n = \begin{cases} \dfrac{n!}{\left[\left(\dfrac{n}{2}\right)!\right]^2}, \text{if } n \text{ is even} \\ \dfrac{n!}{\left(\dfrac{n+1}{2}\right)!\left(\dfrac{n-1}{2}\right)!}, \text{if } n \text{ is odd} \end{cases} \tag{3.3}$$

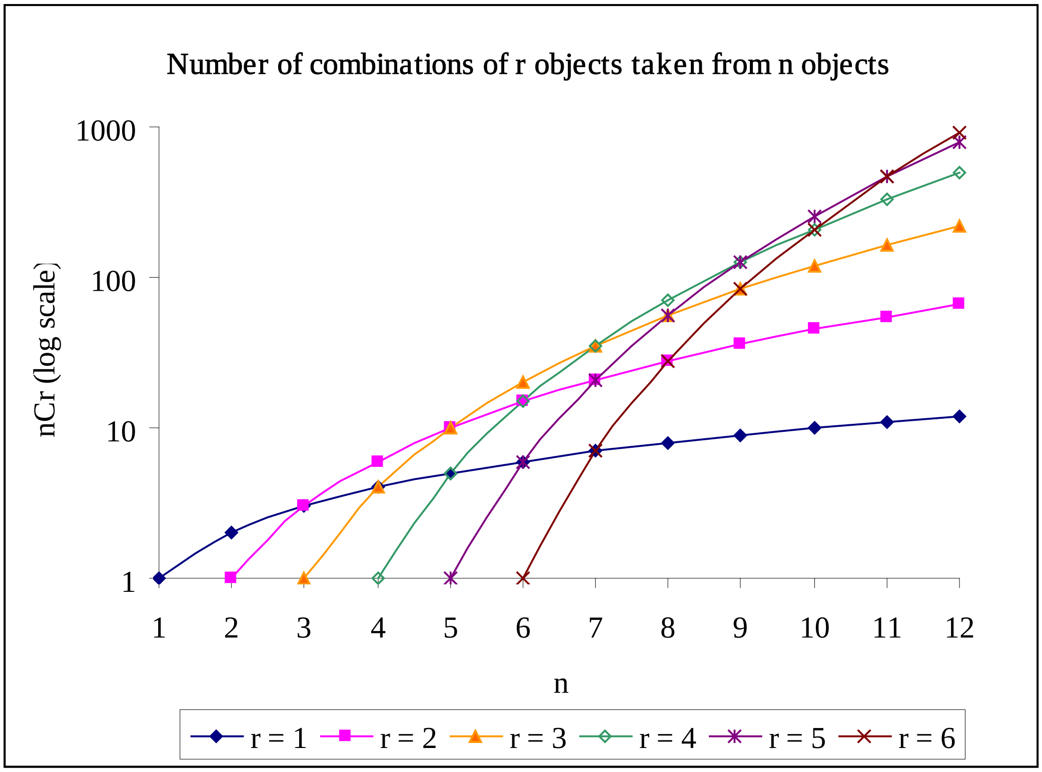


Figure 3.2: Number of combinations of r objects taken from n objects.

A graph that illustrates the more complex problem that involves also time is shown in Figure 3.3. In the figure, the stations $s \in S$ have stage measurements for the times $t \in T$. One approach to solve this problem can be to find the optimal subsets of all given times $t = t^*$ and then decide which of the optimal subsets is the optimal subset of the period T.

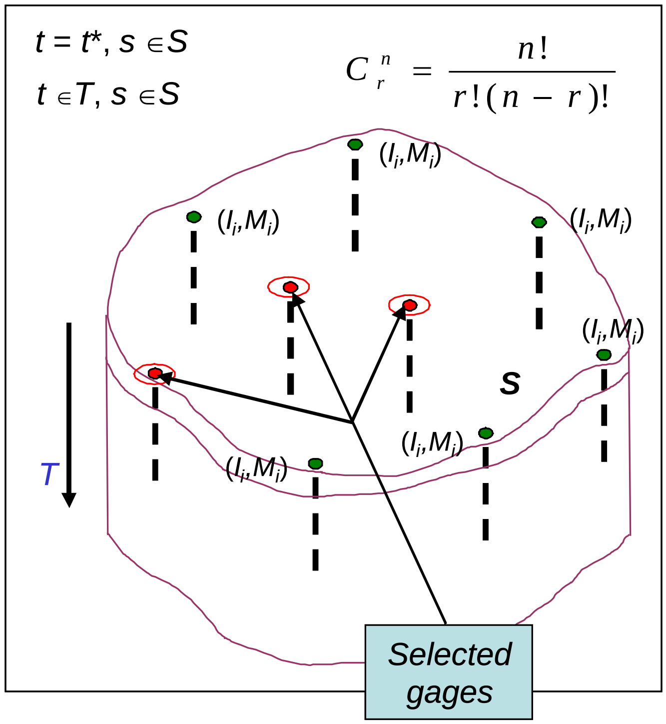


Figure 3.3: Graphic description of the optimization problem.

The minimization of the *RMSE* function is a problem of combinatorial optimization, analogous to the one discussed by Pardo-Igúzquiza (1998). This problem does not have an analytical solution, is not linear, depends on measured data and has multiple local minima. According to the literature review, this problem could be solved by applying simulated annealing or a genetic algorithm. In the solution with simulated annealing, the objective function is the energy, and the temperature is a fictitious parameter calibrated for the given problem. In the case of the genetic algorithm, an individual is a given set of *r* stations taken from *n* stations. The parameter of fitness is the corresponding root mean square error *RMSE*. In a given population, the best-fitted individual (the best set of *r* stations) is that which has the least value of *RMSE*. That set of *r* stations is more likely to be in the next generation, one or more times, compared with less fitted sets. To estimate water elevations in a lake, the method of interpolation should be a spatial interpolation method. From the reviewed methods, a convenient method is the inverse distance weighted method (IDW) (with a power equal to 2). The IDW method was chosen for two reasons. First, it is easy to implement, and second, Sorenson and Maidment (2004) found that any of the four interpolation methods provides a reasonable result for interpolating water surface elevations. Equation 2.2 with $p = 2$ is used

$$I_i = \frac{\sum_{j \neq i} \frac{M_j}{d_{i,j}^2}}{\sum_{j \neq i} \frac{1}{d_{i,j}^2}} \tag{2.2'}$$

where $d_{i,j}$ represents the geographic distance between station *i* and station *j*. All stations *j* are in the subset of *r* stations while station *i* is the subset of $n - r$ stations (Figure 3.4).

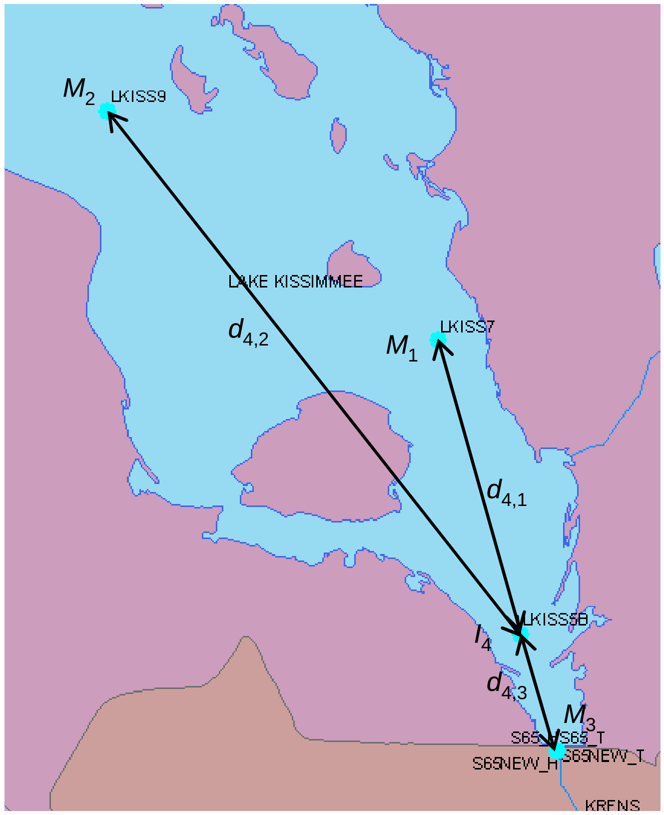


Figure 3.4: Distances between stations

For adopting the best optimization method for the proposed problem, preliminary implementations of both methods were prepared. These implementations were written in Microsoft's Visual Basic for Applications (VBA) to take advantage of this programming language that is included in ESRI's ArcMap software. The inputs to both methods are the (x, y) coordinates and the measured stage value at each monitoring point, n (number of total monitoring points), and r (number of monitoring points to be included in the optimal subset). The output is the subset of monitoring stations with the least *RMSE*.

3.2.1.1 Application of the Simulated Annealing Method

Important details of the application of the simulated annealing method, used in this research are:

- An initial subset is probabilistically chosen. Each subsequent subset is obtained from the current subset interchanging randomly only one station in the subset (of r stations) with a station out of the subset (chosen of $n - r$ stations).
- The energy of a selected set of stations is measured with its value of the root mean square error, *RMSE*.
- An initial temperature, T_o, is obtained choosing the greater value between the variance of the measured stages and the maximum value of the measured stages.
- The maximum number of iterations, it_{max}, is equal to the number of combinations of n stations taken in subsets of r stations. This maximum number of iterations limits the number of iterations to the worse case scenario of performing an exhaustive evaluation of all combinations.
- The maximum number of iterations after the last optimum value obtained, it_{opmax}, is equal to the maximum number of iterations divided by 2.
- A combination or subset of stations is accepted according with the value of the increment in *RMSE* between the current combination and the current optimal combination. That is, *df = RMSE of current combination – RMSE of current optimal combination*. If *df* is greater than zero, the current combination is accepted becoming the current optimal

combination. If df is negative or equal to zero the current combination is accepted with a probability (Equation 2.1) of $\exp\left(-\frac{df}{T}\right)$, where T is the current temperature.

- A change in temperature is given when the number of accepted combinations after the last change of temperature is greater than it_{max} /100 or the number of tried combinations after the last change of temperature is greater than it_{max}/10. The new temperature is equal to 0.25 times the previous temperature ($T \leftarrow 0.25\, T$).
- The process ends when the number of iterations is greater than or equal to it_{max} or the number of iterations after the last optimum is greater than or equal to it_{opmax}.

3.2.1.2 Application of the Genetic Algorithm Method

The genetic algorithm method used in this research is applied as follows:

- The population size, n_p, is computed as a function of the number of combinations with the empirical expression,

$$n_p = \mathrm{int}\left(3\ln\frac{n!}{r!\left(n-r\right)!}\right) \text{, If } n_p \text{ is odd, } n_p = n_p + 1 \tag{3.4}$$

obtained in this study after several numerical experiments were conducted to find small populations that still allow the genetic algorithm to find optimal results.

- Each chromosome or individual is an array of n Boolean variables. In this study, a chromosome represents a combination of stations. If a particular station i, is in a chromosome, the corresponding gene i, has a value of true.

As each Boolean variable has two possible values, n bits with r bits turned on are used, i.e. r values of 1 and $n - r$ values of 0.

- Maximum number of iterations, $it_{max} = 10\ n_p$
- Maximum number of iterations after last optimum, $it_{opmax} = it_{max} / 4$.
- Fitness function: *RMSE*
- Probability of selection of individual j, $P_s(j)$, is assigned using $P_s(j) = (MaxRMSE - RMSE(j))/SumRMSE$, where *MaxRMSE* is the maximum value of the *RMSE* of all individuals and *SumRMSE* is the sum of the *RMSE* of all individuals.
- Each pair of parents in a population could interchange one gene; that is, each pair of subsets of stations could interchange one station with a probability of crossover equal to 0.5.
- Each turn-on gene has a probability of mutation equal to 0.1. When mutations occur, a number of genes are turned off and the same number of genes is turned on.

3.2.1.3 Criteria for Comparison

The developed implementations of a genetic algorithm and simulated annealing were compared according to:

- Computing time. The measurement of computing time is referred to a Dell Workstation PWS360 with a 3.2 GHz Pentium 4 Processor and 1.5 GB of RAM running VBA in Windows XP.
- Whether the method reached the optimal *RMSE* at least once for each r considered.

- Ratio between the obtained maximum and minimum values of the optimal *RMSE*. The true optimal value might not be obtained, only approximations to it bounded by the maximum and minimum values mentioned.
- Ratio of number of successful realizations to the total number of realizations. Given that these methods find the optimal value with a probabilistic search, each one of the iterations to find an optimal subset is called a realization. A realization is successful if the true minima of the objective function is found, and is unsuccessful otherwise.

Once the parameters of both optimization methods were calibrated, the methods were applied to both datasets. All possible sizes of subset, from one to 16 stations were considered. Twenty realizations were computed for each size of subset and method. The genetic algorithm method had a better performance than the simulated annealing method. Results are shown in Chapter 5.

3.2.2 Development of the First Methodology for Lakes

Having a better performance than the simulated annealing method, the genetic algorithm was adopted in the first methodology for optimizing the stage-monitoring network for a lake. At this point in the research, it was not clear that different optimization methodologies were needed for lakes and for streams, so a case study that includes stream and floodplain stations was selected. This is the case study 3; it is located in the subbasin S-65BC and considers stations in the northern and southern portions of the subbasin. The methodology used genetic algorithm as the optimization method and IDW as the interpolation method. Three one-year consecutive periods of daily stage measurement were used to find the optimal subset of water elevation stations: from October 1, 2001, to September 30, 2004. Figure 3.5 shows the subbasin S-65BC's stations, and Table 3.3 shows the names and geographic coordinates of stations.

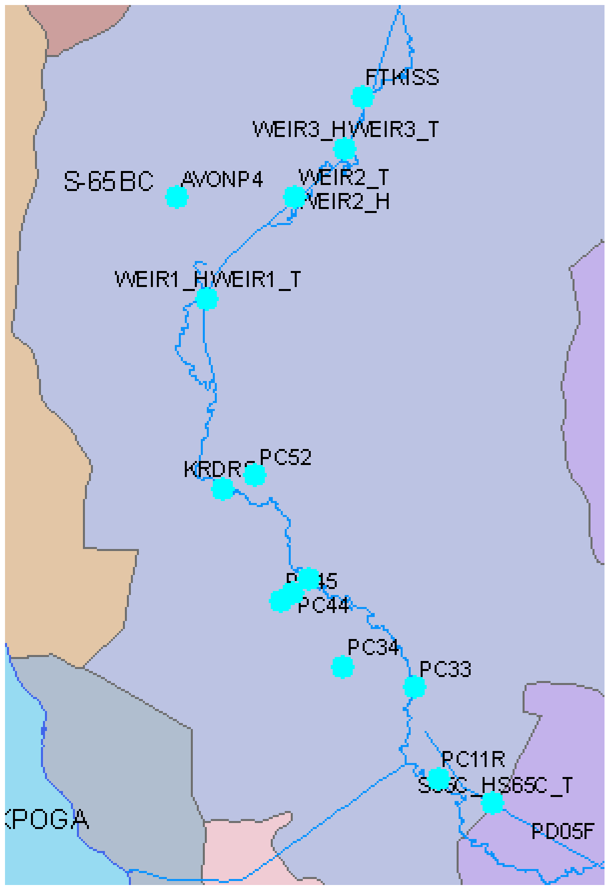


Figure 3.5: Stage-monitoring stations of basin S-65BC.

Station	*Coord-x* (ft)	*Coord-y* (ft)
AVONP4	587609.4	1174516.5
FTKISS	605942.2	1184399.5
KRBNS	600725.5	1136935.0
KRDRS	592258.3	1145920.0
PC11R	613428.7	1117442.6
PC33	611005.3	1126331.2
PC34	604078.7	1128480.4
PC44	599159.6	1135659.8
PC45	597855.5	1134887.6
PC52	595270.6	1147180.0
S65C_H	618817.6	1114966.3
S65C_T	618817.6	1114966.3
WEIR1_H	590606.7	1164605.9
WEIR1_T	590606.7	1164605.9
WEIR2_H	599232.6	1174550.4
WEIR2_T	599232.6	1174550.4
WEIR3_H	604259.1	1179268.4
WEIR3_T	604259.1	1179268.4

Table 3.3: Subbasin S-65BC's station names and coordinates (in Florida State Plane Coordinate System)

After the adoption of the genetic algorithm and still using IDW as the interpolation method, the next step in the creation of the optimization algorithm for stations in lakes was the consideration of a time period instead of a single day and an admissible root mean square error (*RMSE*). Such consideration involves, apart from the identification of the optimal subsets (i.e. selecting r from n) of each day of the period (*daily optimal subsets*), the statistical processing of these subsets to find the optimal subset of the period.

Two options for carrying out the identification of the daily optimal subsets were implemented. One consists of finding daily optimal subsets when varying r, the subset size. The possibility exists of finding optimal subsets, one for each considered size and date. This option is called *Optimization by Range of Sizes*. The other option is aimed at finding only one optimal subset for each date, the subset that has the minimum size and at the same time has a *RMSE* equal to or smaller than the admissible *RMSE* previously fixed according to policy rules developed by the SFWMD (*Optimization by Admissible RMSE*). See flow chart in Figure 3.6.

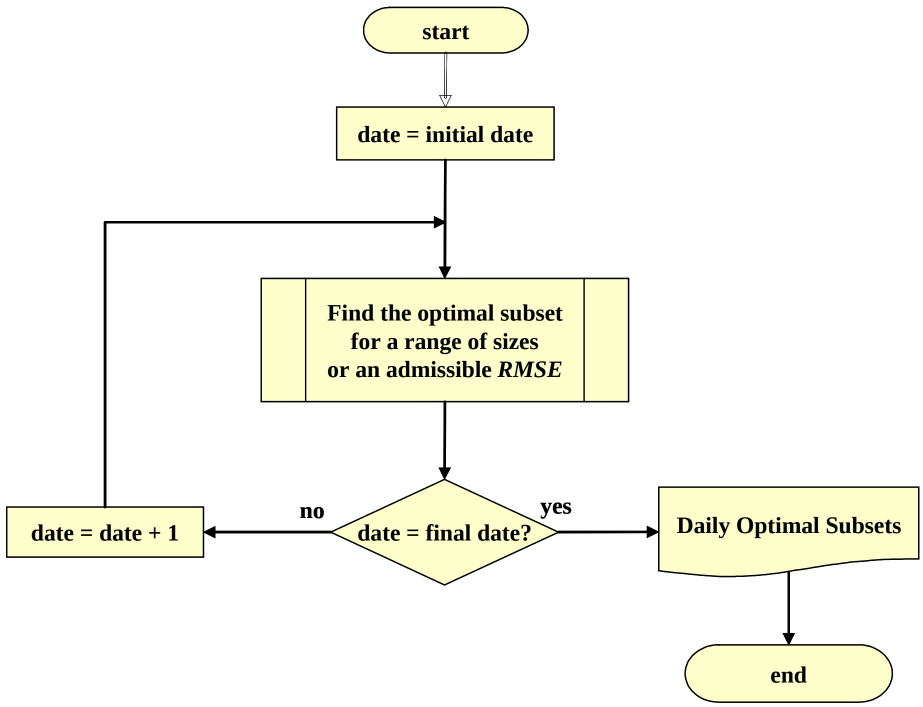


Figure 3.6: Flow chart of the identification of the daily optimal subsets of the period algorithm.

The optimization by range algorithm does not have any complications; it is implemented using a *for-next* loop, finding the optimal subset for each r varying from r_i ($r_i \geq 1$) to r_f ($r_f \leq n - 1$). On the other hand, the optimization by admissible *RMSE* was initially implemented using a method based on the well-known bisection method for finding roots of functions, with r_i and r_f as extreme values assuming that the *RMSE* value monotonically decreases when r increases, making it feasible to find the smallest subset that has an *RMSE* less than or equal to the admissible *RMSE*. Unfortunately, the *RMSE* value does not behave nicely. It is true that *RMSE* is large for small values of r and small for large values of r, but it does not always decrease from one value of r to the next. The optimal subsets of several sizes from a set of 13 stations in subbasin S-65BC illustrate this case (Figure 3.7).

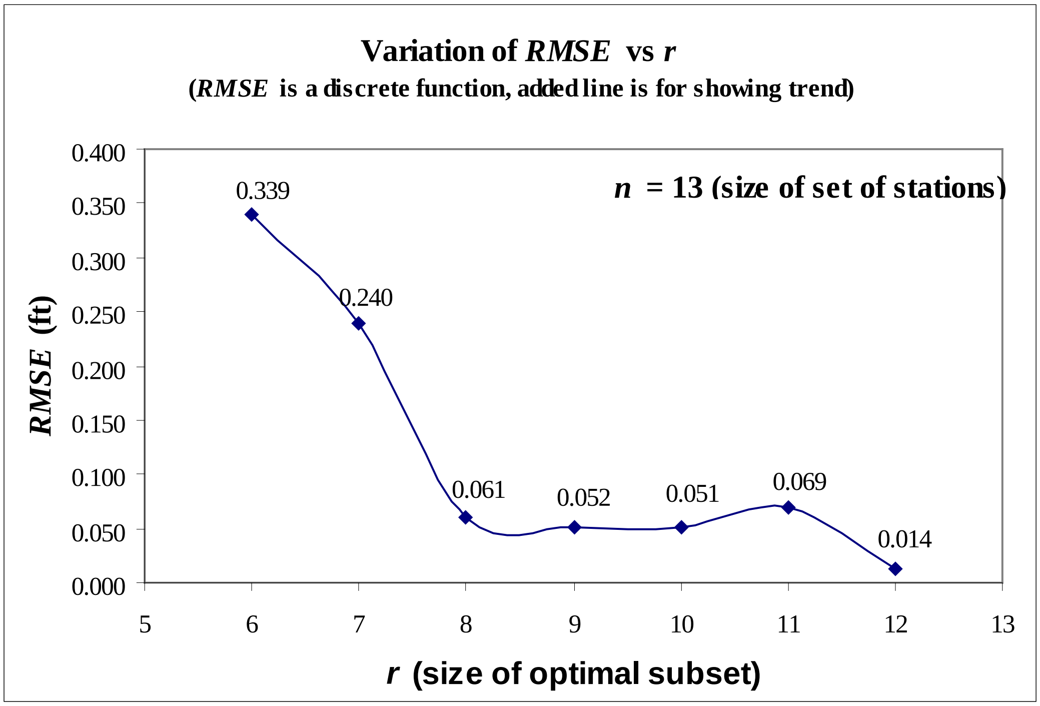


Figure 3.7: Variation of *RMSE* for several optimal subset sizes for the stations of subbasin S-65BC on October 1, 2002.

It was necessary to modify the code and adopt an algorithm similar to that used in the optimization by range of sizes option, but this time beginning from the largest *r* to the smallest *r* and sequentially finding one or more optimal subsets with *RMSE* ≤ admissible *RMSE*. An additional issue that was revealed with this option was that sometimes, due to the probabilistic nature of the optimization method, the true optimal subset was not found. To solve this problem and improve the method each time a subset is chosen from a set of stations, the genetic algorithm must run until an obtained minimum value of *RMSE* is repeated at least three times. This brute force approach proved to be effective in making the genetic algorithm more robust.

Once the daily optimal subsets of the period are obtained, a statistical analysis is performed, finding the most frequent daily optimal subsets and basic properties of their respective root mean square errors. The steps of this process are:

1. Sort the records of identified daily optimal subsets by ascending date and ascending *r*.
2. Exclude records with *RMSE* > Admissible *RMSE* (In the optimization by admissible *RMSE* this step was already performed).
3. Choose the record, from each day, that has the smallest subset of stations and at the same time has a *RMSE* ≤ Admissible *RMSE*.
4. For the identified daily optimal subsets, compute statistics such as frequency, average *RMSE* and standard deviation of *RMSE*.
5. Finally, chose the most frequent subset as the optimal subset of the period.

The case study 3 was used to test the first methodology to optimize its network as if it were a lake. The purpose of using several periods is to have several sets of optimal stations that can be compared. If the sets are essentially the same, that is, the sets contain the same stations; one can conclude that the optimal subset of stations is stable, at least for the periods involved. In such a case, it can be hypothesized that the validity of the optimal subset of stations may be extended to other periods. The results (Chapter 5) were inconclusive, such that no single subset or combination can be considered the optimal subset of the period. An additional criterion must be defined to find the optimal subset.

3.2.3 Development of the Second Methodology for Lakes

The variations in the optimal subset of stations from year to year shown in Chapter 5 are believed to be dependent on the number of operational stations and their corresponding levels. In addition, the presence of hydraulic structures along the streams breaks the continuity in the water profiles affecting the optimization process. These issues can be tackled by choosing periods when the number and working conditions of operational stations do not change; and choosing water control units between hydraulic structures that substantially modify the water levels. At this point of the research, the presence of hydraulic structures along the stream that control or modify the flow made clear that a different methodology should be applied to streams and that the case study 3 (being a river) was not appropriate for the continuation of the development of the methodology for lakes. Additionally, there is a need to develop a criterion to effectively identify the optimal subset of stations.

The case study 4, with data of the Lake Kissimmee stage-monitoring network was selected to test the second methodology (the first methodology complemented with the required additional criterion). Lake Kissimmee is located in the northern part of the Kissimmee River Basin Area of SFWMD (Figure 3.8).

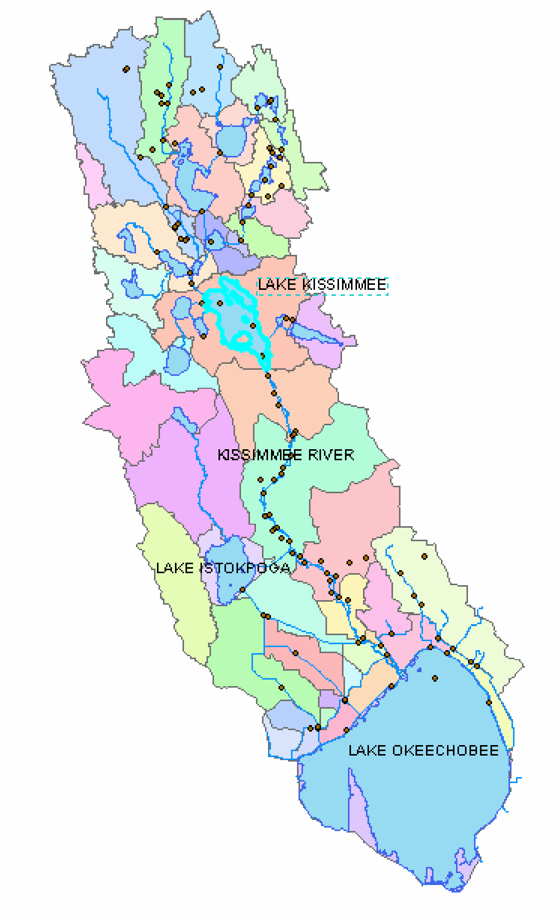


Figure 3.8: Kissimmee River Basin Area. Lake Kissimmee is highlighted.

Located inside Lake Kissimmee are five water stage-measuring stations: LKISS9, LKISS7, LKISS5B, S65_H, and S65NEW_H (Figure 3.9). The suffix _H in the name of a station, like S65_H or S65NEW_H, denotes that the station is measuring the water stage at the headwater side of a control structure. The suffix _T denotes the tail side of a control structure.

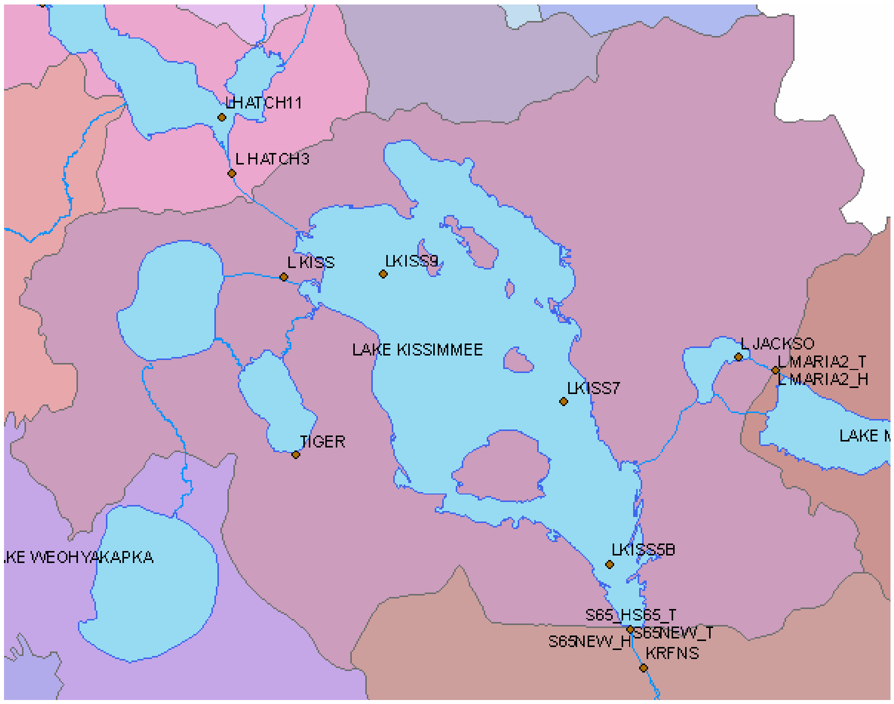


Figure 3.9: Lake Kissimmee. Stage-monitoring stations are shown.

A stable set of stations (a set in which the number of stations and their working conditions remain unchanged), was selected to analyze the data for a period of two years in order to draw meaningful conclusions. A stable set of four stations with daily stage measurements (LKISS9, LKISS7, LKISS5B, and S65_H) was selected for the period 10/01/2001 to 9/30/2003.

An optimal subset of a series of daily optimal subsets cannot be selected with the preliminary methodology because ordinarily there is not a subset that is present in all the days of the period or even in most of the days of the period. There is a need to focus on the behavior of all the daily optimal subsets during the entire period of analysis, not only in the days they are optimal. Therefore, the daily optimal subset that reports better behavior during the entire period would be considered the optimal subset of the period. Assume that the behavior of a subset can be measured by its average *RMSE*, standard deviation of *RMSE* and the number of times (days) the *RMSE* is less than or equal to the admissible *RMSE* previously chosen. The number of days the *RMSE* is less than or equal to the admissible *RMSE*, n_v, can be used to establish a relative frequency, F_r, by dividing n_v by the number of days the given subset is present, f_g. Mathematically:

$$F_r = \frac{n_v}{f_g} \tag{3.5}$$

Then, this empirical probability could be compared with the probability, $P(RMSE \leq AdmRMSE)$, reported by a suitable theoretical probability distribution to see if the relative frequency could be predicted with the theoretical distribution (meaning that the relative frequency and the theoretical probability are approximately equal). If the results for the entire period could be explained with the theoretical distribution, one can assume that the relative frequency or the theoretical probability could constitute a measure of the ranking of a given optimal subset. The greater the relative frequency (or theoretical probability) of a given subset, the higher the rank of the subset. One suitable theoretical probability distribution for the description of *RMSE* is the lognormal distribution because *RMSE* is a strictly positive variate and the lognormal distribution can describe this type of variates. See Ang and Tang (1975).

In summary, the steps of this additional statistical processing of the daily optimal subset results are:

- For each identified daily optimal subset, compute its *RMSE* for all possible days in the period.
- For each subset, count how many days the *RMSE* is ≤ admissible *RMSE.*
- Compute for each subset, statistics for the entire period, such as frequency, minimum *RMSE*, Maximum *RMSE*, average *RMSE* and standard deviation of *RMSE*.
- The relative frequency of a subset (number of days when the *RMSE* of the subset is less than or equal to the admissible *RMSE* divided by the frequency of the subset in the period) is a measure of its rank. The higher the relative frequency, the higher the rank.
- Test the relative frequency of each daily optimal subset comparing it with the P(*RMSE* ≤ admissible *RMSE*) given by a suitable probability distribution (in this case, the lognormal distribution). If the empirical and theoretical frequencies of the subsets are similar, results can be considered adequate.
- Finally, chose as the optimal subset of the period the daily optimal subset with the higher relative frequency.

The application of the second methodology for optimizing stage-monitoring networks in lakes to case study 4 produces the results shown in Chapter 5. Additionally, the complete second methodology is shown in Chapter 4.

3.3 OPTIMIZATION OF STAGE-MONITORING NETWORKS IN STREAMS (STEADY FLOW)

The development of the methodology for optimizing the stage-monitoring network in streams in steady flow considered two steps, one to develop the first methodology and another to develop the second methodology.

3.3.1 Development of First Methodology for streams

The case study 5, a case study located in the Kissimmee River Basin was chosen to develop the first methodology to optimize the stage-measuring network along a stream (Figure 3.10). This case is the Kissimmee River – C-38 Canal, the main stream, of the Pool AE (Subbasins S-65A, S-65BC, S-65D and S-65E), from station S65_T to station S65E_H. This stream has three reaches. The first reach and the third reach are channelized, while the central reach is naturalized. The first reach has a length of 22 miles; this reach ends 1.5 miles downstream of structure WEIR_1. The second reach is 11 miles long, ending 1 mile downstream of station PC33. The final reach has a length of 18.5 miles. There are control structures at the end of each subbasin and three additional weirs in subbasin S-65BC (or Pool BC). The structures control the movement of water in the stream. The most complex movement of water occurs in the naturalized reach, which is located completely in the downstream portion of subbasin S-65BC. Along this reach, there are significant changes of water elevation (Figure 3.11). The period of daily stage measurement data that was chosen is from 10/01/2001 to 12/31/2003. Table 3.4 shows the 20 stations with water elevation data and their distance from the downstream end of C-38 at Lake Okeechobee.

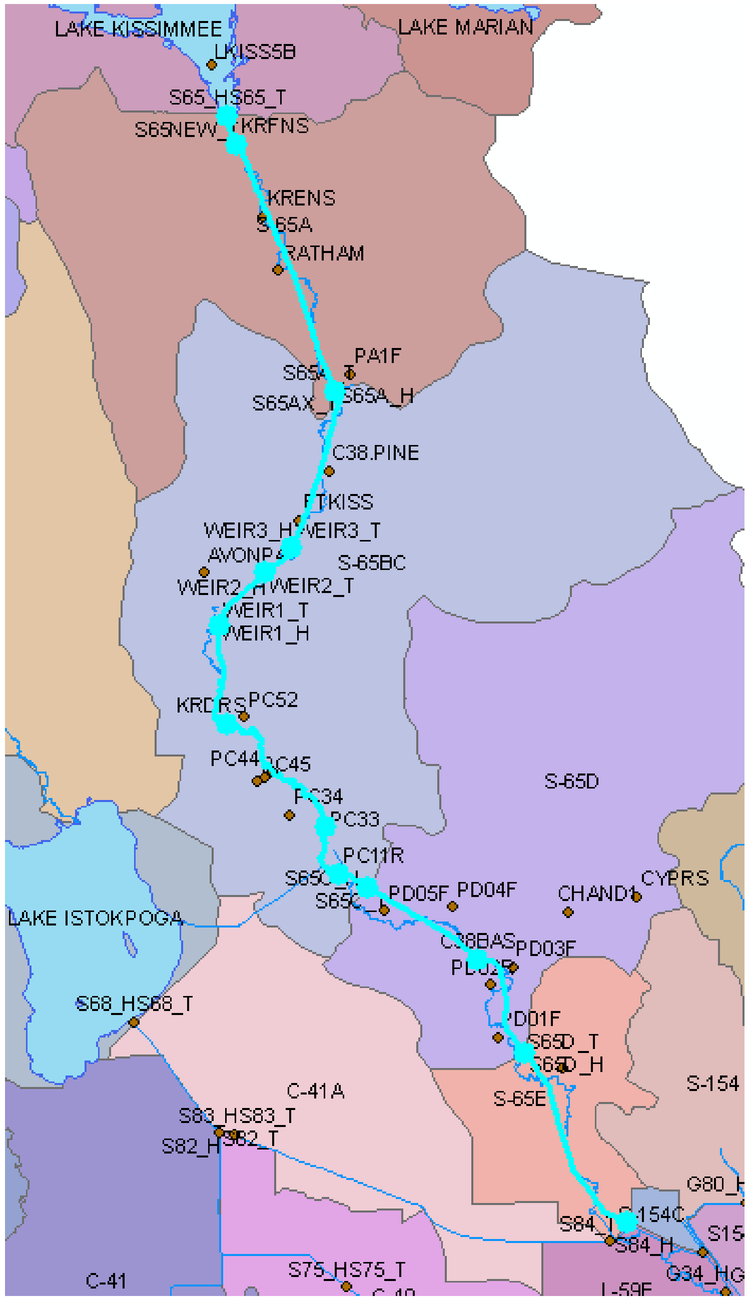


Figure 3.10: Stage-monitoring stations along the main stream of Pool AE.

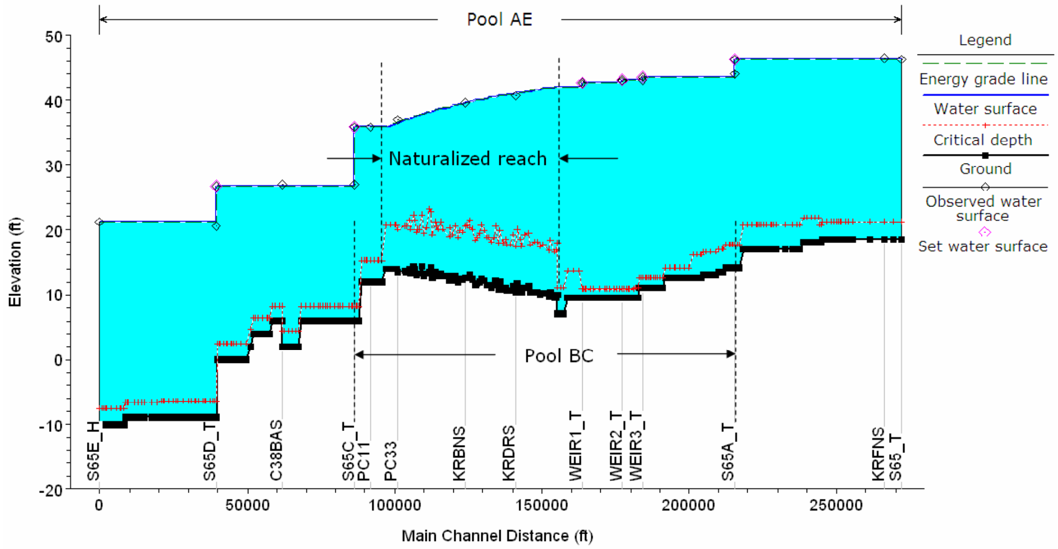


Figure 3.11: Typical profile of main stream of Pool AE.

	Station	Section (Distance, ft)		Station	Section (Distance, ft)
1	S65_T	320105.8	11	KRDRS	188833.2
2	KRFNS	314076.8	12	KRBNS	171777.7
3	S65A_H	263639.2	13	PC33	149107.5
4	S65A_T	263634.2	14	PC11R	139678.6
5	WEIR3_H	232169.0	15	S65C_H	134404.5
6	WEIR3_T	232164.0	16	S65C_T	134399.5
7	WEIR2_H	225189.9	17	C38BAS	110014.5
8	WEIR2_T	225184.9	18	S65D_H	87366.08
9	WEIR1_H	211725.5	19	S65D_T	87361.08
10	WEIR1_T	211720.5	20	S65E_H	47964.26

Table 3.4: Stage-monitoring stations and their distance along the stream. The lowest distance corresponds to the downstream end of the stream.

As the case of stage-monitoring stations of a lake, the solution of the problem of developing a methodology to obtain the optimal subset of stage-monitoring stations along a stream needs measures of error. Two plausible measures of error are the root mean square error of interpolated values of water surface elevation along the stream for a given time (*RMSE*) and the root mean square error of interpolated values of water surface elevation at the same station along a given period of time (*RMSESt*). The definition of *RMSE* was given by equation 3.1, and it is equally applicable to the case of stations along streams. *RMSESt* can be defined by:

$$RMSESt = \sqrt{\frac{1}{n_t}\sum_{j=1}^{n_t}\left(I_j - M_j\right)^2} \tag{3.6}$$

where I_j is the interpolated value and M_j is the measured value, both for the given station for time j., and n_t is the number of dates (or, in general, intervals of time) in which interpolated values are computed (see Figure 3.12). The figure shows a graph that illustrates the case of the optimization of the stage-measuring network in streams. Similarly to Figure 3.3, the stations $s \in S$ have stage measurements for the times $t \in T$ with several missing measurements in some of the stations, resulting in $n_t < T$. One procedure to solve this problem can be based on *RMSESt*. That is, having selected a base subset of r stations, compute the *RMSEst* of all stations in the subset of $n - r$ stations; then select the station that has the maximum value of *RMSESt* to define the base subset of $r + 1$ stations. Repeat until a termination criteria is met.

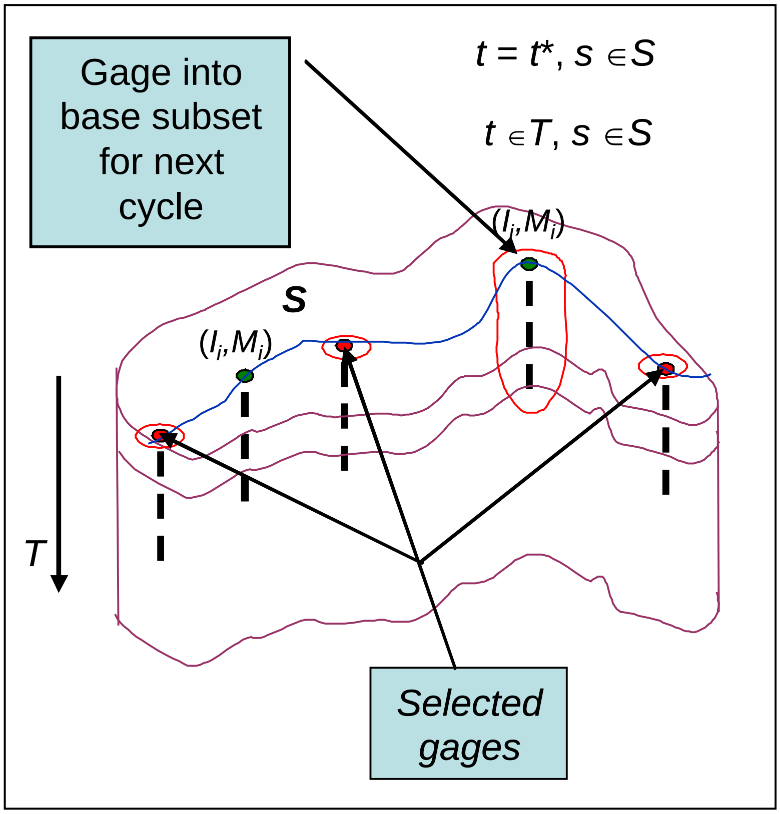


Figure 3.12: Graphic description of the optimization problem.

As in the previous problem, an interpolation method and an optimization method are needed. In principle, this problem can be characterized as a problem of combinatorial optimization, analogous to the lake's problem, but with hydraulic constraints. These constraints make each computation of *RMSE* significantly longer, as high as three orders of magnitude, than in the case of lakes, rendering the application of a genetic algorithm quite onerous. However, the experience of the case of lakes shows that studying the error of subsets of stations (or a single station) along time could lead to optimal or almost optimal solutions.

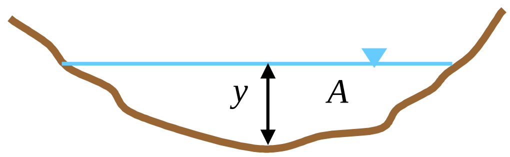
3.3.1.1 One-Dimensional Steady Flow Hydraulic Modeling

The interpolation of water elevations in the case of stations along the streams must be based on a model of the movement of water. The movement of water through a hydraulic system, like a river reach or a network of channels, can be steady or unsteady. It is unsteady if the discharge, Q and the water elevation, h, in any point of the system depend on time; and conversely, it is steady if discharge and water elevation do not depend on time. In actual river reaches or networks of channels, discharges and water elevations are usually changing with respect to time, that is, flows are most of the time unsteady. However, in certain circumstances, when temporal variations of Q and h are small, one can assume that the flow is steady, resulting in analyses that still produce good results. In addition, if one assumes that only the average properties of each cross-section are sufficient to represent the movement of water, neglecting the vertical and horizontal variations of velocity and pressure, one is using a one-dimensional model of the flow. In many situations, one-dimensional modeling is adequate to represent actual flows. Especially in conduits with large lengths compared with the dimensions of their cross-sections, and where differences of velocity among the points of each one of their cross-sections are not acute, the assumption of one-dimensional flow is justified.

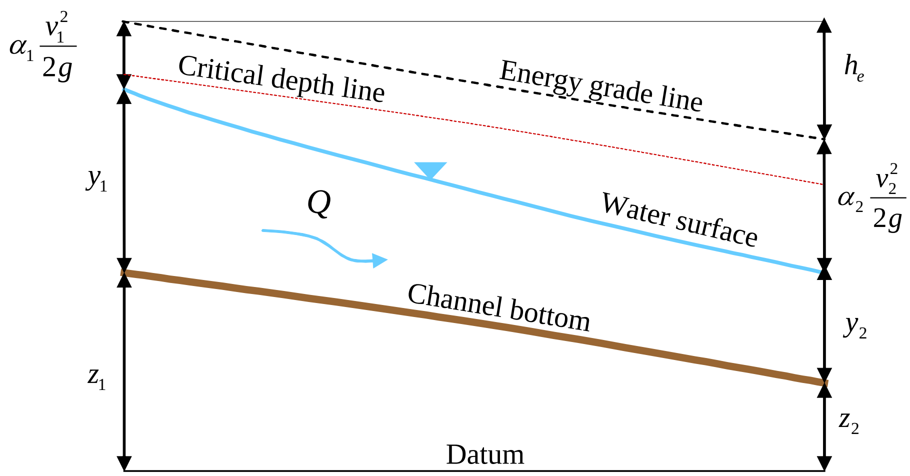
Modeling of gradually varied steady flow is based on the solution of the energy equation. See for example, (Chow 1959). From Figure 3.13, the energy equation can be written:

$$y_1 + z_1 + \alpha_1 \frac{v_1^2}{2g} = y_2 + z_2 + \alpha_2 \frac{v_2^2}{2g} + h_e \tag{3.7}$$

where $y_{1,}$ y_2 are the depths of water at cross sections; $z_{1,}$ z_2 the elevations of the main channel inverts; $v_{1,}$ v_2 the average velocities, computed with total discharge/total flow area; $\alpha_{1,}$ α_2 velocity weighting coefficients; g gravity acceleration, and h_e energy head loss.



Cross Section



Profile

Figure 3.13: Energy equation applied between two sections of a channel.

The energy head loss between two cross sections is composed of friction losses, and contraction or expansion losses:

$$h_e = \overline{s_f} L + k \left| \alpha_2 \frac{v_2^2}{2g} - \alpha_1 \frac{v_1^2}{2g} \right| \tag{3.8}$$

where $\overline{s_f}$ is the representative friction slope between two sections which can be obtained with a friction equation such as Manning's or Chezy's; *L*, the reach length, and *k*, an expansion or contraction loss coefficient. Solving the Energy equation from one cross section (1) to the next section (2), water surface profiles are computed. If the flow is subcritical, computation is from downstream to upstream, while if the flow is supercritical, computation is from upstream to downstream, as in Figure 3.13. Details of the assumptions and methods of computation vary from model to model.

When there is a change from subcritical to supercritical depth or from supercritical to subcritical depth, the energy equation is not considered applicable because it is a situation of a rapidly varying flow. Rapidly varied flow occurs when there is an appreciable change in channel slope, at control structures or in stream junctions. For some of the cases, empirical equations can be used while in others the application of the momentum equation is necessary to obtain the depth. The momentum equation is derived from Newton's second law of motion (Chow 1959), by applying the law to a volume of water enclosed by two sections, such as the volume in Figure 3.14.

3.3.1.2 First Methodology

Three well-proven hydrodynamic models were considered as interpol ation methods in the case of stations in streams for both steady and unsteady flow. The two main relevant characteristics for choosing a model were cost and operating system requirements. The three models include:

- **HEC-RAS.** HEC-RAS can be freely downloaded from the HEC Web site. It is a windows application that can be coupled more or less easily to other windows applications. This software has widespread use in the water resources community.
- **MIKE 11.** Mike 11 is commercial software, relatively expensive, published by DHI Water & Environment. It is also a windows application that can be coupled to other windows applications.
- **FLDWAV.** The National Technical Information Service (NTIS 2005) makes FLDWAV available for a relatively small amount of money . It is not a windows application.

HEC-RAS was chosen for this research because it can be coupled with other windows applications such as ArcMap, and it can be obtained without cost.

Several initial ideas about the optimization of stage -monitoring networks along streams are written as steps to follow in a preliminary study. They are:

- **Step 1.** Model the stream taking into account relevant external and internal boundary conditions.
- **Step 2.** Load geometric data of the stream, discharge and stage data in the appropriate files.

- **Step 3.** Calibrate the model using a handful of representative datasets.
- **Step 4.** Once a good estimation of water elevations at measuring stations that are not boundary conditions is obtained, attempt to calibrate the model for the entire period of analysis adjusting Manning's *n*, as many times as necessary. Once the model has been calibrated, the preliminary optimization process can be initiated.

  Repeat the next three steps several times:
- **Step 5.** With the base subset of stations, compute the water profiles for the days of the given period. The first base subset of stations, is the subset of stations that are required to hydraulically model the considered stream; that is, the stations that set a water elevation from which the water profile is computed in the hydraulic model. In the case of subcritical steady flow, there must be one boundary condition at the downstream end of the stream and one internal boundary condition at the head of each hydraulic structure inside the stream.
- **Step 6.** Compute the root mean errors (*RMSESt*) of all the stations outside the base subset.
- **Step 7.** According to a rule suggested by actual results, choose one station outside the base subset. Add it to the base subset. Go to step 5.
- **Step 8.** Analyze results. See if the algorithm can be superseded. What improvements can be made?

Using this procedure, a preliminary optimization of the main stream of the Pool AE network was obtained. This initial study followed the next steps: creation of geometry file, calibration with a limited number of datasets, calibration with data of the entire period and estimation of water elevations in selected stations.

Creation of geometry file. The creation of the geometry file is needed because HEC-RAS uses this file in the computation of profiles. The geometry of the two C-38 canal reaches was obtained from SFWMD's AHED geodatabase. For the naturalized reach in the Pool BC area, which does not have any cross-section data, the River Channel Morphology Model (RCMM) developed by Merwade and Maidment (2004) was used to create an analytical river channel. The RCMM model uses the curvature of the centerline of the river to create analytical cross-sections in the form of beta probability distribution functions. The AHED geodatabase was used to extract the centerline of the Kissimmee River for input to RCMM. Besides the centerline features, RCMM requires information on average width and depth of the channel to create cross-sections. The width and depth of the canals upstream and downstream of the Kissimmee River were used to rescale the cross-sections in the RCMM model. After the cross-sections for both canals and Kissimmee River were created, the functions in RCMM were used to create the geometry file for HEC-RAS.

Calibration with a limited number of datasets. A manual iterative process of calibration was needed to have good estimates of Manning's *n* and roughness factors for different values of discharge along reaches. Calibration consists in changing Manning's *n* and/or roughness factors to obtain computed profiles similar to observed profiles for given datasets. Additionally, to improve the similarity of the profiles the discharges at structures were moved upstream of the structures. Eight datasets of flows and water stages were used in the initial calibration process. One dataset was generated with the average conditions for the entire period. The resulting profile for the average conditions is shown in Figure 3.15.

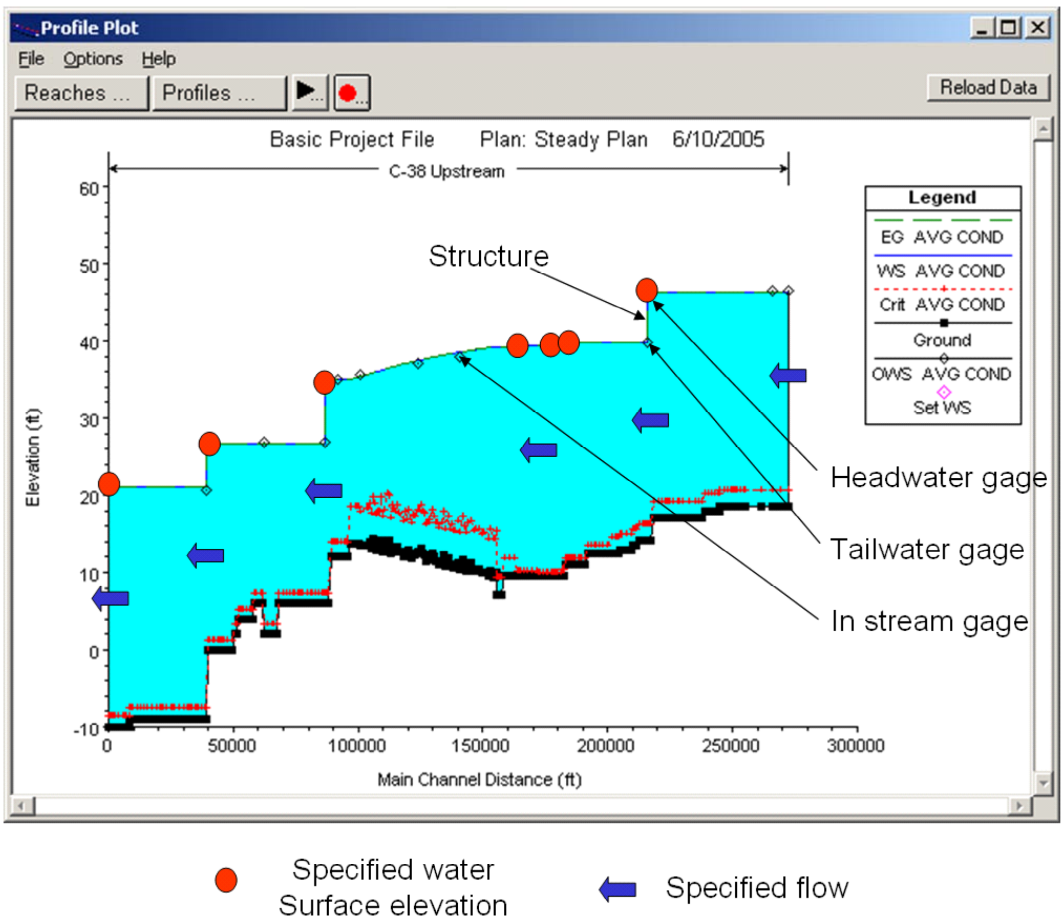


Figure 3.15: Profile for average conditions (annotations are added).

The legend in each HEC-RAS generated profile refers to the symbols that represent several elevations. A green dashed line represents the energy grade line; blue line, water surface elevation; red dotted line with crosses, critical-depth water elevation; black line with solid squares, ground elevation; black line with hollow rhomboids, observed water surface. Finally, tilted hollow squares with a dot in their center are set water elevations. In the profile, they are behind the superimposed large red circles. The origin of the main channel distance is station S65E_T, whose section or longitudinal coordinate is 47964.26 ft. Seven more datasets were chosen to reflect the variation of discharge in the different reaches. The specific dates of these seven datasets are 10/01/2001, 12/28/2003, 9/03/2003, 6/15/2002, 1/20/2002, 1/04/2003, and 9/02/2002.

Calibration with data of the entire period. Once this initial calibration was done, the entire period was used to check the variation of roughness of the three reaches of the stream. From all the datasets of 10/01/2001 to 12/31/2003, only the datasets that have complete data for all the stations at the head side of a control structure or a weir, and no negative flows in the hydraulic structures, were chosen. The reason to choose these datasets is that the water elevations at these stations are used to define internal changes of water elevation, and that HEC-RAS converts negative flows to positive 0.1 values. These two constraints left 428 out of 822 possible daily datasets.

Figure 3.16 shows the discharge in S65C, the structure downstream of the naturalized reach of Kissimmee River, and the water elevation in KRDRS, the extreme upstream stage-monitoring station of the same reach, for the days of the period that can be used to perform the optimization. These stations give a fair idea of the variation of both variables in the entire Pool AE reach.

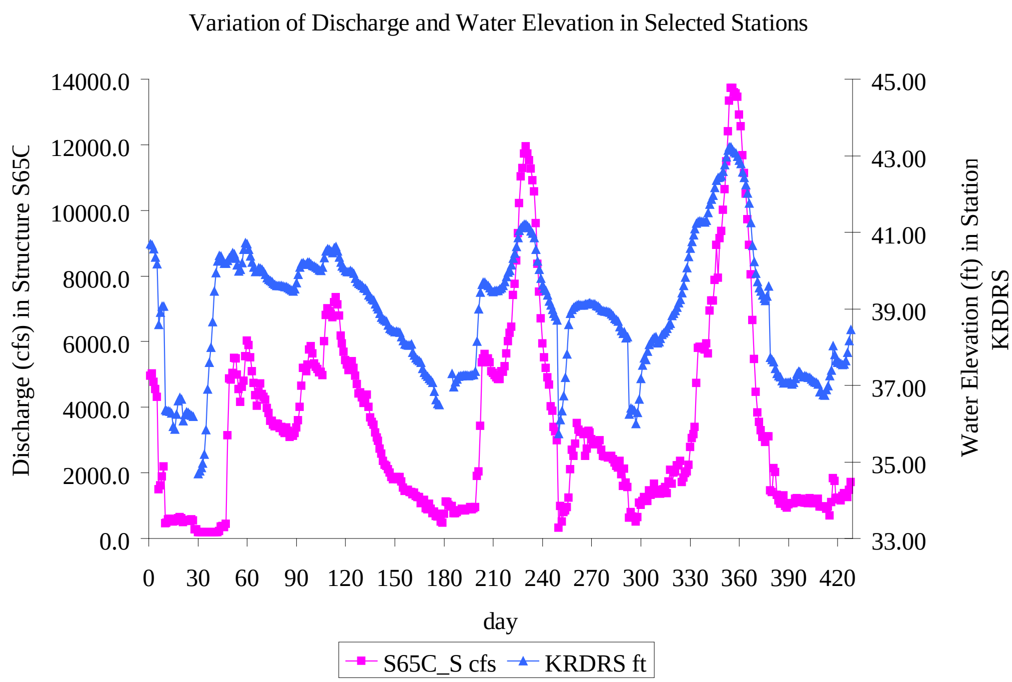


Figure 3.16: Typical variation of discharge and water elevation in the main stream of Pool AE.

The computation of profiles in an automatic way was accomplished with a tool implemented in ArcMap. When the tool is invoked, it prompts for a HEC-RAS project file, previously prepared. This project file refers to the geometric file and the flow and water elevations input of the studied stream. Once the project file is accepted the tool reads from a table (RASTable.dbf) the flow and water elevation data that define the problem of the computation of profiles date by date of the period supplied, sequentially computing the profile for each day. Results are stored in another table (CrossSections.dbf). From the second table the root mean square error for each date is computed using another tool. In addition, another tool computes the root mean square error of each station for the entire period.

After each run of the HEC-RAS automation tool, the errors in the stations are computed. The dates with the biggest errors are identified and used to change the roughness factors to recalibrate the computed profiles. Six runs of calibration were needed before having a negligible reduction of errors; that marked the end of the calibration process. The initial datasets for calibration were complemented with the datasets of 6/27/2002, 7/29/2002, 1/19/2003, 8/10/2003, 8/07/2003, 1/09/2003, 8/26/2003, 1/20/2003, 1/13/2003, making 17 datasets for the calibration of the entire period.

Estimation of water elevations in selected stations. The calibration process is followed by the estimation of water elevations for the selected stations in three different subsets of stations. The results of the last calibration run are valid for the estimation of water elevations for the case of the subset constituted by the seven stations at the head side of the hydraulic structures of Pool AE. To estimate how the errors evolve when more stations are present in the subset of base stations, two more stations were chosen. The first station added into the base subset made a base subset of eight stations. The errors for this subset were computed. The second station made a base subset of nine stations and then the errors were computed again. The stations chosen had simultaneously the maximum *RMSESt* of the previous run and had the same days of data as the first subset of stations. Additionally, to understand the variation of the root square mean error, *RMSE*, plots of probability of no exceedence were prepared for each given base subset. While the empirical probability distribution of *RMSE* offers a way to study the lumped behavior of all the stations outside the base subset, the distribution of errors for a single station could be more useful to judge the particular behavior of that station. The error is defined as the difference between computed and observed water elevation. For the case of nine base stations, the distribution of errors was plotted. It was expected that the errors would tend to decrease when the number of stations in the base subset increases. The process was stopped because the results of the three base subsets of stations made possible the inference of useful conclusions. The results of the three subsets are shown in Chapter 6. Insights gained with the application of the first optimization methodology for stations along streams using data from case study 5 lead to the refinements described in the next section.

3.3.2 Development of the Second Methodology for Streams

The fifth case study, used before in the development of the first methodology for streams, was selected to develop and test the definitive methodology. The optimization results of the Pool AE main stream obtained with the first methodology were encouraging. In each cycle, the station with the largest *RMSESt* and water elevation data for all the days of the period was chosen to enter into the base subset of stations for the next cycle. *RMSESt* represents the deviations among the measured and computed water elevations at a particular station. Consequently, the largest *RMSESt* identifies the station where the interpolation method gives the poorest estimations of water elevation. If that station is set inside the base subset for the next optimization cycle, it is expected that the *RMSESt* values in the remaining stations outside the base subsets will tend to diminish or at least remain the same. This was the result found in the previous study.

Taking into account the preceding discussion, an additional improvement is to consider all stations outside the base subset, even those that do not have complete data for the entire period but still have data for a good number of days. This is to support the results of the optimization process with the largest possible amount of data. With respect to the number of sizes of base subsets, the number of subsets can be from the smallest one that guarantees the correct hydraulic modeling of the stream to the one that leads the computation of the *RMSESt* in only one station. Furthermore, another condition for finishing the optimization process could consider an admissible *RMSESt* value. Steps 7 and 8 of the algorithm of previous section can now be improved:

- **Step 7.** Choose the station that has the largest root mean square error. If the number of stations in the base subset is smaller than the number of stations along the stream minus one, mathematically $r < n - 1$, add the new station to the base subset and go to step five; otherwise finish the optimizing process. This process of including the station that has the largest root mean square error for the next cycle can be called *forward selection*. The process may also be halted when, for example, a given admissible error has been reached.

- **Step 8.** Analyze results. Answer the question: which base subsets can be considered the optimal subset?

The algorithm with the improvements just described, that is the second methodology, was used to re-optimize the stage-monitoring network of the main stream of Pool AE (case study 5). Many of the steps made in the initial optimization study of this stream are already explained. The remaining steps are the optimization process, and the interpretation of results and finding of the optimal base subset.

The data was the same used in the initial optimization study but instead of optimizing only subsets of 7 to 9 stations with complete data, the range of sizes of the base subset considered were from 7 to 19 stations, considering all stations with complete or incomplete data. All the stations outside the base subset could be considered to go inside it because the optimization tool (Network optimization in streams: NetOptStreams) can now identify the dates when all base stations have data (valid dates) and proceed using only data for those dates. The initial base subset is shown in Table 3.5. Once the optimization tool found the base subsets for sizes 8 to 19, an analysis of the results was conducted and conclusions were drawn. Results are shown in Chapter 6.

Station	Section
S65A_H	263639.2
WEIR3_H	232169
WEIR2_H	225189.9
WEIR1_H	211725.5
S65C_H	134404.5
S65D_H	87366.08
S65E_H	47964.26

Table 3.5: Initial base subset (size, *r* = 7).

3.4 Optimization of Stage-Monitoring Networks in Streams (Unsteady Flow)

The development of the methodology for optimizing the stage-monitoring network in streams with unsteady flow considered two steps, one devoted to the calibration process and another to develop the methodology.

3.4.1 Calibration Study

A new case study, case study 6, that comprises the main stream of Pool AE and the first two days of stage measured data considered in the development of the methodology for optimizing the network optimization in streams in steady flow was selected to carry out this study. Figure 3.17 shows the surface water profile in the studied stream on October 1, 2001. Table 3.6 shows discharge and stage data for the two days considered.

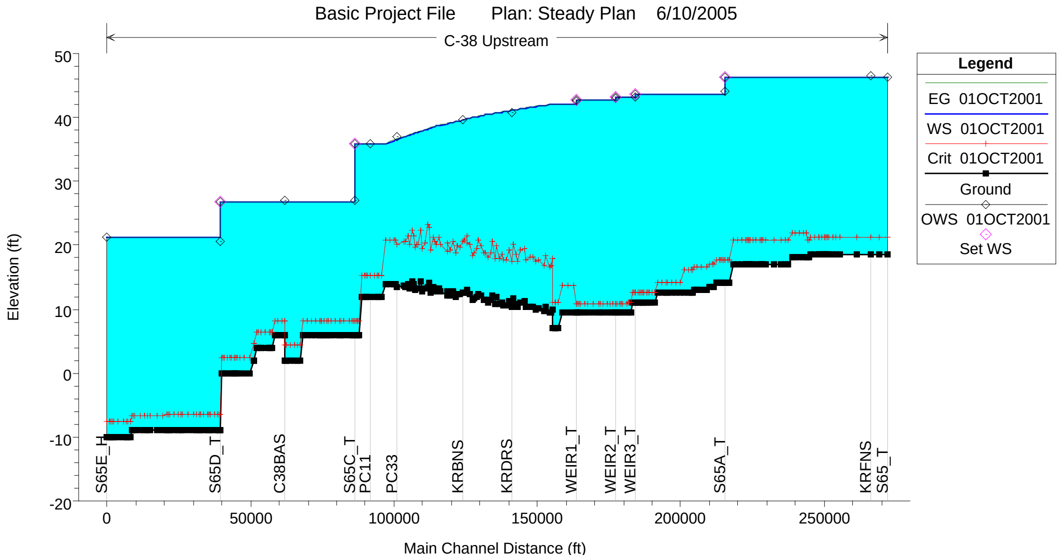


Figure 3.17: Water elevation profile along the main stream of Pool AE on October 1, 2001.

Station	Section (distance, ft)	1-Oct-01	2-Oct-01
		Discharge (cfs)	
S65_S	320106.8	2317.03	2098.77
S65A_S	263639.2	3957.35	3525.39
WEIR3_W	232169	1190.28	1188.33
WEIR2_W	225189.9	739.34	606.04
WEIR1_W	211725.5	927.91	884.91
S65C_S	134404.5	4956.728	5019.228
S65D_S	87366.08	5440.938	5441.391
S65E	47964.26	4661.55	4703.3
		Stage (ft)	
S65_T	320105.8	46.32	46.21
KRFNS	314076.8	46.36	46.25
S65A_H	263639.2	46.163	46.073
S65A_T	263634.2	43.972	43.772
WEIR3_H	232169	43.65	43.55
WEIR3_T	232164	43.1	42.99
WEIR2_H	225189.9	43.15	43.04
WEIR2_T	225184.9	42.97	42.92
WEIR1_H	211725.5	42.78	42.68
WEIR1_T	211720.5	42.39	42.33
KRDRS	188833.2	40.7	40.67
KRBNS	171777.7	39.62	39.61
PC33	149107.5	36.97	36.88
PC11R	139678.6	35.92	35.78
S65C_H	134404.5	35.804	35.639
S65C_T	134399.5	26.90	26.87
C38BAS	110014.5	26.98	26.94
S65D_H	87366.08	26.69	26.653
S65D_T	87361.08	20.62	20.661
S65E_H	47964.26	21.11	21.14

Table 3.6: Discharge and stage data for calibration study.

3.4.1.1 One-Dimensional Unsteady Flow Hydraulic Modeling

Modeling of unsteady flow is more complex compared to that of steady flow, making flow calibration difficult. Modeling of unsteady flow is governed by the principles of conservation of mass and momentum. These laws are expressed in mathematical form using two partial differential equations, named after Barre de Saint-Venant, who first developed them in 1871 (Chow et al 1988). The equations' derivation considers a differential control volume defined in a channel reach (Figure 3.18). Discharge and water elevation are functions of distance along the channel and time. A lateral flow, q, may be entering the control volume.

Whereas, the principle of conservation of momentum for a control volume can be written: *the net momentum entering the volume plus the sum of all external forces acting on the volume must be equal to the rate of accumulation of momentum inside the volume*. Its corresponding equation, neglecting the contribution of momentum of lateral flow is

$$\frac{\partial Q}{dt} + \frac{\partial (Qv)}{\partial x} + gA\left(\frac{\partial h}{dx} + s_f\right) = 0 \tag{3.11}$$

Equations 3.10 and 3.11 constitute a simplified version of the one-dimensional unsteady open-channel flow partial differential equations. Variations of these equations can be found in Chow (1959), Chow et al (1988) or Lai et al (2002). The Saint-Venant equations are hyperbolic partial differential equations and they are traditionally solved by the method of characteristics, the method of finite differences and the method of finite elements (Kutija 2002). However, the most successful and accepted procedure for solving equations 3.9 and 3.10 is the four-point implicit scheme (USACE 2004) shown in Figure 3.19.

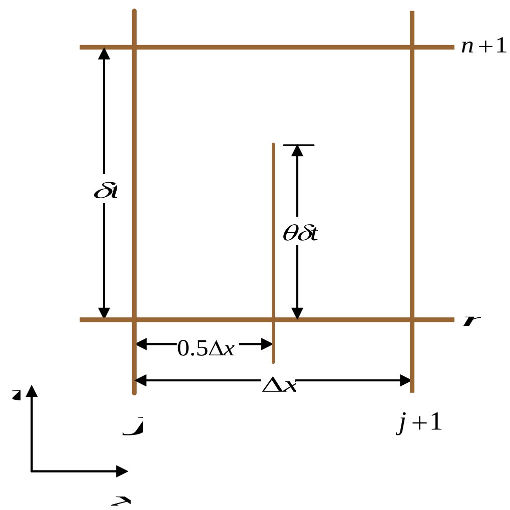


Figure 3.19: Cell illustrating the Four-point method's θ.

Using this scheme, space derivatives and function values are evaluated at an interior point, $(n + \theta)\ \Delta t$. Therefore, values at $(n + 1)\ \Delta t$ enter into all terms in the equations. Applying the scheme for a river reach a system of simultaneous equations is established, generating an implicit scheme. Contrary to an *explicit* scheme in which each unknown value is being computed sequentially along a time line from one spatial point to the next, in an *implicit* scheme like the four-point scheme, the unknown values on a given time line are all determined simultaneously (Chow et al 1988). For a given time step, solutions obtained with an implicit scheme are more stable than those obtained with an explicit scheme. In consequence, significantly larger time steps can be used in an implicit scheme than those used in an explicit numerical scheme. Analyses of stability made by Fread and by Ligget and Cunge in the 1970s showed that the implicit scheme is, in theory, unconditionally stable for $0.5 < \theta \leq 1.0$, but conditionally stable if $\theta = 0.5$ and unstable for $\theta < 0.5$ (USACE 2004). Additionally, in convergence analyses performed by the same authors, it was shown that numerical damping increases as the ratio between the length of a wave in the hydraulic system and the length of a reach decreases. For stream flow problems where the wavelengths are long with respect to spatial distances, convergence is not a serious problem. However, in practice other factors may also contribute to the non-stability of the solution scheme. These factors include dramatic changes in channel cross-section properties, abrupt changes in channel slope, characteristics of the flow wave itself, and the existence of complex hydraulic structures. In fact, these other factors often overwhelm any stability considerations associated with θ (USACE 2004). The Hydrologic Engineering Center (USACE 2004) recommends that any model application should be accompanied by a sensitivity study, where the accuracy and the stability of the solution are tested with various time and distance intervals. Other methods to solve the one-dimensional equations are proposed by Kutija (1996), who proposed adaptive schemes of computation, or Jin and Fread (1997) who developed an explicit schema for flows having strong shocks such as those created by near instantaneous dam-breaks or mixed-flow regimes where the four-point method is likely to fail. Furthermore, Sobey (2001) proposed a series of benchmark test problems for the evaluation of the performance of numerical models of flow and tide propagation in channels, recommending also routine reviews of any numerical model study.

3.4.1.2 Procedures of the Calibration Study

This calibration study benefits from the experience gained in the solution of the optimization of stage-monitoring network along streams in steady flow.

The calibration process considers successive improvements of the HEC-RAS model of the stream. Apart from the roughness values, other factors such as the computational interval value or the time to reach steady flow were investigated. Results of previous models influenced features of the next model. The procedure followed is:

First model. Using the calibrated model developed for the steady case, a first simulation of unsteady flow was made. Any period of simulation of one or several days (between 10/01/2001 and 12/31/2003, the period used in the steady flow case) was thought to be useful to begin the development of the network optimization in streams with unsteady flow. The only condition was that the data must be complete in all stations to be able of setting possible boundary conditions in all of them. The first days of the period have complete data and to reduce the time spent in computing water profiles, the two first days of data were selected (Table 3.5). It was assumed that the data for Oct 1, 2001 corresponds to the last moment of Sep 30, 2001, codified in HEC-RAS 30SEP2001 2400 (or, alternatively, first moment of Oct 1, 2001) and the data for Oct 2, 2001 corresponds to the last moment of Oct 1, 2001 (01OCT2001 2400). That is, the 24 hr interval of unsteady flow from 30SEP2001 2400 to 01OCT2001 2400 was simulated. To be able to set water elevations at the upstream side of each internal control structure, it was necessary to define inline structures. The computation interval was set equal to 15 minutes.

Second Model. After running the first model, it was realized that the unsteady flow module of HEC-RAS needs its own table of flow roughness factors, not the table that the steady flow module uses. Flow roughness factors take into account that roughness varies with flow. Therefore, the model was complemented with the required table and a new simulation was run.

Third Model. A larger number of cross sections could improve the simulation's results. Applying a cross section interpolation tool, several new cross sections were created along the stream. No single reach with a length of 1,000 ft or more is left in the geometry of the stream. In addition, all cross sections were checked to make sure that the range of possible water elevations is included in the computation of tables of geometric and hydraulic properties that are used to speed up the unsteady flow computations.

Fourth Model. An adjustment of the roughness factors of the naturalized river was made.

Fifth Model. Model 4 was changed from using internal boundaries in the inline structures to use properties of these structures, namely, length of crest, elevation of crest and discharge coefficient. In all cases, it was considered that the flow is uncontrolled.

Sixth model. An interesting thing to consider is the time that the flow needs to become steady inside the stream when the stream is subjected to constant inflow pattern and water elevation at the downstream boundary.

Influence of the computation interval's size. A remaining issue to be addressed was the influence of the size of Δt, the computation interval of integration. Using the fifth model a series of runs was done for Δt = 12, 10, 6, and 5 min.

Results of this calibration study are shown in Chapter 7.

3.4.2 Development of Methodology

From the interpretation of the results of the preliminary calibration study, several issues must be considered to perform a definitive calibration of the HEC-RAS model of Pool AE main stream and the formulation of the optimization methodology for this case. These issues are:

- No significant improvement was achieved using additional interpolated cross-sections. To make the computation faster they were deleted for subsequent work.
- Computational interval time must be as small as possible to achieve meaningful results.
- In the steady flow model all input flows were lateral flows that were entering (or exiting) the stream in given cross-sections. A more realistic approach that can be modeled in the unsteady flow version is to change the input flows from lateral to lateral-uniform flows.
- To be able to compute errors it is necessary to set water elevation values in all the stations that are inside the base subset, even when the resulting profiles are not physically sound.
- In the HEC-RAS unsteady flow module, adding a new station into the base subset requires the creation of a *fictitious* inline structure. Such a structure divides in two, a sub-reach between two existing inline structures or one inline structure and an external boundary condition. In the head side of the new structure, a stage-downstream boundary condition must be

set to ensure the correct computation of errors in the next cycle of optimization.

- Instead of using only one day of simulation, it is better to use five or more days. This could make the optimizing results more stable and conclusions derived from the results more robust.
- In principle, the basic characteristics (i.e. the evaluation of *RMSESt* for each station and the selection of the station with the largest *RMSESt* for the next optimization cycle), of the optimization methodology used in the case of steady flow can be used also for unsteady flow. The differences are more in the modeling of unsteady flow inside HEC-RAS and in how these differences should be approached when developing the computational tools needed to solve this problem.

3.4.2.1 First Methodology

According to results of the calibration study and the statements made in the previous section, a new attempt of calibration of the HEC-RAS model for the main stream of Pool AE (case study 6) was performed. The same data of 30SEP2001 2400 to 01OCT2001 2400 were used –considering as before, a linear variation of stages and flows. Uniform input flows were considered. The computational time interval considered was equal to 2 minutes. Errors were computed at intervals of 1 hour. A new calibration was made. Then, the optimization using the first methodology was carried out. Results are presented in Chapter 7.

3.4.2.2 Second Methodology

The results of the optimization of the stations of the main stream of Pool AE using one-day data made apparent the need to apply the same procedure to the optimization of the same stations but with a longer simulation time. The consideration of a longer simulation time converted the first methodology in the second methodology. The seven-day data period from Oct 1, 2001, to Oct 8, 2001, was used to obtain results that could be more robust. This new time period defines study case 7. Table 3.7 complements Table 3.5 showing the data of the remaining days.

A new calibration study to adjust the Manning's roughness coefficient of the naturalized river for the seven days data was made. Once the calibrated values were found, the optimization of the network using the second methodology was carried out. Results of this study case 7, presented in Chapter 7, were used to draw conclusions of the possible utility of using unsteady flow in the optimization of the stage-monitoring network along streams.

Station	Section (distance, ft)	3-Oct-01	4-Oct-01	5-Oct-01	6-Oct-01	7-Oct-01	8-Oct-01
		Discharge (cfs)					
S65_S	320106.8	1772.85	1616.06	1589.44	1453.5	1386.59	1358.22
S65A_S	263639.2	2933.83	2762.31	2683.81	2244.85	1933.34	1975.03
WEIR3_W	232169	1056.52	937.5	888.5	752.33	545.34	538.51
WEIR2_W	225189.9	392.54	320.72	315.11	-40.64	-147.47	-97.43
WEIR1_W	211725.5	786.44	695.39	655.6	589.71	476.39	470.86
S65C_S	134404.5	4767	4550.444	4306.027	3866.163	3534.261	3434.488
S65D_S	87366.08	5196.472	4980.395	4595.523	4244.377	3940.622	3866.625
S65E	47964.26	4510.51	4333.06	3991.4	3734.53	3533.58	3460.27
		Stage (ft)					
S65_T	320105.8	46.19	46.31	46.23	46.1	46.15	46.26
KRFNS	314076.8	46.24	46.35	46.27	46.15	46.2	46.31
S65A_H	263639.2	46.111	46.241	46.164	46.069	46.142	46.257
S65A_T	263634.2	43.346	42.931	42.724	42.392	41.934	41.853
WEIR3_H	232169	43.18	42.77	42.56	42.24	41.78	41.69
WEIR3_T	232164	42.71	42.37	42.19	41.96	41.63	41.55
WEIR2_H	225189.9	42.76	42.43	42.24	42	41.66	41.57
WEIR2_T	225184.9	42.72	42.4	42.21	42	41.67	41.58
WEIR1_H	211725.5	42.43	42.12	41.93	41.71	41.4	41.32
WEIR1_T	211720.5	42.14	41.88	41.72	41.53	41.28	41.2
KRDRS	188833.2	40.56	40.35	40.17	40.02	39.85	39.77
KRBNS	171777.7	39.52	39.34	39.15	38.99	38.84	38.8
PC33	149107.5	36.85	36.76	36.64	36.53	36.46	36.5
PC11R	139678.6	35.76	35.76	35.73	35.68	35.71	35.83
S65C_H	134404.5	35.645	35.66	35.63	35.598	35.65	35.778
S65C_T	134399.5	26.84	26.80	26.78	26.89	26.76	26.88
C38BAS	110014.5	26.93	26.89	26.87	26.99	26.87	26.98
S65D_H	87366.08	26.635	26.598	26.585	26.706	26.59	26.696
S65D_T	87361.08	20.641	20.617	20.606	20.687	20.657	20.691
S65E_H	47964.26	21.13	21.1	21.1	21.17	21.15	21.19

Table 3.7: Additional discharge and stage data for the optimization study of the Pool AE main stream's network.

3.5 ADDITIONAL STUDY TO PROPOSE IMPROVED METHODOLOGIES FOR LAKES AND FOR STREAMS WITH STEADY STATE

A second look at the second proposed methodologies suggested the need to test the tabu search method in the case of both lakes and streams with steady flow. Additional study of the optimization in lakes was needed to verify if tabu search could find the optimal subset of the period in less computational time. In the case of streams with steady flow, additional study was needed to implement tabu search to find the best combinations of stations for each possible size of subset, and not only the subsets chosen in the forward selection method.

3.5.1 Procedure for Comparing Optimization Methods in the case of Lakes

A new, more complex, case study 8 has been chosen to compare the results of the second methodology, that uses a genetic algorithm, and the third methodology, that uses tabu search. This case study is Lake Okeechobee (Figure 3.20). Lake Okeechobee is a lake located in the South Florida Water Management District. It is the biggest lake in the district with a surface area on the order of 647 square miles. In contrast, the area of Lake Kissimmee is only 54 square miles. Fourteen stations that monitor the water elevation inside Lake Okeechobee are shown in Figure 3.20. The stations are CULV10A_H, CULV5A_H, L_OKEE, L_OKEE.M_G, L001, L005, L006, LZ40, S127_T, S129_T, S133_T, S135_T, S191_T and S3_T. The daily average stage in the stations of Lake Okeechobee from 10/01/2001 to 9/30/2003 is shown in Figure 3.21. Stage in the first year is smaller than in the second year; this difference may influence the result of the optimization, producing a different optimal subset for each different year. In this research, it is assumed that a year with low levels and a year with high levels are sufficient to determine the optimal subset of stations of all possible sizes.

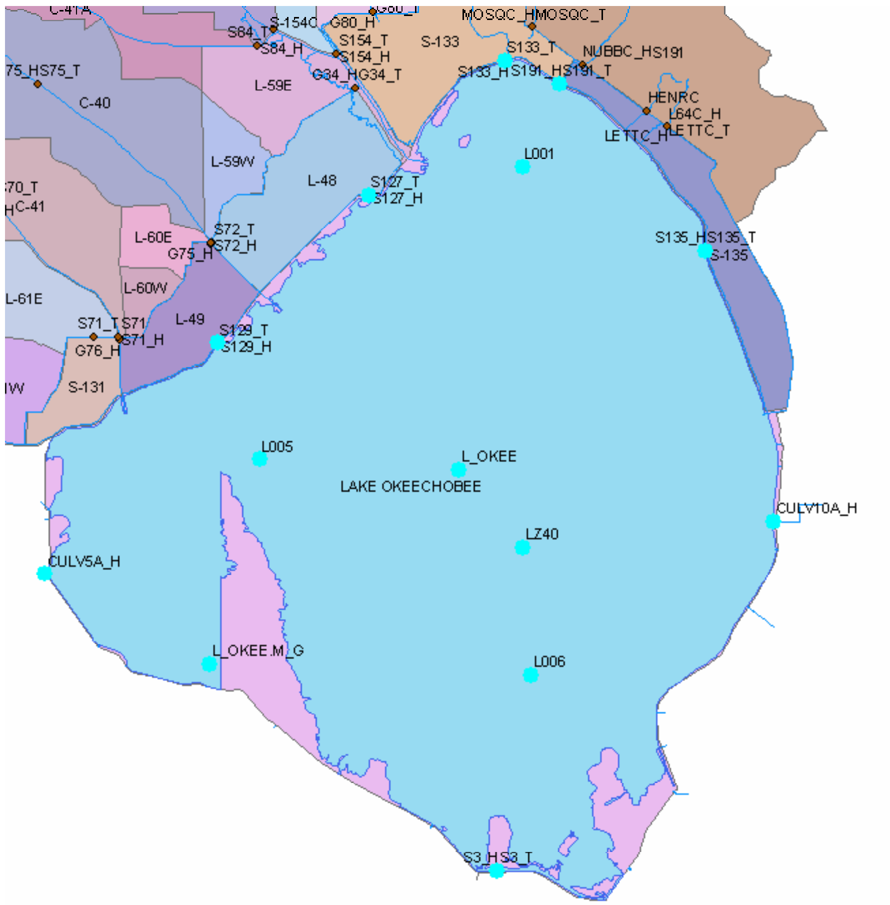


Figure 3.20: Lake Okeechobee's stage-monitoring stations.

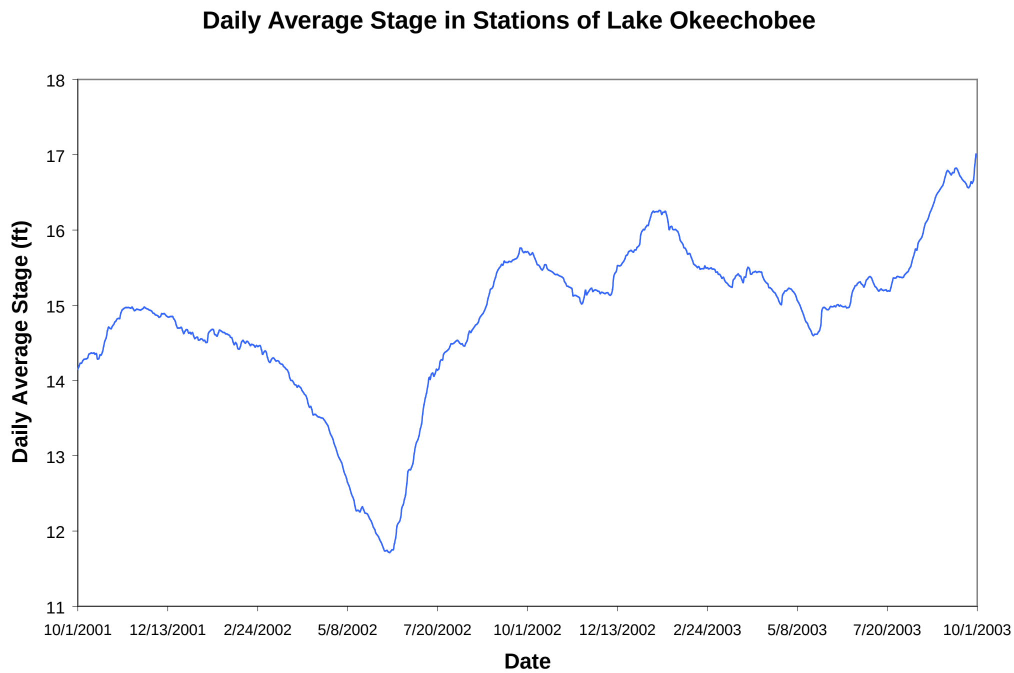


Figure 3.21: Daily average stage in the stations of Lake Okeechobee from 10/1/2001 to 9/30/2003.

3.5.1.1 Third Methodology for Lakes

The third methodology, a new methodology for lakes based on tabu search, should generate results directly comparable to those of the second methodology for lakes. Therefore, in the third methodology, it should be necessary to choose an admissible *RMSE* (i.e. 0.1 ft) and identify a required reliability, R_e (i.e. 0.8). The first condition determines for each day of the period whether the *RMSE* is adequate (*RMSE* ≤ admissible *RMSE*). The second condition sets the fraction of the time the subset should give adequate *RMSE* values. The actual fraction of the time or relative frequency, F_r, should be as close as possible to the required reliability, that is, minimizes $R_e - F_r$ *with* $F_r \leq R_e$. In summary, the subset that minimizes $R_e - F_r$ with $F_r \leq R_e$ and at the same time has the smaller size (smaller r) is selected as the optimal subset of the period.

Important details of the application of the tabu search method used in this methodology are presented next in the form of a pseudocode. Some important variables present in the pseudocode are:

- r_i, r_f: minimum and maximum value, respectively, of r to be considered
- *AdmRMSE*: *admissible RMSE*
- R_e: reliability, is the fraction of the time that the RMSE should be smaller than or equal to *AdmRMSE*.
- F_r: relative frequency = number of days that a subset has *RMSE* <= *admRMSE* divided by the number of days in which the subset operates.
- *it*, *itmax*, *itBtOpmax* and *itBtChgRmax*: iteration, maximum number of iterations, maximum number of iterations between consecutive best values of the objective function, and maximum number of iterations between changes of r.
- *xbest*, *xnow* and *xnext*: arrays of Boolean variables that represent several subsets of stations. A true value in the array's i position means that the station i is selected.
- *ObfFn_best*, *ObfFn_now* and *ObfFn_next*: objective functions of the respective subset of stations.
- *PrevSol*, *PrevBest* and *PrevNeigh*: collections (computational structures that use hashing to improve the retrieving of solutions previously found).

PrevSol stores previous solutions, *PrevBest* previous best solutions and *PrevNeigh* previous neighbors.

Pseudocode

```
Give the features of the network to be optimized
  (Such as n, the number of stations, and the coordinates and time series of the
   stage values associated with the stations).
Give ri, r, and rf
Set it = 0, itop = 0, itBtop = 0
Init visitedR array of booleans
VisitedR(r) = True

Randomly obtain the first subset and associated objective function: xnow and
ObjFn_now
  (The objective function is computed for all the valid days of the period
   considered for the optimization of the stage-measuring network optimization).
   The objective function is equal to 1 – Fr, where Fr = ncomplydays/nvaliddays,
   ncomplaydays is the number of days when RMSE ≤ admissible RMSE and nvaliddays
   is the number of days when the given subset is present).

Set xbest = xnow, ObfFn_best = ObjFn_now
Add xbest to PrevSol collection
Add xbest to PrevBest collection

Initialize the History matrix (n x n matrix)

minTabuTenure = 1
maxTabuTenure = n - 1
TabuTenure = (minTabuTenure + maxTabuTenure) / 2

Possible_change_r = 0
Do
  it = it + 1
  itop = itop + 1

  Identify neighborhood of xnow: Neigh()
    (Neigh() is an array of r * (n – r) neighboors of xnow)
  Add all new neighbors to PrevNeigh Collection

  Identify neighbors present in previous solutions

  Compute MoveVal for all neighbors with: MoveVal(i) = ObjFn(i) – ObjFn_now

  Update Frequency
    (Sum of the cell values below the principal diagonal of the History matrix)

  First round of identification of minMoveVal
    (The minMoveVal is the minimum MoveVal(i) corresponding to a neighbor that is
     not in the previous solutions and it is not tabu; that is, its tenure is less
     than it (tenure is recorded in the cells above the diagonal of the History
     matrix). If a MoveVal(i) is tabu, then if Moveval(i) + ObjFn_now < ObjFn_best
     holds, then minMoveVal = MoveVal(i) (This constitutes an aspiration
     criterion).

  Correction by frequency
  If minMoveVal > 0 then
```

```
   All MoveVal(i) = (SumFreq/corresponding cell of History) * MoveVal(i))
   Second round of identification of minMoveVal
     (The minMoveVal is the minimum corrected MoveVal(i) corresponding to a
      neighbor that is not in the previous solutions and is not tabu)
End if
The next solution is: xnext = Neigh(imin) and ObjFn_next = ObjFn(imin), where
  imin corresponds to minimum MoveVal(i).

Add xnext to PrevSol Collection

If it >= itmax or itBtOp >= itBtOpmax then
  If ObjFn_next < ObjFn_best then
    xbest = xnext and ObjFn_best = ObjFn_next
    Add xbest to PrevBest collection
    itop = it
    Compute number of function evaluations to find optimum
      (The number of previous neighbors plus the number of previous solutions)
    Obtain time to find optimum
  End if
  Exit Do
End if

xnow = xnext and ObjFn_now = ObjFn_next

Update frequency
  (Add 1 to the corresponding cell of History matrix)

Update recency
  (Assign it + TabuTenure to the corresponding cell of History matrix)

If ObjFn_now < ObjFn_best then
  xbest = xnow and ObjFn_best = ObjFn_now
  Add xbest to PrevBest collection
  itop = it
  Compute number of function evaluations to find optimum
  Obtain time to find optimum
  itBOp = 0
  If TabuTenure - 1 >= minTabuTenure then
    TabuTenure = TabuTenure - 1
  End if
  Possible_change_r = 0
Else
  If TabuTenure + 1 <= maxTabuTenure then
    TabuTenure = TabuTenure + 1
  End if
  Possible_change_r = Possible_change_r + 1
  If Possible_change_r >= itBtChgRmax then
    If 1 – ObjFn_best < Re then
      If r + 1 <= rf then
        If visitedR(r + 1) = False then
          visitedR(r + 1) = True
          r = r + 1
          Select from xnow a new xbest with an additional station
            (There are n – r + 1 neighbors to choose from. Add the new neighbors
             to PrevNeigh collection. Choose neighbor with the minimum objective
             Function value).
          Add xbest to PrevSol collection
          Add xbest to PrevBest collection
          it = it + 1
          itop = it
          Compute number of function evaluations to find optimum
          Obtain time to find optimum
```

```
          itBOp = 0
        End if
      End if
    Else
      If r - 1 >= ri then
        If visitedR(r - 1) = False then
          visitedR(r - 1) = True
          r = r - 1
          Deselect a station from xnow to obtain a new xbest
            (There are r + 1 neighbors to choose from. Add the new neighbors to
             PrevNeigh collection. Choose neighbor with the minimum objective
             Function value).
          Add xbest to PrevSol collection
          Add xbest to PrevBest collection
          it = it + 1
          itop = it
          Compute number of function evaluations to find optimum
          Obtain time to find optimum
          itBOp = 0
        End if
      End if
    End if
   End if
  End if
Loop

Compute total number of function evaluations
Obtain total time

Extract xbest from PrevBest collection
  (xbest is the subset of stations that have the maximum Fr smaller than or equal
   to the given Reliability and at the same time the smallest size)

Display/Print xbest and derived parameters
```

The previous pseudocode can be written in a simplified algorithmic form as:

1. A first subset of r stations is randomly chosen.
2. For the given size of subset, find the minimum of $1 - F_r$.
3. If F_r passes from $F_r > R_e$ to $F_r \leq R_e$ or from $F_r \leq R_e$ to $F_r > R_e$, the optimal subset has been found, it corresponds to the best value for the smaller size of the transition.
4. If $F_r > R_e$ then $r = r - 1$ otherwise $r = r + 1$. Go to Step 2.

An example that graphically explains the algorithm is shown in Figure 3.22. The algorithm, base of the third methodology, was applied as a VBA macro to the problem of finding the optimal subset of the Lake Okeechobee network (case study 8) with $F_r \leq R_e = 0.8$ and minimum size, considering an *admissible RMSE* = 0.1 ft. The results of the third methodology application are shown in Chapter 8. In addition, the results of the screening of the effect of two different annual periods of data in the solution of the same problem are also shown in Chapter 8. Finally, the results of this methodology to the case of the optimal subsets of Lake Okeechobee compared to those obtained with the second methodology for lakes are also presented in Chapter 8.

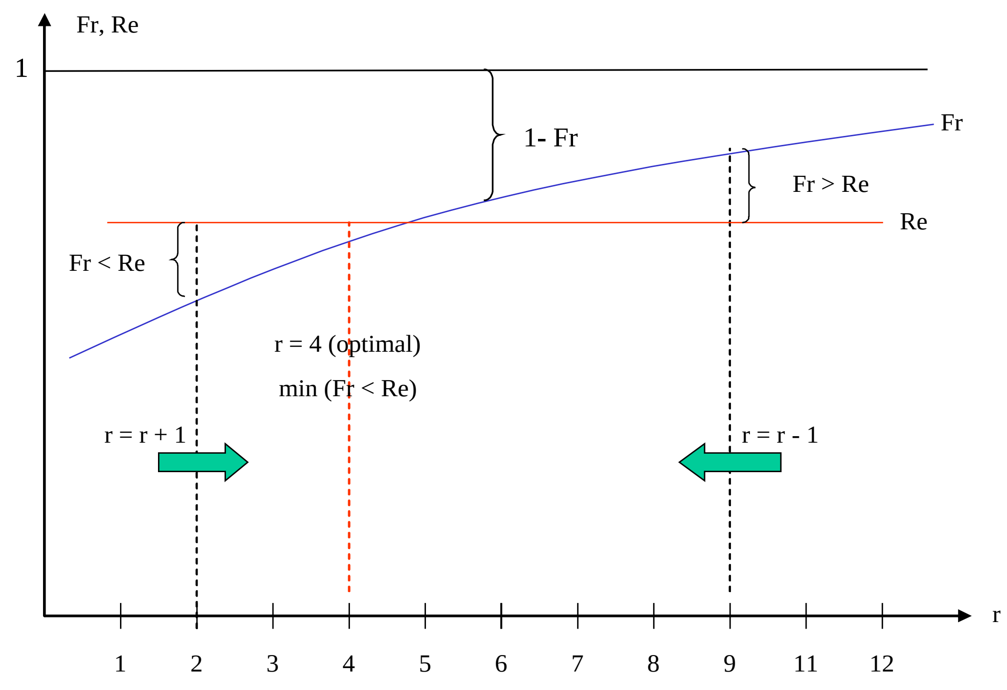


Figure 3.22: Algorithm shown in graphic form.

3.5.1.2 Fourth Methodology for Lakes

The use of an admissible *RMSE* value to identify the optimal subset for a given reliability, R_e, makes it necessary to repeat the process when a different admissible *RMSE* value is chosen. To simplify the process of optimizing the network and obtain in one run the optimal subsets of all possible sizes *r*, the previous pseudocode was changed. This change gives origin to the fourth methodology for lakes. The new pseudocode was implemented in a VBA macro and it finds the subsets of stage-monitoring stations of all possible sizes that minimize their respective average *RMSE*. To define *average RMSE* let us assume the existence of a set of *n* water level measuring stations with daily data from a time period of *T* days from which is needed to choose a subset of *r* stations. The root mean square error $RMSE_t$ for a given day *t*, defined over the $n - r$ omitted stations is:

$$RMSE_t = \sqrt{\frac{1}{n-r}\sum_{i=1}^{n-r}\left(I_{t,i} - M_{t,i}\right)^2} \tag{3.12}$$

where $I_{t,i}$ is the interpolated value on day *t* and $M_{t,i}$ is the measured value on day *t*, both for station *i*. Equation 3.12 is equation 3.1 but with *t* made explicit. Then, the *average RMSE* for the entire period of analysis, is

$$avgRMSE = \frac{1}{n_t}\sum_{t=1}^{n_t} RMSE_t = \frac{1}{n_t}\sum_{t=1}^{n_t}\sqrt{\frac{1}{n-r}\sum_{i=1}^{n-r}\left(I_{t,i} - M_{t,i}\right)^2} \tag{3.13}$$

where, $n_t \leq T$ is the number of valid days of the period of analysis. A day is valid day if the subset of *r* stations has measurements in all of them. The subset of *r* stations that minimize the value of the *average RMSE* is the optimal subset.

Results of the application of the fourth methodology to Lake Okeechobee (case study 8) are shown in Chapter 8.

3.5.1.3 Ordered list of Stations in the case of Lakes

A closely related problem to the problem of the optimal subsets of stations is the problem of the optimal ordered list of stations. The successive selection of stations to define subsets of size $r = 1$ to n, without changing any already selected station, can define an ordered list of stations. The steps to define an ordered list are:

1. Choose one station and compute the error of the remaining $n - 1$ stations.
2. Choose another station from the remaining $n - 1$ stations. Compute the error of the $n - 2$ stations.
3. Choose another station from the remaining $n - 2$ stations. Compute the error of the $n - 3$ stations.
4. Repeat the process for all the remaining stations until completing the ordered list of n stations.
5. Compute the sum of errors. That is, it is defined an ordered list and its associated global error.

A result that can be interesting to know is whether the global error of the optimal ordered list of stations is smaller than the sum of errors of the optimal combinations of stations of all possible sizes. The problem of finding the ordered list with the least global error can be solved using tabu search (see section 2.2.6). In this problem, contrary to the problem of the optimal subsets (where swap, insertion and deletion moves are required), only swap moves are needed. *Swap* *moves* are moves that interchange the order of two stations in a given ordered list and, following the tabu search algorithm, help create the set of neighbor ordered lists that must be evaluated to find the next best ordered list, until the optimal ordered list is found. An example of a swap move can be illustrated with the ordered list of five members (2, 5, 1, 4, 3), if the move (5, 4) is performed the new ordered list (2, 4, 1, 5, 3) is created. In fact, the total number of neighbors of a given ordered list of n members is $\frac{n(n-1)}{2}$; for the example, a total of 5 × (5 – 1) / 2 = 10 neighbors can be defined. In essence, this problem involves permutations. If a partial ordered list of r members were of interest, the total number of ordered lists of r members is

$$P_r^n = \frac{n!}{(n-r)!}. \tag{3.13}$$

The identity that relates the number of permutations with the number of combinations of given r and n is

$$C_r^n = \frac{n!}{r!(n-r)!} = \frac{P_r^n}{r!}. \tag{3.14}$$

Equation 3.14 can be interpreted stating that for each subset of r stations, $r!$ ordered lists of r members exist. However, the problem involves the identification of the optimal ordered list of n stations, so its *cardinality C*, the number of possible solutions, is equal the number of permutations of n objects taken from n objects, that is

$$C = P_n^n = n!\ . \tag{3.15}$$

In contrast, the problem of finding the best subsets of all possible sizes has cardinality equal to

$$C = \sum_{r=1}^{n-1} C_r^n = 2^n - 2 \; . \tag{3.16}$$

Figure 3.23 shows a graph of n versus the cardinality of both problems; the cardinality of the ordered list of stations grows faster (when n grows) than that of the subsets of stations.

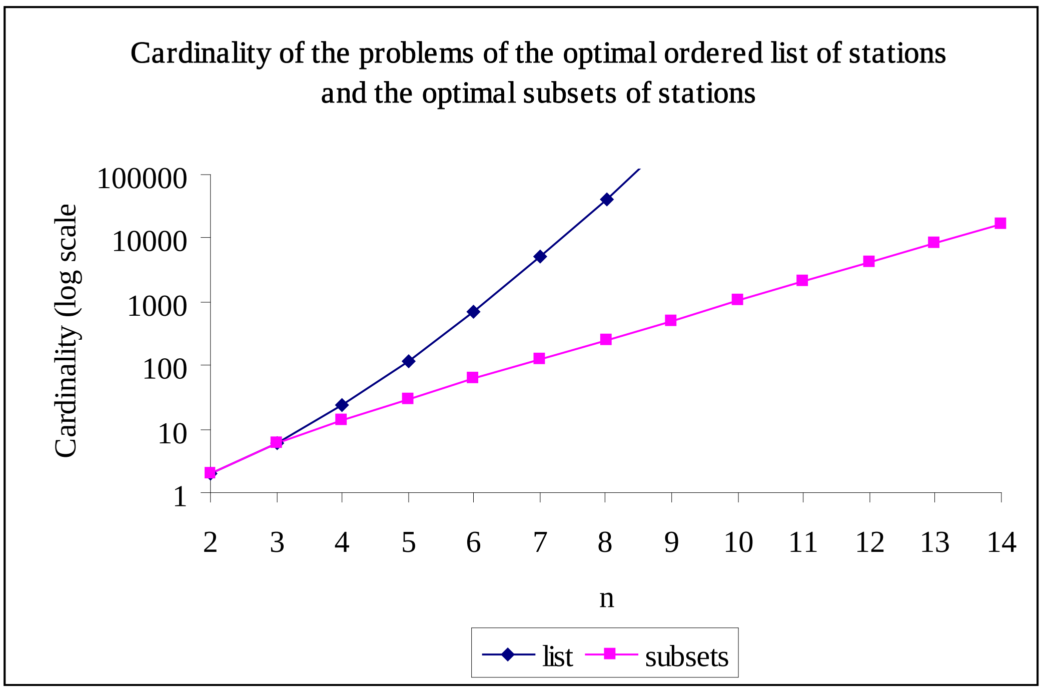


Figure 3.23: Cardinality of the problems of the optimal ordered list of n stations and the optimal subsets of $r = 1$ to $r = n - 1$ stations.

An alternative way to check the existence of an optimal ordered list of stations, without actually find it, is exemplified by the next test.

Test for checking the existence of a true optimal ordered list of gages

Given a set of n gages and a particular objective function (such as the average error, the relative frequency, etc) that needs to be optimized. Follow the next procedure:

- Find the optimal subset of size $r = 1$, call it S_1^*.
- Find the best subset of size $r = 2$, keeping in it the station in S_1^*, call it S_2^*. Here the second station must be chosen from the stations that are not in S_1^*.
- Find the best subset of size $r = 2$ considering all the n gages. Call it S_2^{**}. If in S_2^{**} the station in S_1^* is not present, the optimal ordered list of stations does not exist. If it is present then the size $r = 3$ must be checked.
- Find the best subset of size $r = 3$, keeping in it the stations in S_2^*, call it S_3^*. Here the third station must be chosen from the stations that are not in S_2^*.
- Find the best subset of size $r = 3$ considering all the n stations. Call it S_3^{**}. If in S_3^{**} the gages in S_2^* are not present, the optimal ordered list of gages does not exist, If they are present then the size $r = 4$ must be checked.
- Repeat for $r = 4, 5, \ldots, n - 1$, until the existence or non-existence of the optimal ordered list of gages is proved.

This test avoids finding the optimal ordered list of stations with the formulation previously stated. This test only needs the combinatorial approach and the identification of only one partial or complete ordered list.

However, to test the application of tabu search on the solution of the optimal ordered list of stations problem, a program was written for the case of lakes. Results of that program applied to Lake Okeechobee (case study 8) are shown in Chapter 8. Comparison with the resulting optimal subsets of stations of the same case study is also shown.

3.5.2 Third Methodology for Streams with Steady Flow

In the second methodology for streams with steady flow previously discussed, the optimization method used was *forward selection*. The selection of this method was done assuming that the time needed to find an optimal subset using a genetic algorithm would be too long because the evaluation of the objective function needs the repetitive computation of flow profiles (in the case of Kissimmee River each profile computation takes around 7 seconds in the computer used). The experience with tabu search in the case of lakes suggests the feasibility of implementing this method in the case of streams with steady flow. Based on the implementation of tabu search in the case of lakes and computer code used in the tools of the second methodology for lakes, a VBA macro was created to perform the optimization in streams with steady flow using tabu search. This macro is the core of the third methodology for streams with steady flow.

The reach of Kissimmee River in the PoolAE basin was again selected now as case study 9. Its description is in Section 3.3 (Figure 3.10, Figure 3.11, and Table 3.4). To screen the feasibility of optimization using the new optimization methodology only the first available 60 days (taken from 10/01/2001 to 07/15/2002) of the original period (from 10/01/2001 to 12/31/2003) were considered in this new analysis. The time of travel of water at average conditions along the reach is of the order of 13.6 days. The study of two consecutive periods of 30 days, roughly two times of travel and a combined period of 60 consecutive days may provide results that could include at least a complete flood wave. The available data does not provide consecutive days because a number of them were discarded in the calibration process of the first and second methodologies for streams with steady flow; however, for the purposes of this study it is assumed that they are consecutive. Figure 3.24 shows discharge in the structure S65C and water elevation in KRDRS, for the considered 60 days (water elevation on days 28 and 29 is missing). These stations located in the naturalized reach give a fair idea of the variation of both variables in the entire Pool AE stream.

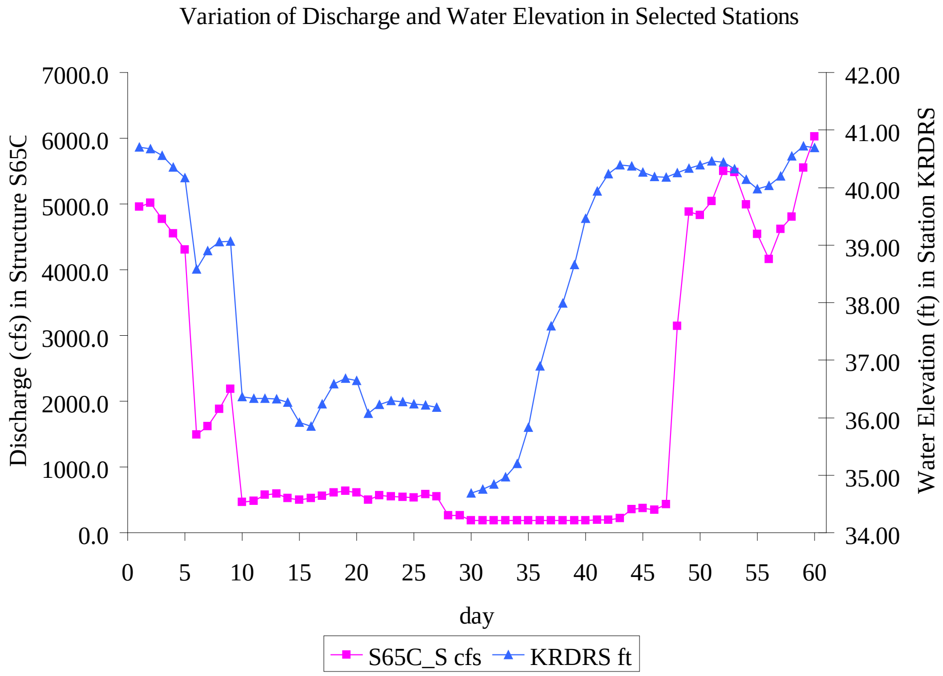


Figure 3.24: Variation of discharge and water elevation in stations related to the naturalized reach of Kissimmee River.

Discussion of the details that must be performed before applying the third optimization methodology for streams with steady flow are included in Section 3.3 and will not be repeated here. The new optimization methodology considers the average *RMSE* as the objective function and is able to identify the optimal subsets of stations from the subset defined with the number of stations that define the hydraulics of the reach plus an additional station to the subset with $n - 1$ stations. The corresponding VBA macro was applied to the reach of Kissimmee River located in Pool AE (case study 9). The possible existence of optimal ordered list of stations was also checked in the same case study. Results of this methodology are shown in Chapter 8.

3.6 CONCLUDING REMARKS

The development of the optimization methodologies began with the consideration that a single methodology was useful for lakes and streams. After applying the first methodology, based on IDW interpolation and genetic algorithm optimization, to the stage-monitoring network of Pool BC (case study 3), it was clear that a different methodology was needed for the case of stage-monitoring networks in streams because the spatial variation of water level in lakes and in streams is different. In lakes, with very small water velocity, water level changes smoothly from point to point, and in streams water level can abruptly change from point to point obeying the hydraulic conditions along the course. In addition, the failure of the identification of the optimal subset of stations of the Pool BC subbasin prompted the creation of an additional step that completed the process of identification of the optimal subset of stations in a lake. This addition completed the second methodology for lakes; this methodology was applied to the Lake Kissimmee stage-monitoring network (case study 4).

The main stream of Pool AE, considering a period of 15 months of daily data (case study 5), was used to develop and test the methodology for streams with steady flow. The development process of this methodology considered the proposition of the first methodology for streams, which begins the optimization of the base subset of stations that hydraulically models the stream and then subjectively adds one station to the base subset for the next cycle of optimization. Results of the first methodology suggested that the station with the maximum root mean square error in a cycle of optimization must be part of the base subset in the next cycle. This refinement was incorporated in the second methodology. In this methodology, the HEC-RAS steady flow module was used as interpolation method and forward selection, a sequential method, as the optimization method. The creation and calibration of the HEC-RAS model of the stream are performed before the network is optimized. The River Channel Morphology Model was used to create an analytical river channel of the naturalized reach located in the Pool BC area.

The main stream of Pool AE with data of one and seven days, respectively case studies 6 and 7, were used to develop and test the methodology for streams with unsteady flow. Initially a calibration study was performed. Results of the calibration study influenced the proposition of the first methodology for streams with unsteady flow. From the first methodology, the second methodology was proposed. This second methodology is similar to the second methodology for streams with steady flow. The main difference is that in this case, the HEC-RAS unsteady flow module is used as the interpolation method.

An additional study of the second methodologies for lakes and streams tested the possible change of optimization method. The change from genetic algorithm to tabu search seeks the reduction of computational time in the optimization of networks in lakes and the possibility of using tabu search instead of forward selection, in the case of networks in streams with steady flow. The two new developed methodologies, the third and four methodologies, based on tabu search were applied to Lake Okeechobee (case study 8). Case study 9 located in

Kissimmee River, was used in the case of the third optimization methodology for streams with steady flow.

A summary of the proposed optimization methodologies is shown in Table 3.8. The table includes the optimization method, the objective function used in each methodology, the section where the methodology is described, and (in the case of being assigned) the number of the equation that defines the objective function. Nine methodologies were proposed and were tested, four for lakes, three for streams with steady flow, and two for streams with unsteady flow. The methodologies numbered 8 and 9 in Table 3.8, were adopted and named (LOSNO and KRSNO) after analyzing the entire set of results obtained (see section 8.4).

Number	Methodology	Optimization method	Objective function	Section/ (equation)
1	First, Lakes	Genetic Algorithm	max number of days with $RMSE \leq admRMSE$	3.2.2
2	Second, Lakes	Genetic Algorithm	max $P(RMSE \leq admRMSE)$ or max F_r	3.2.3/(3.5)
3	First, streams, steady flow	Forward selection	max *RMSESt*	3.3.1/(3.6)
4	Second, streams, steady flow	Forward selection	max *RMSESt*	3.3.1/(3.6)
5	First, streams, unsteady flow	Forward selection	max *RMSESt*	3.4.2/(3.6)
6	Second, streams, unsteady flow	Forward selection	max *RMSESt*	3.4.2/(3.6)
7	Third, Lakes	Tabu search	min $(F_r - R_e)$ with $F_r \leq R_e$	3.5.1.1/(3.6)
8	Fourth, Lakes (LOSNO)	Tabu search	min *average RMSE*	3.5.1.2/(3.13)
9	Third, streams, steady flow (KRSNO)	Tabu search	min *average RMSE*	3.5.2/(3.13)

Table 3.8: Summary of studied methodologies

Finally, a summary of the case studies used to develop and test the optimization methodologies is shown in Table 3.9. The table includes the geographic area, time period, and methodology in which each case study was used.

Case study	Geographic area	Time Period	Studied with methodology
1	Synthetic	1 day	Initial comparison of simulated annealing and genetic algorithm
2	Pool BC	08/08/2002	
3	Pool BC	10/01/2001 - 9/30/2004	First, lakes
4	Lake Kissimmee	10/01/2001 - 9/30/2003	Second, lakes
5	Kissimmee River in Pool AE	10/01/2001 - 12/31/2003	First and second, streams, steady flow
6	Kissimmee River in Pool AE	10/01/2001 - 10/02/2001	First, streams, unsteady flow
7	Kissimmee River in Pool AE	10/01/2001 - 10/08/2001	Second, streams, unsteady flow
8	Lake Okeechobee	10/01/2001 - 9/30/2003	Third and fourth, lakes
9	Kissimmee River in Pool AE	10/01/2001 - 07/15/2002	Third, streams, steady flow

Table 3.9: Summary of case studies

Chapter 4: Procedures of Implementation

4.1 Introduction

Several toolsets were developed as a result of this research, however, only the three toolsets that implement the second methodology for lakes (based on genetic algorithm), the second methodology for streams with steady flow (based on forward selection), and the second methodology for streams with unsteady flow (also based on forward selection), are presented. Each toolset is closely related to its respective optimization methodology. The description of each methodology also contains the description of its corresponding toolset.

4.2 Second Methodology for Optimization of Stage-Monitoring Networks in Lakes

The proposed methodology considers the use of a series of GIS Tools. These tools were developed as the problem was being solved. In the next two sections, a description of both tools and methodology is presented.

4.2.1 Network Optimization in Lakes Toolset

The toolset for stations in lakes was developed in VBA and is implemented in an ArcMap document. The tools are invoked from a toolbar. The SFWMD's DBHydro Web Page <http://www.sfwmd.gov/org/ema/dbhydro/index.html> was the main source of the data. Later the data were loaded into a geodatabase (Figure 4.1).

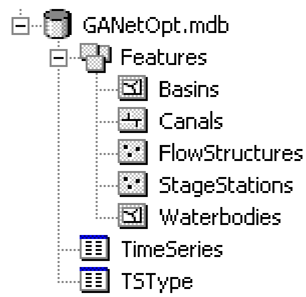


Figure 4.1: Network optimization in lakes' geodatabase structure.

The geodatabase has several feature classes; the most important is a feature class of points called StageStations whose attributes are those of the water elevation stations of the study area. A time series table was also created using the ArcHydro format; it has water elevation data from October 1, 2001, to September 30, 2004. The application of the tools produces results that are stored in several text files.

Users access the Network Optimization in Lakes toolset through its toolbar (Figure 4.2). The toolset has four tools.

Figure 4.2: Network optimization in lakes toolbar

The tool OptByRange (Optimization by Range) and the tool OptByAdmRMSE (Optimization by Admissible RMSE) identify optimal subsets of stations. The tool StatsOfDailyOptSubsets (Statistics of Daily Optimal Subsets) and the tool StatsOfOptSubsetsAllPeriod (Statistics of Optimal Subsets for the entire period) compute statistics of the daily optimal subsets. Users could use either of the optimization tools once they interactively selected the set of stations to optimize. A period of analysis, a range of *r* (the size of the optimal subset), and in the case of the OptbyAdmRMSE tool, an admissible *RMSE*, must be provided. The StatsOfDailyOptSubsets tool computes the statistics of the daily optimal subsets of the period using the results of any one of the optimization tools. The StatsOfOptSubsetsAllPeriod tool uses the results of the optimization tool and those of the first statistical tool to identify the optimal subset of stations. A complete description of this toolset, illustrated with an example can be found in the Users Manual (Appendix A)

4.2.2 Second Methodology for Network Optimization in Lakes

Once all the station and stage data from DBHydro has been prepared and loaded into the corresponding geodatabase, the method used to solve the problem of the spatial optimization of the stations in a lake involves four main steps (Figure 4.3). The user manually does the first two steps before using the created tools.

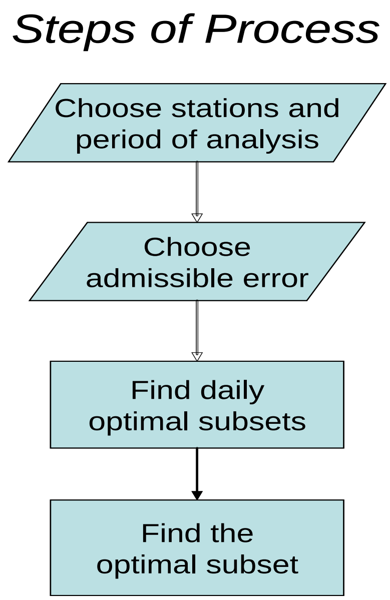


Figure 4.3: Network optimization in lakes process steps.

Choose stations and period of analysis. The first step requires the selection of a set of stations and the period of analysis. It is important to choose stations that share the same period of record, and whose operating measuring conditions remain the same during the period of analysis. To have good results it is recommended to choose at least two years of data.

Choose admissible error. The second step requires choosing an admissible error. For preliminary assessments, the admissible *RMSE* suggested by SFWMD is between 0.05 and 0.1 ft.

Find daily optimal subsets. The third step identifies the optimal sets of stations for each day of the period of analysis. The genetic algorithm method of optimization is used each time a change occurs in the set of stations or in the required number of stations in the subset of minimum *RMSE*.

Find the optimal subset. Finally, the optimal daily subsets of station results obtained in the third step are analyzed to get the optimal subset of stations for the entire period. This analysis is a two-step process. The first process identifies day by day the optimal station subset that has simultaneously an *RMSE* less than or equal to the admissible *RMSE* and the smallest number of stations. The second process computes the relative frequency for the entire time series period where the station subset complies with the constraint of the admissible *RMSE*. Finally, the station subset with the highest relative frequency becomes the optimal subset of stations for the period of analysis.

4.3 Second Methodology for Optimization of Stage-Monitoring Networks in Streams (Steady Flow)

The proposed methodology considers, as in the case of stations in lakes, the use of a series of GIS Tools. The origin of the tools was parallel to the solution of the problem. In the next two sections a description of both tools and methodology are included.

4.3.1 Network Optimization in Streams with Steady Flow Toolset

The toolset for stations in streams has two tools, the NetOptStreams tool was developed in VBA while the NetOptStreamsGr tool was developed using Visual Basic (VB). Both tools are implemented in an ArcMap document. The tools are invoked from a toolbar. The flow and water elevations that come within the ArcMap document were provided by the SFWMD or downloaded from DBHydro. Later the data were loaded into two dbf tables (RASTABLE.dbf and CrossSections.dbf) ready to be queried when the NetOptStreams tool is activated.

The RASTABLE.dbf file provides the tool with the data to build the HEC-RAS flow file needed to compute profiles. While the CrossSecctions.dbf file stores measured and computed water elevations that are used to compute errors in each cycle of optimization. These tables have data from October 1, 2001, to December 30, 2003. The application of the NetOptStreams tool produces results that are stored in several text files. From the produced text files, the NetOptStreamsGr tool creates a series of graphs of results.

Users access the Network Optimization in Streams toolset through its toolbar (Figure 4.4).

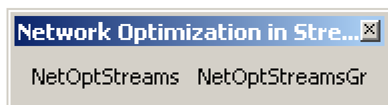


Figure 4.4: Network optimization in streams toolbar.

The NetOptStreams tool (Network Optimization in Streams) implements the sequential optimization process outlined before. It computes profiles using HEC-RAS, root mean square errors by date (*RMSE*), errors by station, and root mean square errors by station (*RMSESt*), for the range of sizes in the base subset of stations. The user can interactively choose the name of the HEC-RAS project in which the streams have been previously modeled, and the prefix to use to store results in several files. In addition, the user can set the initial base subset of stations and the maximum size of the base subset. Finally, the user can launch the optimization process, which implies the interaction between the customized code in ArcGIS and HEC-RAS, and has the following steps:

Step 1. Erase any previous computed water elevations from the table CrossSections.dbf. The remaining water elevations in table CrossSections.dbf are observed water elevations in the locations of all the stations of the stream for the period of analysis.

Step 2. From the base subset of stations stored in the table CrossSections.dbf, set the required internal boundaries of change of water elevation in the flow file of the HEC-RAS project used to model the stream.

Step 3. For each valid date of the period, write in the HECRAS flow file the values of flow in flow changes, the stage in the downstream stage boundary condition, and the stage in internal changes of stage conditions. All this daily values are stored in the table RASTable.dbf. Call HEC-RAS to compute the water profile. Call the routine of HEC-RAS that exports geo-referenced results to an sdf File. See USACE (2004). From the sdf file, obtain an xlm file. From that xlm file, read the water elevations computed at the stations outside the base subset and store them in the table CrossSections.dbf.

Step 4. For each date, compute *RMSE* from the water elevations that are stored in table CrosssSections. Store results in appropriate text file.

Step 5. For each station, compute *RMSESt* from the water elevations stored in table CrosssSections. Store results in appropriate text file. Identify the station with the maximum *RMSESt*.

Step 6. In table CrossSections, set the station with maximum *RMSESt* because it goes into the base subset. If the new size of base subset is smaller than or equal to the maximum size of base subset go to Step 1, otherwise, end process.

Once the optimization process is finished, the user can use the NetOptStreamsGr tool to plot three types of graphs: errors by station, *RMSE* for all stations, and *RMSESt* by size of base subset. Further details of this toolset are in the Users Manual (Appendix A).

4.3.2 Second Methodology for Network Optimization in Streams (Steady Flow)

The method used to solve the problem of optimization of the stations in streams involves the following main steps:

Choose stations and period of analysis. Select the set of *n* stations along a stream and the period of analysis. These stations and period of analysis should have the same conditions that are mentioned for the case of stations in lakes.

Choose admissible error. The second step requires choosing an admissible error. Here it is important to consider that the square of the flow's error is proportional to the error of the head at the hydraulic structure. The study of the *RMSE* of the base subset of stations might be complemented with the study of the errors in individual stations.

Model the system in HEC-RAS. In the HEC-RAS model, all the necessary internal known water elevations must be set. In subcritical flow, these elevations are at the headwater side of all the hydraulic structures inside the studied stream. These stations constitute the first subset of *r* stations taken from a set of *n* stations. The HEC-RAS modeling process must include the calibration of the computed profiles changing the Manning's roughness and the roughness factors to obtain computed profiles as similar as possible to the observed profiles. The modeling process can be initially done with a limited number of daily datasets. Once all the data of the period is loaded in the appropriate files, a second round of calibration can be done until the average *RMSE* becomes stable.

Optimize the station network along streams. This process begins with the initial subset of r stations. In each loop, a new station is added to the subset until an exit condition holds. Figure 4.5 shows a flow chart that explains the process.

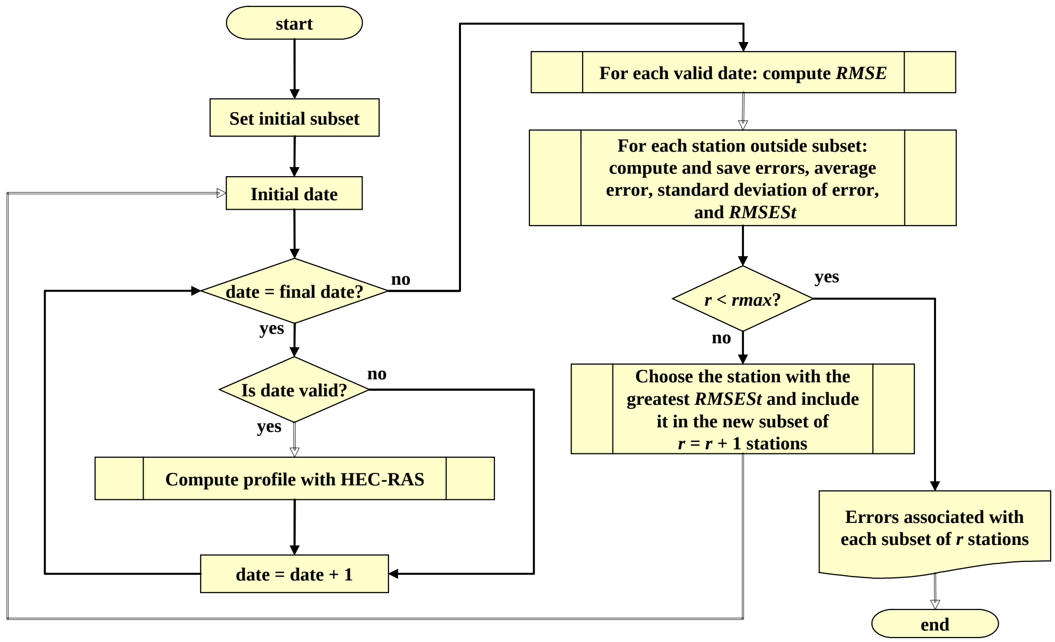


Figure 4.5: Flow chart of the optimization of stations in streams process.

A valid date is a date where all the stations in the base subset have data. The maximum number of stations in the base subset, r_{max}, must be smaller than n, the number of stations. In this context, as is defined by equation 3.6, *RMSESt* is the root mean square error of the estimated water elevation in a given station along the period of analysis. Similar to the optimization method for stations in lakes, users may establish an exit condition for this optimization process. Instead of considering a range of subset sizes, users may end the process when an admissible *RMSESt* is reached.

Interpret Results and Find the Optimal Base Subset. According to the chosen admissible measure of error, interpret the optimization results and decide if one of the identified base subsets could be the optimal base subset of stations.

4.4 SECOND METHODOLOGY FOR OPTIMIZATION OF STAGE-MONITORING NETWORKS IN STREAMS (UNSTEADY FLOW)

The proposed methodology considers, as in the two previous cases, the use of a series of GIS Tools. As before, the origin of the tools was parallel to the solution of the problem. In the next two subsections a description of both tools and methodology are included.

4.4.1 Network Optimization in Streams with Unsteady Flow Toolset

The toolset for stations in streams in unsteady flow has two tools, the UNetOptStreams and UNetOptStreamsGr tools. To distinguish them from the tools for steady flow their names are prefixed with a “U”. Both tools were developed from the corresponding tools for steady flow, are implemented in an ArcMap document, and are invoked from a toolbar. The flow and water elevations that come within the ArcMap document were provided by the SFWMD or downloaded from DBHydro. Later the data were loaded into two dbf tables (RASTABLE.dbf and CrossSections.dbf) ready to be queried when the UNetOptStreams tool is activated. The RASTABLE.dbf file provides the tool with the data to build the HEC-RAS unsteady flow file needed to compute profiles. The CrossSecctions.dbf file stores measured and computed water elevations that are used to compute errors in each cycle of optimization. Currently, these tables have data from October 1, 2001 0:00 hr, to October 8, 2001 0:00 hr. The application of the UNetOptStreams tool produces results that are stored in several text files. From the produced text files, the UNetOptStreamsGr tool creates a series of graphs of results.

Users access the Network Optimization in Streams toolset through its toolbar (Figure 4.6).

Figure 4.6: Network optimization in unsteady flow streams toolbar.

The UNetOptStreams tool (Network Optimization in Streams in Unsteady Flow) implements the sequential optimization process outlined before. It computes profiles using HEC-RAS, root mean square errors by date (*RMSE*), errors by station, and root mean square errors by station (*RMSESt*), for the range of sizes in the base subset of stations. The user can interactively choose the name of the HEC-RAS project in which the streams have been previously modeled, and the prefix to use to store results in several files. In addition, the user can set the initial base subset of stations and the maximum size of the base subset. Finally, the user can launch the optimization process. This optimization process is similar to that of the case of streams with steady flow. Once the optimization process is finished, the user can use the UNetOptStreamsGr tool. This tool plots three types of graphs: errors by station, *RMSE* for all stations, and *RMSESt* by size of base subset. Details of the use of this toolset are completely analogous to those of the toolset for steady flow (See the NetOptStreams Toolset User Manual in Appendix A). The difference is in the preparation of data to be loaded in the two dbf-tables mentioned. As well as in HEC-RAS, instead of using steady flow computations the user must use unsteady flow computations (USACE 2004).

4.4.2 Second Methodology for Network Optimization in Streams (Unsteady Flow)

The method used to solve the problem of optimization of the stations in streams with unsteady flow involves the following main steps.

Choose stations and period of analysis. Select the set of *n* stations along a stream and the period of analysis. All the stations should have data for the entire selected period. The period may consider one or two weeks with the widest possible variation of flows and stages, to try to asses the behavior of errors for an ample range of conditions. If a high quality of results is required, data should be instantaneous data, not hourly or daily data.

Choose admissible error. The second step requires choosing an admissible error. Here it is important to consider that the square of the flow's error is proportional to the error of the head at the hydraulic structure. The study of the *RMSE* of the base subset of stations might be complemented with the study of the errors in individual stations.

Model the system in HEC-RAS. In the HEC-RAS model, all the necessary internal known water elevations must be set. Each sub reach defined between two structures (or one structure and an external boundary condition) needs at least two boundary conditions: one water elevation condition at the downstream end, and a flow condition at the upstream end. The water elevation conditions are those at the headwater side of all the hydraulic structures inside the studied stream. The flow boundary conditions are given by the flows crossing the upstream structures and/or the uniform lateral flows entering the stream. The stations that hydraulically model the stream constitute the first subset of *r* stations taken from a set of *n* stations. The HEC-RAS modeling process must include the calibration of the computed profiles changing the Manning's roughness, roughness factors or whatever necessary parameter, to obtain computed profiles as similar as possible to the observed profiles. The modeling process can be initially done with a limited period of simulation. Once all the data of the period are loaded in the appropriate files, a second round of calibration is performed until the average *RMSE* becomes stable.

Optimize the station network along streams. This process begins with the initial subset of *r* stations. In each loop, a new station is added to the subset until an exit condition holds. Figure 4.7 shows a flow chart that explains the process.

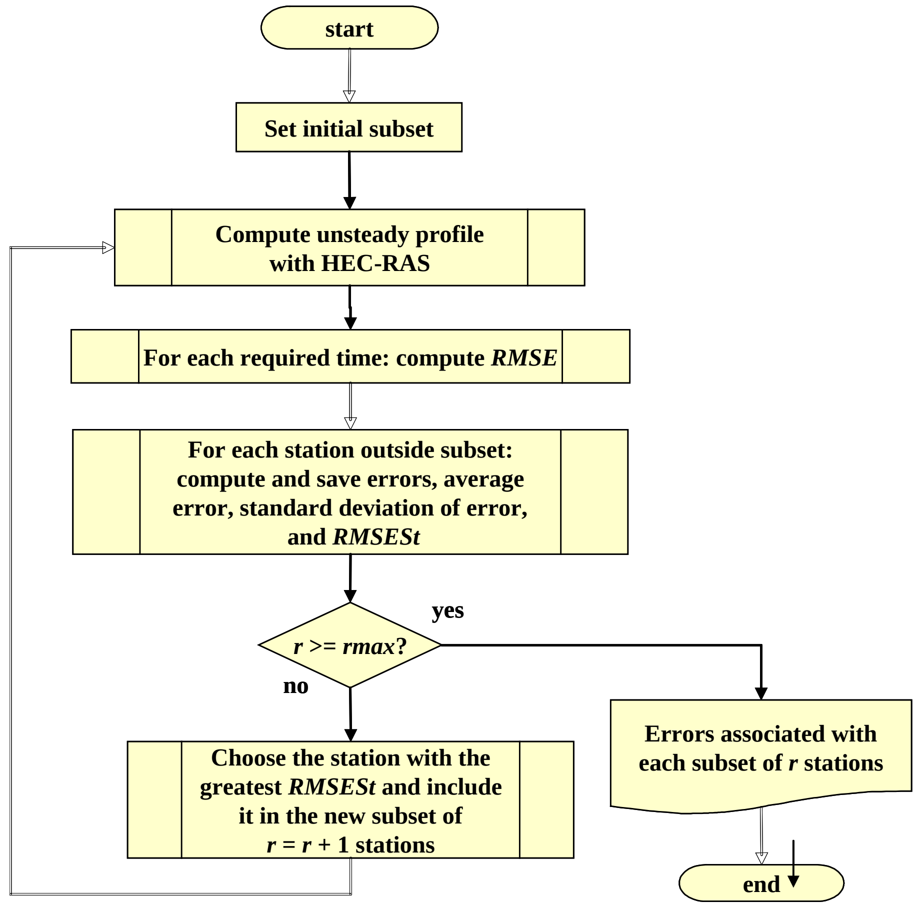


Figure 4.7: Flow chart of the optimization of stations in streams (unsteady flow) process.

The maximum number of stations in the base subset, r_{max}, must be smaller than n, the number of stations. In this context, as is defined by equation 3.6, *RMSESt* is the root mean square error of the estimated water elevation in a given station along the period of analysis. Similar to the previous optimization methods, users may establish an exit condition for this optimization process. Instead of considering a range of subset sizes, users may end the process when an admissible *RMSESt* is reached.

Interpret Results and Find the Optimal Base Subset. According to the chosen admissible measure of error, interpret the optimization results and decide if one of the identified base subsets could be the optimal base subset of stations.

Chapter 5: Results of the Network Optimization in Lakes

5.1 INTRODUCTION

Results are ordered closely following the development of Chapter 3. They consider the main results and the general history of this research. First, the results of the comparison of simulated annealing and genetic algorithm using two datasets are presented. Second, the most important results of the application of the first methodology to stations of Pool BC are described. Finally, the results of the application of the second methodology to the Lake Kissimmee stage-monitoring stations are presented.

5.2 RESULTS OF THE COMPARISON OF SIMULATED ANNEALING AND GENETIC ALGORITHM

Simulated annealing and genetic algorithms are considered good optimization methods for the problem of this research. Nevertheless, only one was considered to be used, the better suited to solve the problem. Both methods were applied to the optimization problem of the two 17 stage-monitoring station datasets (case studies 1 and 2) described in Chapter 3. Twenty realizations were made for each dataset and each method for all possible sizes of subsets (r = 1 to 16). Results are summarized in Tables 5.1 to 5.4. In each of these tables, the first column indicates the size of subset, r. The second column is the true minimum *RMSE* (ft). Column 3 is the smallest value of the searched minimum *RMSE* obtained in 20 realizations (the total number of realizations). Column 4 is the number of realizations in which the smallest value was found. Column 5 is the number of realizations in which the smallest value was found divided by 20. Column 6 is the greatest value of the searched minimum *RMSE* obtained in 20 realizations. Columns 7 and 8 are analogous to columns 4 and 5 but for the greatest value of the searched minimum *RMSE*. Column 9 presents the ratio of the greatest value divided by the smallest value of minimum *RMSE*.

1	2	3	4	5	6	7	8	9
r	True *RMSE* min	*RMSE* min	reali-zations	real./ total real.	*RMSE* max	reali-zations	real./ total real.	Max/min
1	6.187	6.187	10	0.500	6.431	1	0.050	1.040
2	2.031	2.031	9	0.450	2.945	1	0.050	1.450
3	1.368	1.368	11	0.550	1.889	1	0.050	1.381
4	1.201	1.201	11	0.550	1.313	1	0.050	1.093
5	0.926	0.926	10	0.500	1.342	1	0.050	1.449
6	0.840	0.840	11	0.550	1.138	2	0.100	1.355
7	0.864	0.864	2	0.100	1.005	2	0.100	1.162
8	0.638	0.638	13	0.650	0.736	1	0.050	1.153
9	0.425	0.425	9	0.450	0.626	2	0.100	1.472
10	0.358	0.358	10	0.500	0.569	1	0.050	1.588
11	0.460	0.460	8	0.400	0.563	1	0.050	1.223
12	0.414	0.414	14	0.700	0.716	1	0.050	1.730
13	0.331	0.331	13	0.650	0.688	1	0.050	2.076
14	0.446	0.446	10	0.500	0.711	1	0.050	1.596
15	0.382	0.382	9	0.450	0.685	1	0.050	1.795
16	0.132	0.132	4	0.200	1.727	1	0.050	13.064

Table 5.1: Simulated annealing. Case study 1. Total time 514.2 s.

1	2	3	4	5	6	7	8	9
r	True *RMSE* min	*RMSE* min	reali-zations	real./ total real.	*RMSE* max	reali-zations	real./ total real.	max/min
1	6.187	6.187	10	0.500	6.431	1	0.050	1.040
2	2.031	2.031	13	0.650	3.862	1	0.050	1.902
3	1.368	1.368	16	0.800	1.654	4	0.200	1.209
4	1.201	1.201	16	0.800	1.303	1	0.050	1.085
5	0.926	0.926	12	0.600	1.339	1	0.050	1.446
6	0.840	0.840	11	0.550	1.138	2	0.100	1.355
7	0.864	0.864	12	0.600	1.042	1	0.050	1.206
8	0.638	0.638	15	0.750	0.662	5	0.250	1.039
9	0.425	0.425	16	0.800	0.541	1	0.050	1.271
10	0.358	0.358	13	0.650	0.527	1	0.050	1.472
11	0.460	0.460	18	0.900	0.479	1	0.050	1.040
12	0.414	0.414	16	0.800	0.538	2	0.100	1.300
13	0.331	0.331	17	0.850	0.564	1	0.050	1.702
14	0.446	0.446	17	0.850	0.600	1	0.050	1.346
15	0.382	0.382	20	1.000	0.382	20	1.000	1.000
16	0.132	0.132	20	1.000	0.132	20	1.000	1.000

Table 5.2: Genetic algorithm. Case study 1. Total running time 164.6 s.

1	2	3	4	5	6	7	8	9
r	True *RMSE* min	*RMSE* min	reali-zations	real./ total real.	*RMSE* max	reali-zations	real./ total real.	max/min
1	1.349	1.349	5	0.250	1.416	2	0.100	1.050
2	0.506	0.506	11	0.550	0.596	1	0.050	1.178
3	0.346	0.346	7	0.350	0.348	7	0.350	1.005
4	0.178	0.178	13	0.650	0.187	1	0.050	1.048
5	0.106	0.106	7	0.350	0.143	1	0.050	1.346
6	0.073	0.073	8	0.400	0.108	1	0.050	1.469
7	0.086	0.086	11	0.550	0.089	1	0.050	1.039
8	0.076	0.076	7	0.350	0.091	1	0.050	1.190
9	0.063	0.063	8	0.400	0.068	1	0.050	1.090
10	0.059	0.059	9	0.450	0.060	1	0.050	1.031
11	0.039	0.039	10	0.500	0.068	1	0.050	1.761
12	0.037	0.037	12	0.600	0.061	1	0.050	1.648
13	0.038	0.038	9	0.450	0.052	1	0.050	1.365
14	0.046	0.046	10	0.500	0.057	1	0.050	1.243
15	0.012	0.012	13	0.650	0.067	1	0.050	5.475
16	0.009	0.009	6	0.300	0.168	1	0.050	19.711

Table 5.3: Simulated annealing. Case study 2. Total running time 539.5 s.

1	2	3	4	5	6	7	8	9
r	True *RMSE* min	*RMSE* min	reali-zations	real./ total real.	*RMSE* max	reali-zations	real./ total real.	Max/min
1	1.349	1.349	13	0.650	1.397	2	0.100	1.036
2	0.506	0.506	14	0.700	0.596	1	0.050	1.178
3	0.346	0.346	12	0.600	0.349	1	0.050	1.008
4	0.178	0.178	14	0.700	0.186	1	0.050	1.043
5	0.106	0.106	9	0.450	0.155	1	0.050	1.457
6	0.073	0.073	15	0.750	0.108	1	0.050	1.469
7	0.086	0.086	6	0.300	0.090	1	0.050	1.052
8	0.076	0.076	8	0.400	0.093	1	0.050	1.212
9	0.063	0.063	12	0.600	0.068	2	0.100	1.090
10	0.059	0.059	7	0.350	0.061	1	0.050	1.037
11	0.039	0.039	13	0.650	0.069	1	0.050	1.795
12	0.037	0.037	13	0.650	0.046	1	0.050	1.254
13	0.038	0.038	18	0.900	0.051	1	0.050	1.337
14	0.046	0.046	17	0.850	0.049	3	0.150	1.062
15	0.012	0.012	18	0.900	0.034	2	0.100	2.766
16	0.009	0.009	20	1.000	0.009	20	1.000	1.000

Table 5.4: Genetic algorithm. Case study 2. Total running time 174.5 s.

From the comparison of columns 2 and 3 in Tables 5.1 to 5.4, it can be concluded that for all sizes of optimal subset both optimization methods reached the optimal value of *RMSE* at least once. The genetic algorithm was faster than simulated annealing, with an average computing time (running VBA in a Dell Workstation PWS360 with a 3.2 GHz Pentium 4 Processor and 1.5 GB of RAM) of 170 s for genetic algorithm and 527 s for simulated annealing. The average ratio between the maximum and the minimum value of the optimum obtained was greater in simulated annealing (2.415) compared with that of genetic algorithm (1.288). This means, compared to simulated annealing, the optimal values from genetic algorithm are closer to the true minima of the objective function. Finally, the genetic algorithm procedure is more likely to obtain the true minimum than simulated annealing. In these two case studies, on average, genetic algorithm obtained the true minimum in 70% of the realizations compared with 47% in the case of simulated annealing. Table 5.5 summarizes these findings.

Criterion	Simulated Annealing	Genetic Algorithm
Average computing time (s)	527	170
Was the optimal value reached at least once for each size of optimal subset?	Yes	Yes
Average ratio between minimum and maximum value of obtained optimum	2.415	1.288
Average ratio of successful realizations in case study 1 (Synthetic data) (level of confidence 95%)	0.481 ± 0.075	0.756 ± 0.074
Average ratio of successful realizations in case study 2 (Pool BC 8/08/2002) (level of confidence 95%)	0.456 ± 0.060	0.653 ± 0.100

Table 5.5: Comparison of optimization procedures.

The case studies used to test both optimization algorithms have more variation than the variation expected in lakes, particularly that of subbasin S-65BC, because in that subbasin, along the canalized Kissimmee River, several hydraulic structures break the water surface producing noticeable discontinuities. These discontinuities make interpolation methods more prone to report high water elevation estimation errors. Consequently, these datasets offer good testing ground for the estimation of the robustness of both methods. Finally, based on the results of the comparison of performance of both methods, the genetic algorithm method was adopted for the optimization methodology for lakes.

In Figure 5.1, the optimal subset of r = 10 gauging stations (in color blue) of n = 17 operating stage-monitoring stations on 08/08/2002 in Subbasin S-65BC (Case study 2, station PC41 has no data for that day) is shown. The *RMSE* value is on the order of 0.06 ft.

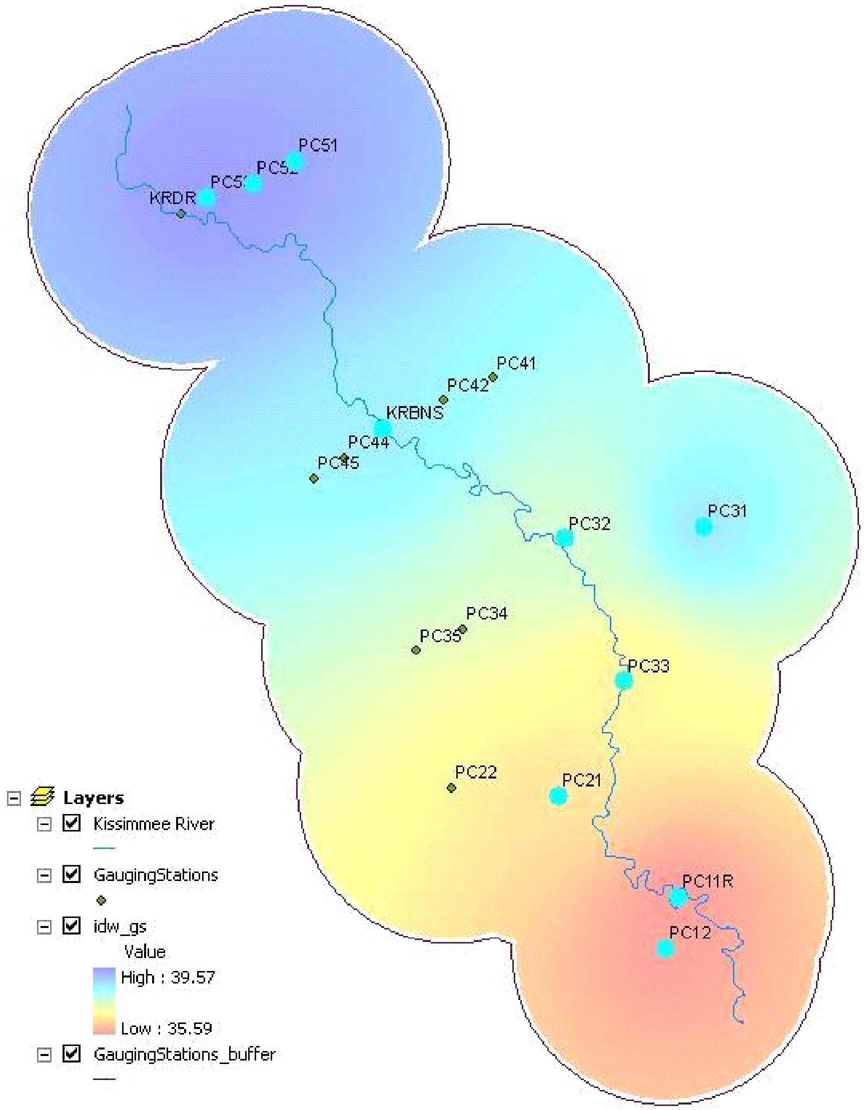


Figure 5.1: Example of an optimized stage-monitoring network using IDW (Case study 2, Kissimmee River in subbasin S-65BC).

5.3 ATTEMPT OF OPTIMIZATION OF THE SUBBASIN S-65BC STAGE MONITORING NETWORK

The first methodology for lakes was applied to case study 2. Two one-year periods were first used to find the optimal subset of stage-monitoring stations in the sub basin S-65BC: from October 1, 2001, to September 30, 2002, and from October 1, 2002, to September 30, 2003. Because the most frequent subsets of stations were different for the first two years, the next water year, October 1, 2004, to September 30, 2004, was also considered.

Once the optimal daily subsets were obtained for each day of the three periods of analysis, the statistical procedure explained earlier was used to find relevant statistics of each subset. The number of different optimal combinations or subsets of stations (with *RMSE* $<\cong$ 0.1 feet), was 45 (in 302 days) in the first period, 72 (in 343 days) in the second period, and 87 (in 317 days) in the third period.

The 10 most frequent combinations of stations for the three periods are shown in Table 5.6 to 5.8. In these tables, the first column shows *r* (the number of stations in the combination), the second column is the number of occurrences of that combination or subset, and the third and fourth columns show the mean (in feet) and standard deviation (in feet) of the root mean square error, respectively for each combination. Finally, the fifth column shows the stations that take part in each combination (identified with their SFWMD names). Stations labeled such as S65C_H_S65C_T are, in fact, two stations that share the same coordinates.

r	*Freq*	*RMSE Avg*	*RMSE StDev*	STATIONS
9	49	0.0644	0.0142	FTKISS,KRBNS,KRDRS,PC11R,PC33,PC34,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR2_H_WEIR2_T
8	40	0.0665	0.0146	FTKISS,KRBNS,KRDRS,PC11R,PC34,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR2_H_WEIR2_T
9	19	0.0876	0.0104	FTKISS,KRDRS,PC11R,PC33,PC44,PC52,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR3_H_WEIR3_T
12	16	0.0609	0.0315	FTKISS,KRBNS,KRDRS,PC11R,PC33,PC34,PC44,PC45,PC52,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR3_H_WEIR3_T
9	15	0.0578	0.0166	FTKISS,KRBNS,KRDRS,PC11R,PC33,PC34,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR3_H_WEIR3_T
11	13	0.0652	0.0126	FTKISS,KRBNS,KRDRS,PC11R,PC34,PC45,PC52,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR2_H_WEIR2_T,WEIR3_H_WEIR3_T
9	10	0.0599	0.0108	FTKISS,KRBNS,KRDRS,PC11R,PC33,PC52,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR3_H_WEIR3_T
12	10	0.0737	0.0103	FTKISS,KRBNS,KRDRS,PC11R,PC33,PC34,PC44,PC52,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR2_H_WEIR2_T,WEIR3_H_WEIR3_T
10	9	0.0745	0.0039	FTKISS,KRBNS,KRDRS,PC11R,PC34,PC45,PC52,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR2_H_WEIR2_T
11	8	0.0631	0.0201	FTKISS,KRBNS,KRDRS,PC11R,PC33,PC34,PC45,PC52,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR3_H_WEIR3_T

Table 5.6: The ten most frequent combinations or subsets of stations in the subbasin S-65BC with *RMSE* <≅ 0.1 for Oct 2001 - Sep 2002.

r	*Freq*	*RMSE Avg*	*RMSE StDev*	STATIONS
9	33	0.0717	0.0155	FTKISS,KRBNS,PC11R,PC33,PC44,PC52,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR2_H_WEIR2_T
10	30	0.0895	0.0075	FTKISS,KRBNS,KRDRS,PC11R,PC33,PC44,PC52,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR2_H_WEIR2_T,,
9	18	0.0697	0.0263	FTKISS,KRBNS,KRDRS,PC11R,PC33,PC34,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR3_H_WEIR3_T
8	15	0.0695	0.0161	FTKISS,KRBNS,KRDRS,PC11R,PC34,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR2_H_WEIR2_T
9	15	0.0917	0.0061	FTKISS,KRBNS,KRDRS,PC11R,PC33,PC44,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR2_H_WEIR2_T
8	13	0.0894	0.0083	FTKISS,KRBNS,KRDRS,PC11R,PC34,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR3_H_WEIR3_T
10	12	0.0836	0.0096	FTKISS,KRBNS,KRDRS,PC11R,PC34,PC44,PC52,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR2_H_WEIR2_T
10	12	0.0680	0.0213	FTKISS,KRBNS,KRDRS,PC11R,PC34,PC45,PC52,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR3_H_WEIR3_T
9	11	0.0787	0.0128	FTKISS,PC11R,PC34,PC44,PC45,PC52,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR2_H_WEIR2_T
11	10	0.0764	0.0061	FTKISS,KRBNS,KRDRS,PC11R,PC33,PC34,PC45,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR2_H_WEIR2_T,WEIR3_H_WEIR3_T

Table 5.7: The ten most frequent combinations or subsets of stations in the subbasin S-65BC with *RMSE* <≅ 0.1 for Oct 2002 - Sep 2003.

r	*Freq*	*RMSE Avg*	*RMSE StDev*	STATIONS
11	30	0.0660	0.0116	AVONP4,KRBNS,KRDRS,PC11R,PC34,PC44,PC52,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR2_H_WEIR2_T,WEIR3_H_WEIR3_T
10	15	0.0822	0.0112	AVONP4,FTKISS,KRBNS,KRDRS,PC11R,PC34,PC44,S65C_H_S65C_T,WEIR1_H,WEIR3_H
11	14	0.0668	0.0173	AVONP4,KRBNS,KRDRS,PC11R,PC34,PC44,PC52,S65C_H_S65C_T,WEIR1_H,WEIR2_H,WEIR3_H
8	13	0.0905	0.0052	AVONP4,FTKISS,KRBNS,KRDRS,PC11R,PC34,PC44,S65C_H_S65C_T
11	13	0.0664	0.0128	AVONP4,FTKISS,KRDRS,PC11R,PC33,PC44,PC52,S65C_H_S65C_T,WEIR1_H,WEIR2_H,WEIR3_H
9	11	0.0822	0.0106	AVONP4,KRBNS,PC11R,PC34,PC45,PC52,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR2_H_WEIR2_T
11	10	0.0821	0.0058	AVONP4,FTKISS,KRBNS,KRDRS,PC11R,PC34,PC44,PC52,S65C_H_S65C_T,WEIR1_H,WEIR3_H
11	9	0.0497	0.0255	AVONP4,KRBNS,KRDRS,PC11R,PC34,PC45,PC52,S65C_H_S65C_T,WEIR1_H_WEIR1_T,WEIR2_H_WEIR2_T,WEIR3_H_WEIR3_T
10	8	0.0863	0.0101	AVONP4,KRBNS,PC11R,PC34,PC44,PC52,S65C_H_S65C_T,WEIR1_H,WEIR2_H,WEIR3_H
12	8	0.0356	0.0284	AVONP4,FTKISS,KRBNS,KRDRS,PC11R,PC33,PC34,PC45,S65C_H_S65C_T,WEIR1_H,WEIR2_H,WEIR3_H

Table 5.8: The ten most frequent combinations or subsets of stations in the subbasin S-65BC with *RMSE* $<\cong 0.1$ for Oct 2003 - Sep 2004.

The 10 most frequent combinations for the first, second and third periods occur in 189, 169 and 131 days, respectively. The five most frequent combinations for the first, second, and third periods occur in 139 of 189, 111 of 169 and 85 of 131 days, respectively. *Only two combinations are present simultaneously in the 10 most frequent combinations and in the five most frequent combinations of the first two periods. They are*:

- FTKISS, KRBNS, KRDRS, PC11R, PC33, PC34, S65C_H_S65C_T, WEIR1_H_WEIR1_T, WEIR3_H_WEIR3_T. 9 stations. First period: 15

days, average *RMSE* = 0.0578, StDev *RMSE* = 0.0166. Second period: 18 days, average *RMSE* = 0.0697, StDev *RMSE* = 0.0263.

- FTKISS, KRBNS, KRDRS, PC11R, PC34, S65C_H_S65C_T, WEIR1_H_WEIR1_T, WEIR2_H_WEIR2_T. 8 stations. First period: 40 days, average *RMSE* = 0.0665, StDev *RMSE* = 0.0146. Second period: 15 days, average *RMSE* = 0.0695, StDev *RMSE* = 0.0161.

No single subset or combination is repeated in each of the three-year's 10 most frequent combinations. Therefore, it is difficult to select the combination or subset of stations that can be considered the optimal subset of stations in the subbasin S-65BC because the most frequent combinations of one of the periods are not the most frequent combinations of the other periods. It seems that the combination that is optimal for a period depends, apart of the behavior of the water cycle in the period, *on whether each station is in operation* and, in general, on the operation of the water management district. For example, the AVONP4 station begun its operation on 08/21/2003 and appears in 39 optimal combinations of the last 41 days of the second period. In addition, it appears in all of the 10 most frequent optimal combinations and in 315 of 317 of the optimal combinations of the third period.

5.4 Optimization of the Lake Kissimmee Stage-Monitoring Network

Once the results of the first methodology were analyzed, an additional statistical procedure was generated, included in the second methodology and then applied to case study 4 (presented in Chapter 3) (Figure 5.2).

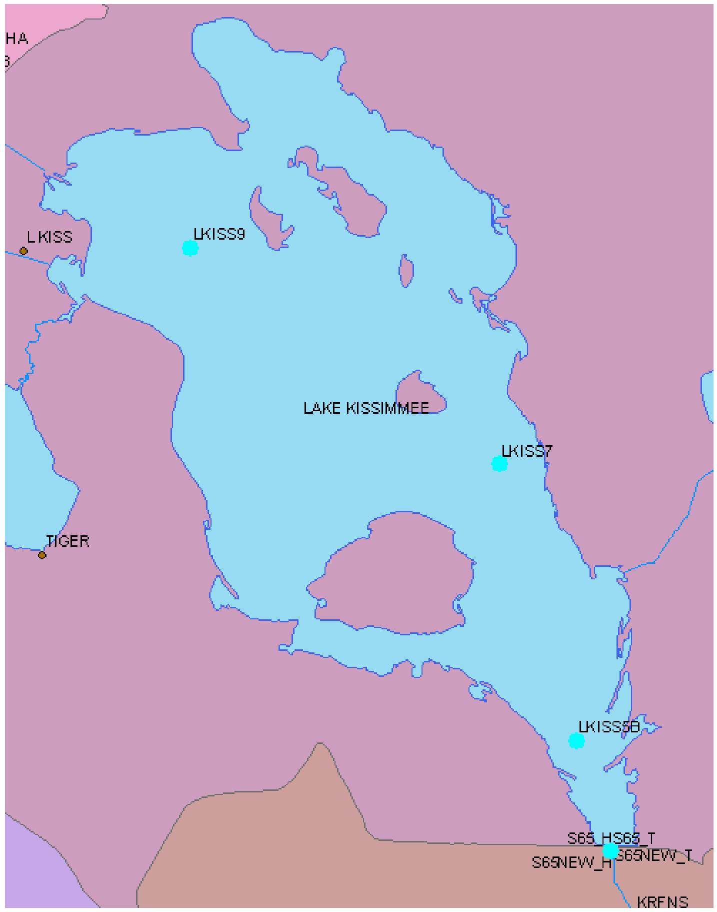


Figure 5.2: Case study 4. Lake Kissimmee. Selected stations in Lake Kissimmee (LKISS9, LKISS7, LKISS5B, and S65_H).

Applying the optimization by range of sizes option, the statistics of daily optimal subsets process with an admissible *RMSE* of 0.1 ft and the statistics of daily optimal subsets for the entire period to this set of stations, the results shown in Table 5.9 and Figure 5.3 were found.

	Daily Optimal Subset	F_r	$P(RMSE \leq \text{adm } RMSE)$	rank
1	LKISS5B	0.4356	0.4868	3
2	LKISS9	0.2890	0.3040	
3	S65_H	0.3685	0.4244	
4	LKISS7, LKISS9	0.3781	0.3883	
5	LKISS5B, LKISS9	0.4096	0.3741	
6	LKISS7, S65_H	0.3438	0.3875	
7	LKISS5B, LKISS7, S65_H	0.3233	0.3087	
8	LKISS7, LKISS9, S65_H	0.8890	0.8648	1
9	LKISS5B, LKISS7, LKISS9	0.7973	0.7609	2

Table 5.9: Lake Kissimmee's daily optimal subsets of stations. The optimal subset of the period is LKISS7, LKISS9, S65_H.

In Table 5.9 are shown the nine daily optimal subsets identified with the statistics of daily optimal subsets process. Once these subsets were identified, the relative frequency of all subsets and their corresponding $P(RMSE \leq \text{admissible } RMSE)$ were computed and then ranked accordingly to the values of the probability, one for the largest probability, two for the second largest probability, and so on. The values of relative frequency and probability are similar for each subset. The root mean square error of F_r compared with the theoretical probability distribution is 0.0355. One can conclude without further analysis that the fit of the relative frequency to the theoretical probability distribution is good and that the optimal subset is LKISS7, LKISS9, S65_H. This subset reports 86.5 % of the time an *RMSE* smaller than or equal to 0.1 ft.

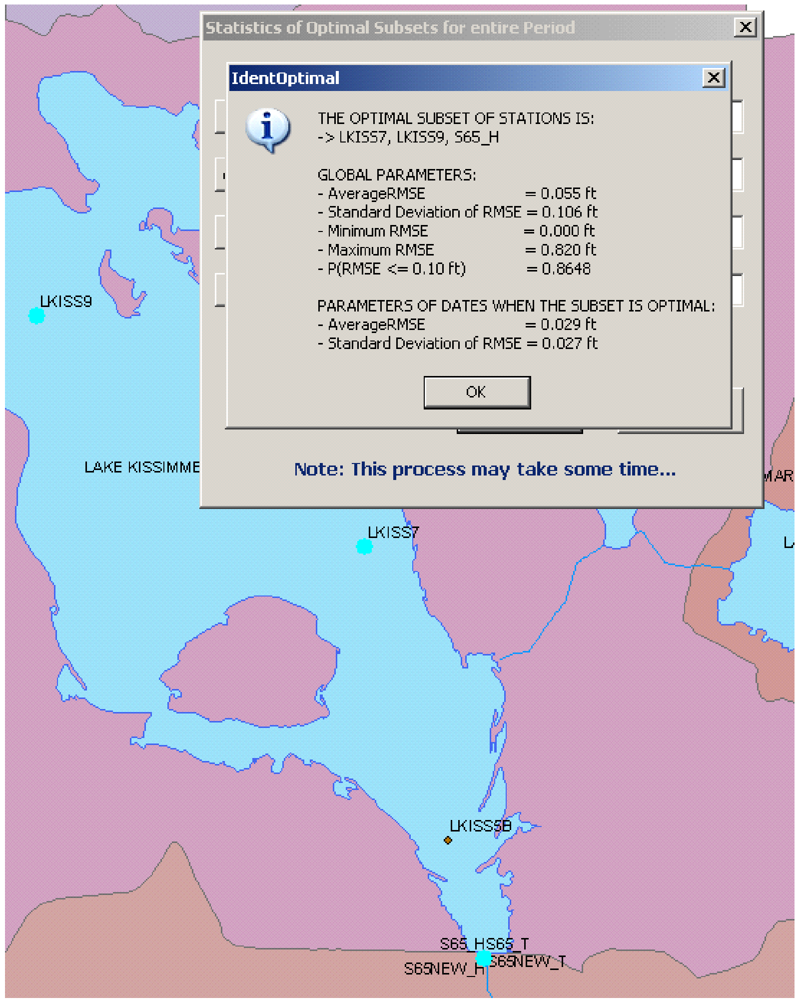


Figure 5.3: The optimal subset of stations of Lake Kissimmee is LKISS7, LKISS9 and S65_H.

Finally, in Figure 5.3 are shown the results that the GIS Tools developed for this purpose present. The application of the methodology for optimizing stations in lakes for this new case study was satisfactory. The optimal subset of stations was identified and a modification of current steps was not needed.

5.5 CONCLUDING REMARKS

The results of the application of the successive steps of the methodology explained in Chapter 3 lead to the refining of the optimization methodology for stations in lakes. The initial results lead to the adoption of the genetic algorithm as the optimization method. The results of the first optimization methodology applied to the Pool BC stations (case study 3) made clear that streams and lakes need different methodologies and that an additional statistical step was needed to identify the optimal subset of stations in lakes. The application of the second methodology, which includes the additional statistical step, to the network of Lake Kissimmee (case study 4) was satisfactory. It was able to identify the optimal subset.

Chapter 6: Optimization of Stage-Monitoring Networks in Streams (Steady flow)

6.1 INTRODUCTION

In this chapter, the main results of the development of the methodology for optimizing stage-monitoring networks in streams with steady flow are presented. First, the results of the application of the first methodology to the optimization of the main stream of the Pool AE stage-monitoring network (case study 5) are presented. The analysis of the previous results led to the proposition of the second methodology. Subsequently, the results of the application of the second methodology to the same network are presented.

6.2 FIRST OPTIMIZATION OF THE MAIN STREAM OF POOL AE STAGE-MOTORING NETWORK

The first methodology was applied to the stage-monitoring network of the main stream, of Pool AE (Figure 6.1), already described in Chapter 3. This initial optimization experiment followed the next steps: creation of geometry file, calibration with a limited number of datasets, calibration with data of the entire period and estimation of water elevations in selected stations. Results of each step are presented in the following paragraphs.

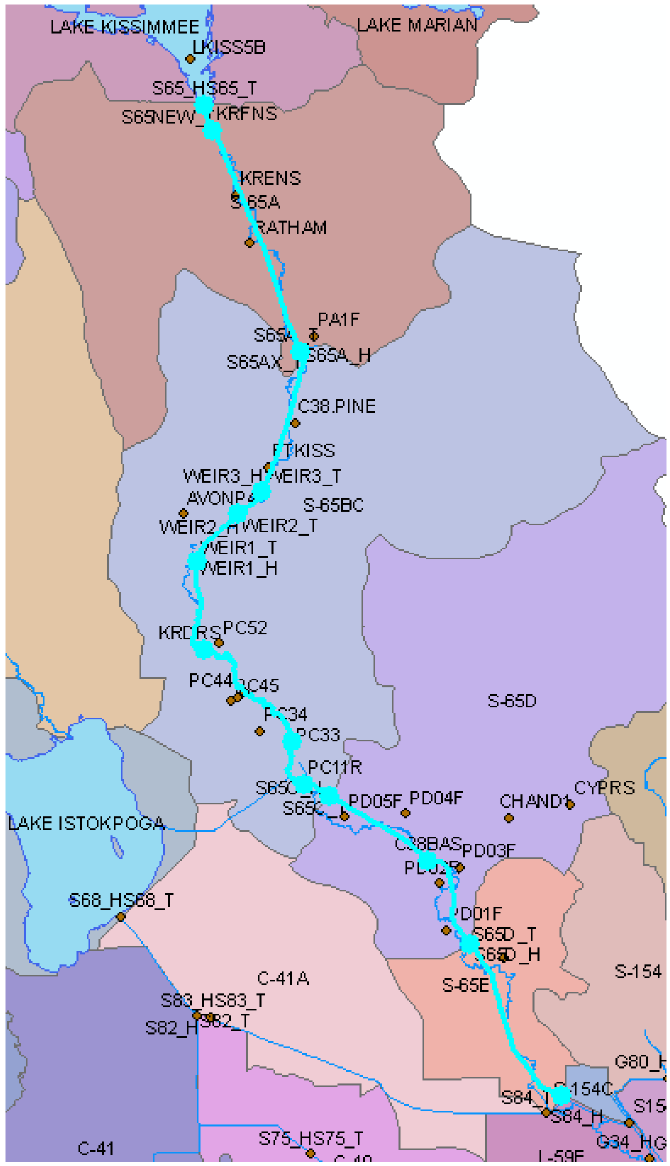


Figure 6.1: Case study 5. Stage monitoring stations along the main stream of Pool AE

Creation of geometry file. Following the procedure described in Chapter 3, the geometry of the stream shown in Figure 6.2 was obtained. A typical section of the naturalized reach is also shown in Figure 6.3.

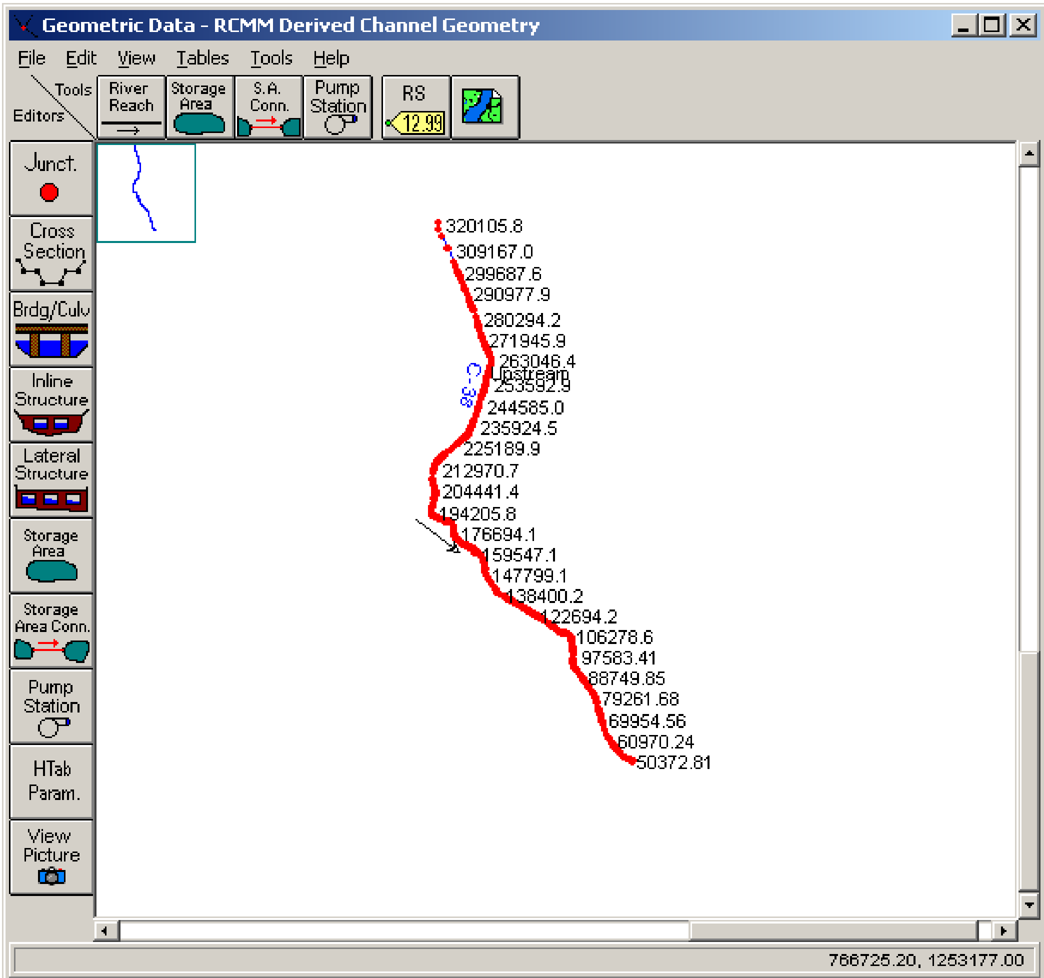


Figure 6.2: Pool AE main stream represented inside HEC-RAS.

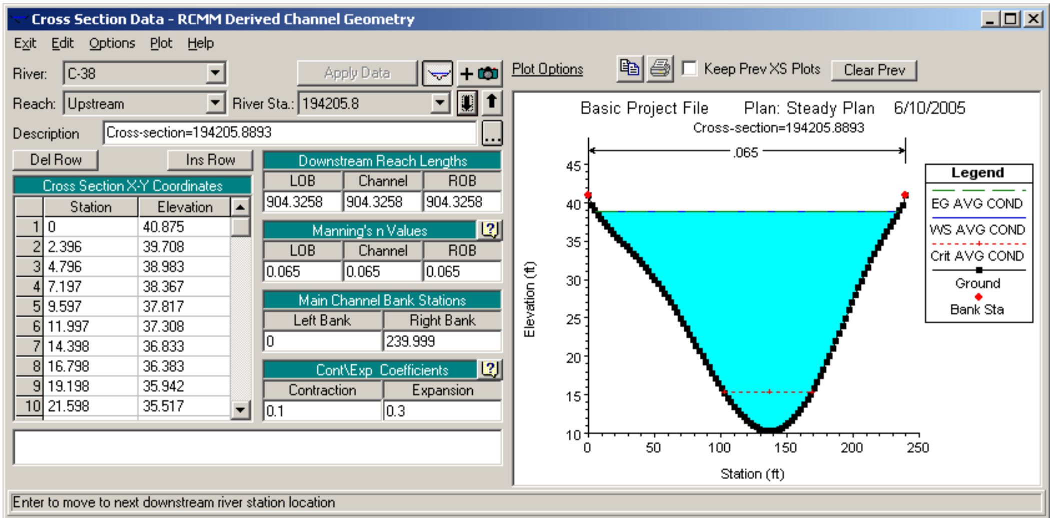


Figure 6.3: A Pool AE main stream section represented inside HEC-RAS.

Calibration with a limited number of datasets. The reach between stations WEIR1_H and S65C_H that includes the Kissimmee River was the most difficult to calibrate, as it has the biggest difference of water elevation between its initial and final cross sections. In addition, when discharge varies, its water elevations show the greatest variation. The first calibration results were those of the average conditions, a Manning's n for the channel's sections equal to 0.015 and the corresponding n for the naturalized sections (Kissimmee River) equal to 0.065 were found. Then the roughness factors necessary for flow conditions different from the average conditions were found with the data of the other seven days. As an example, measured and estimated water elevations for the average conditions are summarized in Table 6.1. The data in Table 6.1 is also represented in Figure 3.12. The root mean square error for this data is 0.24 ft. As another example, the profile for Oct 1, 2001, is shown in Figure 6.4.

Station	Section (ft)	Main channel distance (ft)	Observed water elev. (ft)	Computed water elev. (ft)
S65_T	320105.8	272141.54	46.36	46.31
KRFNS	314076.8	266112.54	46.38	46.31
S65A_H	263639.2	215674.94	46.29	46.29
S65A_T	263634.2	215669.94	39.89	39.72
WEIR3_H	232169.0	184204.74	39.72	39.72
WEIR3_T	232164.0	184199.74	39.63	39.64
WEIR2_H	225189.9	177225.64	39.64	39.64
WEIR2_T	225184.9	177220.64	39.62	39.39
WEIR1_H	211725.5	163761.24	39.39	39.39
WEIR1_T	211720.5	163756.24	39.31	38.95
KRDRS	188833.2	140868.94	37.96	38.26
KRBNS	171777.7	123813.44	37.08	37.24
PC33	149107.5	101143.24	35.57	35.34
PC11R	139678.6	91714.34	34.98	34.92
S65C_H	134404.5	86440.24	34.92	34.92
S65C_T	134399.5	86435.24	26.86	26.64
C38BAS	110014.5	62050.24	26.91	26.63
S65D_H	87366.08	39401.82	26.63	26.63
S65D_T	87361.08	39396.82	20.6	21.09
S65E_T	47964.26	0.00	21.09	21.09

Table 6.1: Average conditions' water elevation observed and computed values.

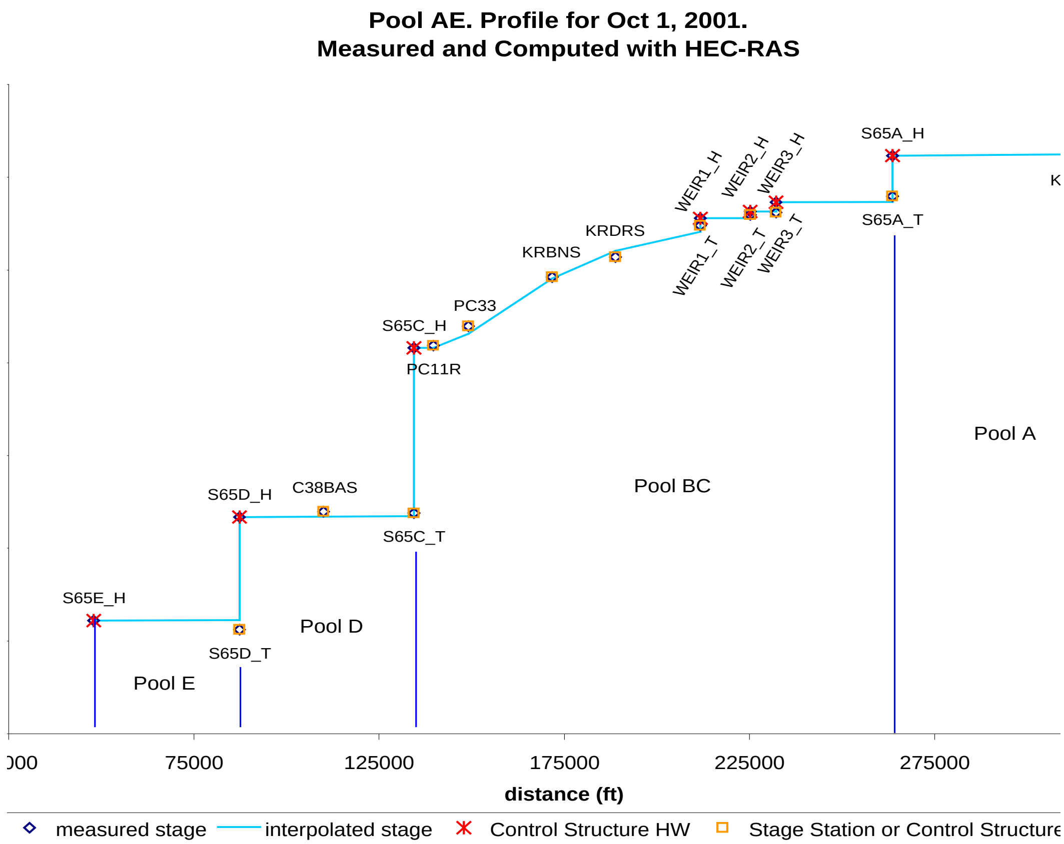


Figure 6.4: Profile for Oct 1, 2001. Stations along the stream are indicated.

Calibration with data of the entire period. A summary of the properties of the root mean square errors for the calibration with data of the entire period for the six calibration runs described in Chapter 3 is shown in Table 6.2.

	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6
Max	2.58	2.58	2.58	2.58	2.58	2.58
Min	0.21	0.21	0.21	0.21	0.21	0.21
Average	0.55	0.47	0.44	0.44	0.43	0.43
Std. Dev.	0.43	0.32	0.31	0.31	0.31	0.31

Table 6.2: Summary of the calibration process with the original subset of seven base stations and data of the entire period (Daily *RMSE* in ft)

Estimation of water elevations in selected stations. The results of the last calibration run are valid for the estimation of water elevations for the case of the subset constituted by the seven stations at the head side of the hydraulic structures of Pool AE. To form the base subset with eight stations, station WEIR1_T was added because it has 428 days of data and the maximum *RMSESt* of the seven base stations' run. Then, to create the nine base stations set, station KRNBS was selected because it has data on 428 days and has the maximum *RMSESt* in the run of eight base stations. The resultant *RMSESt* values of all the stations of the three subsets are shown in Tables 6.3. Table 6.4 shows a summary of the statistical properties of the *RMSE* values of the three studied subsets. Table 6.5 shows the probability of no exceedence of *RMSE* for the same three subsets. The probability of no exceedence of *RMSE* shown in Table 6.5 is plotted in Figure 6.5. This figure shows how the root mean square error diminishes as more base stations are used.

no.	Station Name	Section	number of.days	Subsets of 7 stations *RMSESt* (ft)	8 stations *RMSESt* (ft)	9 stations *RMSESt* (ft)
1	S65_T	320105.8	428	0.07	0.07	0.07
2	KRFNS	314076.8	419	0.08	0.08	0.08
3	S65A_H	263639.2	428			
4	S65A_T	263634.2	428	0.21	0.21	0.21
5	WEIR3_H	232169	428			
6	WEIR3_T	232164	428	0.05	0.05	0.05
7	WEIR2_H	225189.9	428			
8	WEIR2_T	225184.9	428	0.22	0.22	0.22
9	WEIR1_H	211725.5	428			
10	WEIR1_T	211720.5	428	1.24		
11	KRDRS	188833.2	419	0.90	0.90	0.43
12	KRBNS	171777.7	428	0.65	0.65	
13	PC33	149107.5	353	0.71	0.71	0.71
14	PC11R	139678.6	428	0.14	0.14	0.14
15	S65C_H	134404.5	428			
16	S65C_T	134399.5	428	0.21	0.21	0.21
17	C38BAS	110014.5	428	0.26	0.26	0.26
18	S65D_H	87366.08	428			
19	S65D_T	87361.08	427	0.49	0.49	0.49
20	S65E_H	47964.26	428			

Table 6.3: *RMSESt* values of each station for the three studied subsets. Highlighted stations are the initial base stations.

	Subsets of 7 stations *RMSE* (ft)	8 stations *RMSE* (ft)	9 stations *RMSE* (ft)
Max	2.58	1.91	0.80
Min	0.21	0.18	0.16

Average	0.43	0.36	0.29
Std. Dev.	0.31	0.22	0.11

Table 6.4: Statistical properties of *RMSE* for the entire period.

rmse (ft)	$P(RMSE \leq rmse)$ for subsets of		
	7 stations	8 stations	9 stations
0.15	0.000	0.000	0.000
0.2	0.000	0.077	0.157
0.25	0.084	0.329	0.514
0.3	0.318	0.537	0.624
0.35	0.458	0.643	0.769
0.4	0.593	0.715	0.834
0.45	0.706	0.801	0.895
0.5	0.811	0.874	0.944
0.55	0.853	0.909	0.956
0.6	0.900	0.953	0.974
0.65	0.930	0.970	0.984
0.7	0.944	0.972	0.993
0.75	0.949	0.972	0.995
0.8	0.956	0.972	1.000
0.85	0.963	0.974	1.000
0.9	0.967	0.974	1.000
0.95	0.970	0.974	1.000
1	0.972	0.974	1.000
1.5	0.974	0.995	1.000

Table 6.5: Probability of no exceedence of *RMSE* of the three subsets for the entire period.

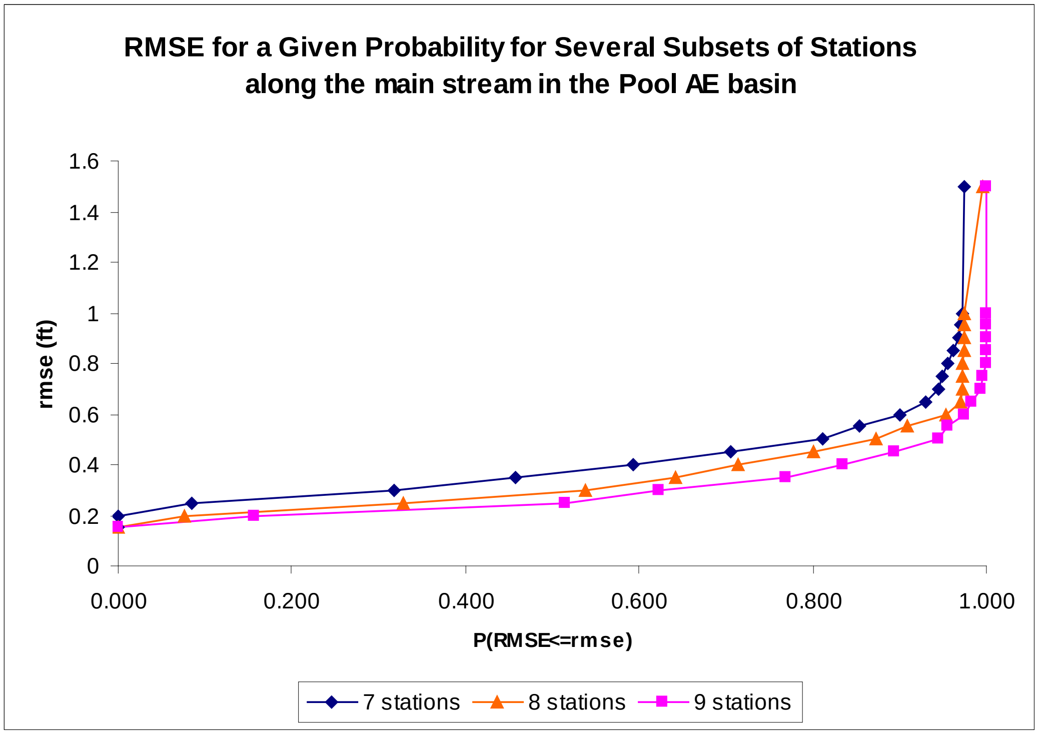


Figure 6.5: Probability of no exceedence of *RMSE* for the subsets of seven, eight and nine base stations in Pool AE.

From Tables 6.3 to 6.5 and Figure 6.5, one can conclude that if the number r of stations in the base subset increases, the average errors, in the remaining $n - r$ stations tends to decrease. The optimization process could be stopped when a satisfactory distribution of *RMSE* is reached. To show how error varies in a given station, a plot of error versus probability of no exceedence can be plotted. For the case of the subset of nine base stations, the distribution of errors for some of the stations was plotted in Figures 6.6 to 6.16. The first optimization of the stations of the case study 5, the mainstream of Pool AE basin, shows that the optimization of the stage-monitoring network along streams is feasible.

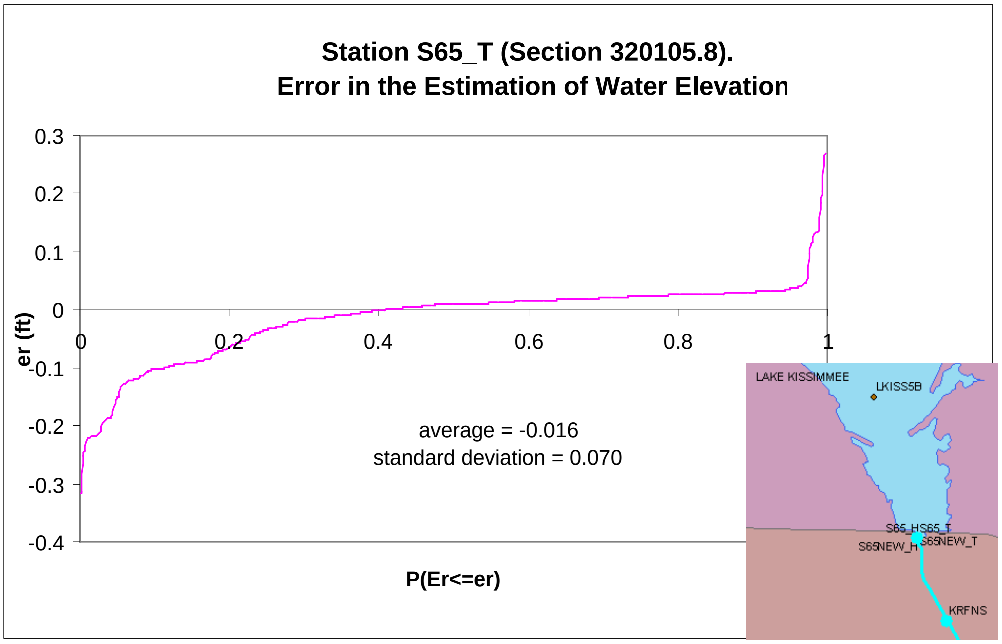


Figure 6.6: Station S65_T's error distribution.

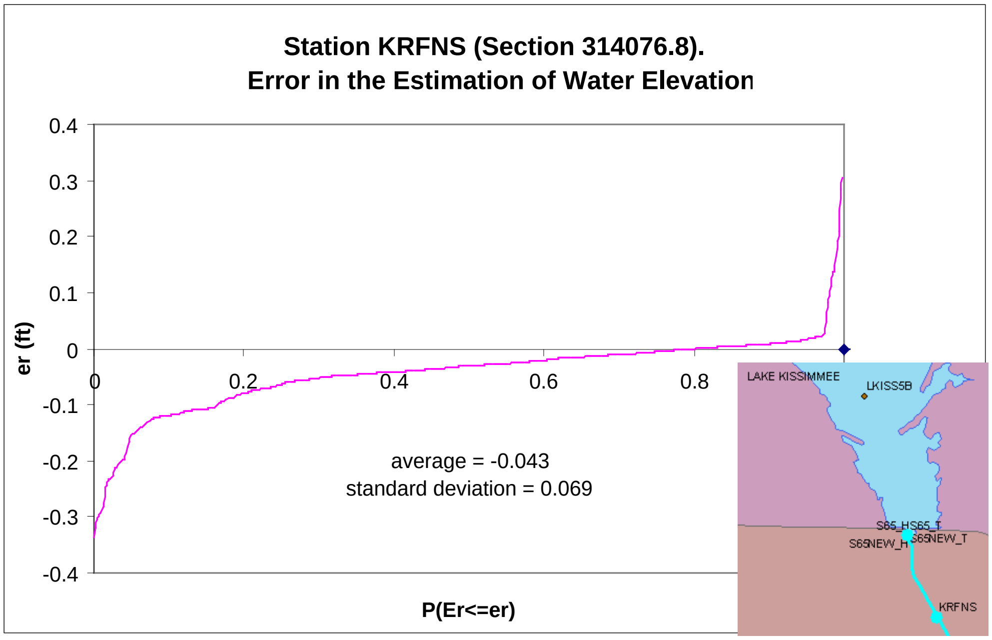


Figure 6.7: Station KRFNS's error distribution.

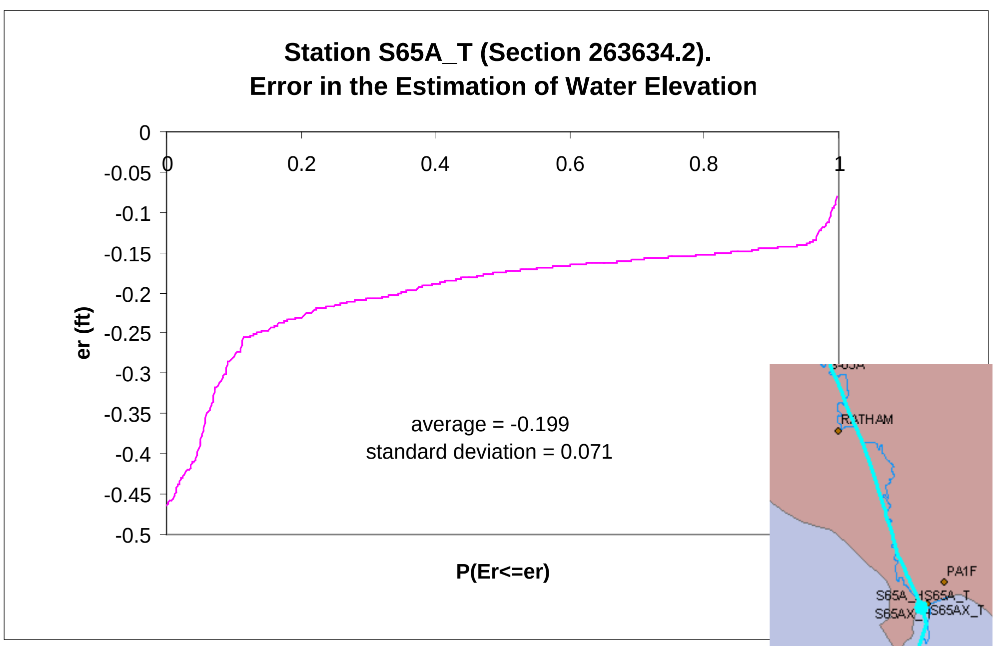


Figure 6.8: Station S65A_T's error distribution.

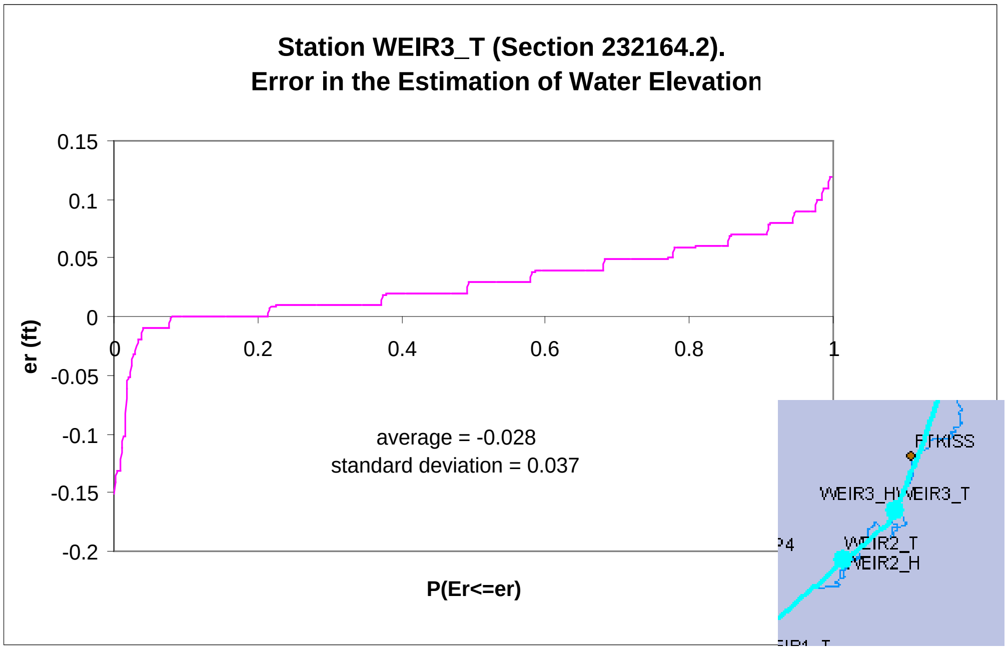


Figure 6.9: Station WEIR3_T's error distribution.

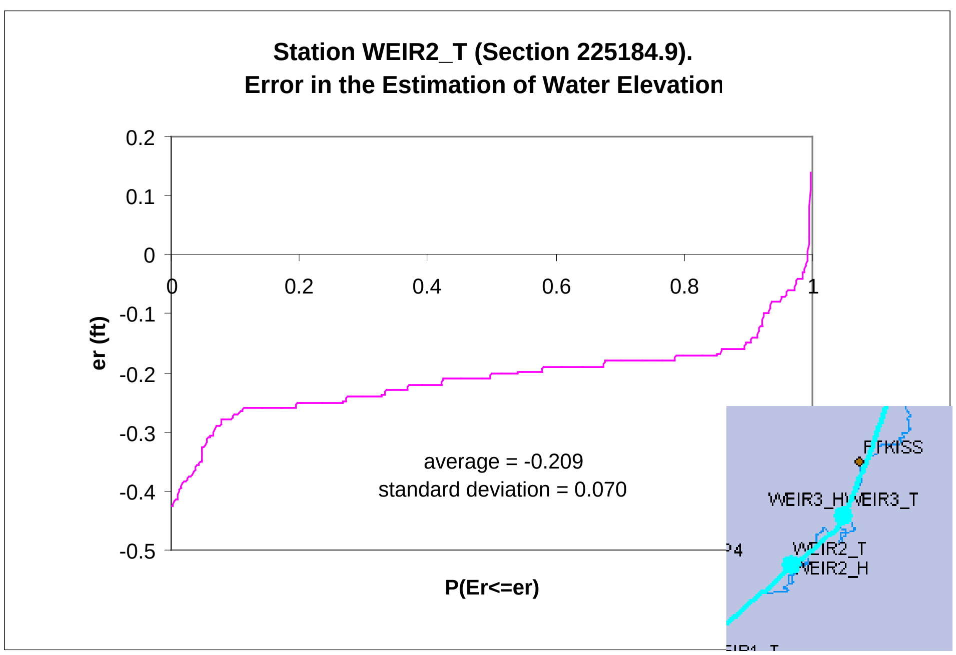


Figure 6.10: Station WEIR2_T's error distribution.

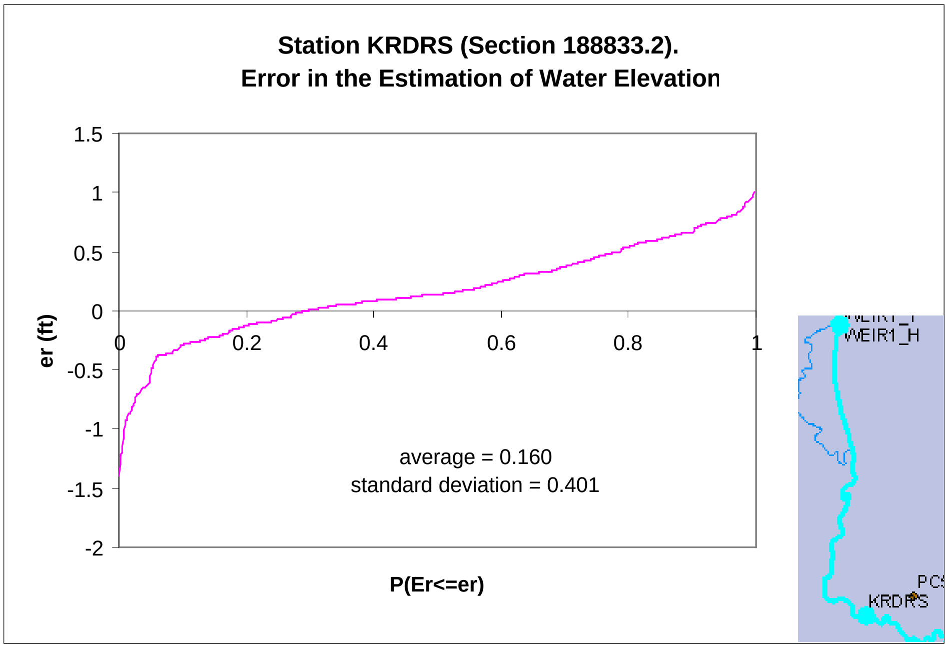


Figure 6.11: Station KRDRS's error distribution.

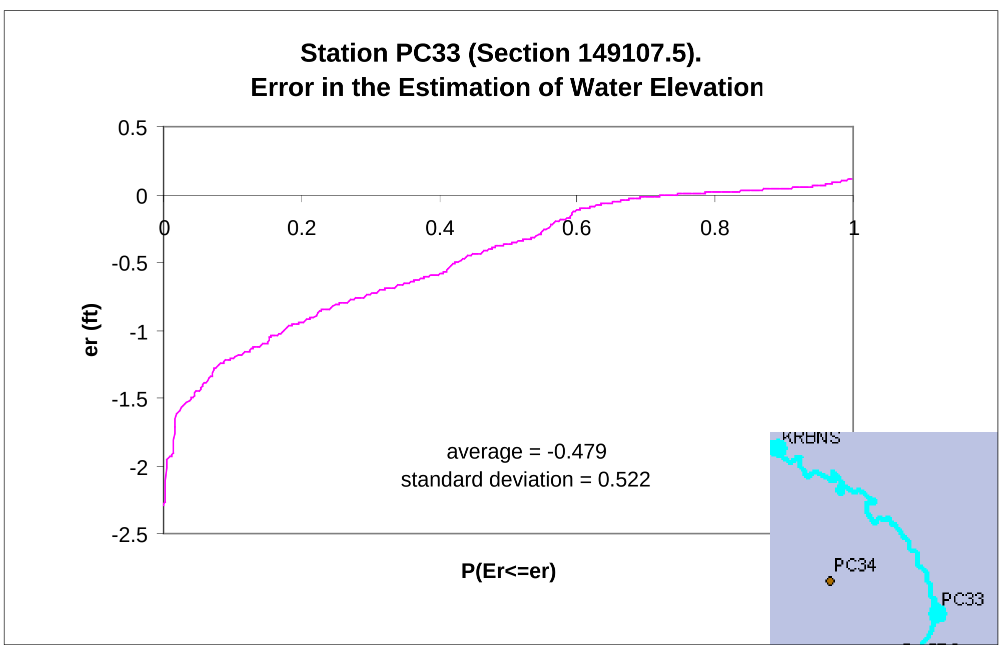


Figure 6.12: Station PC33's error distribution.

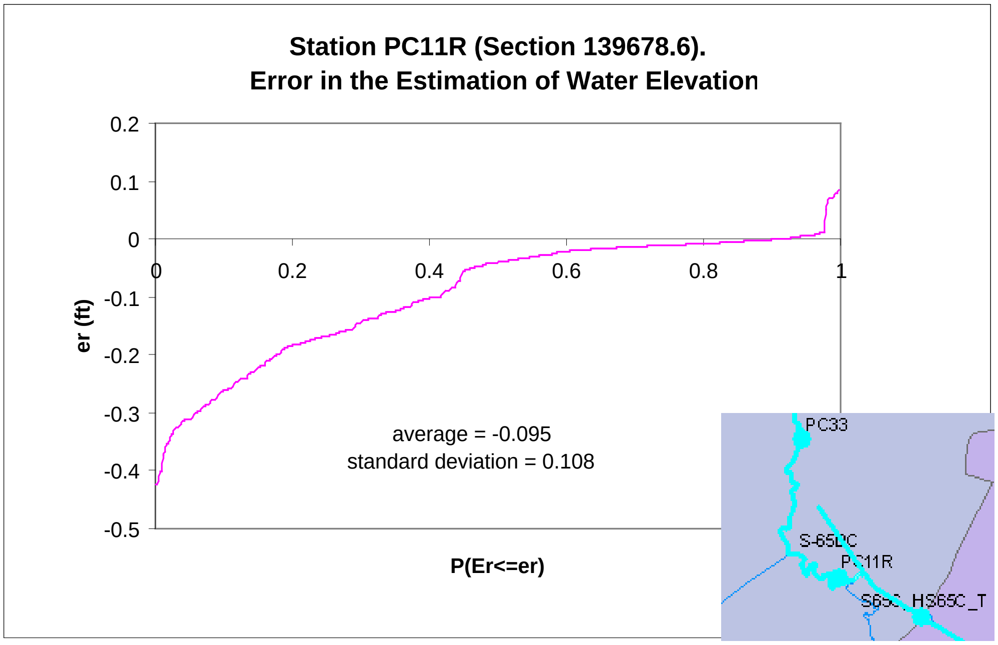


Figure 6.13: Station PC11R’s error distribution.

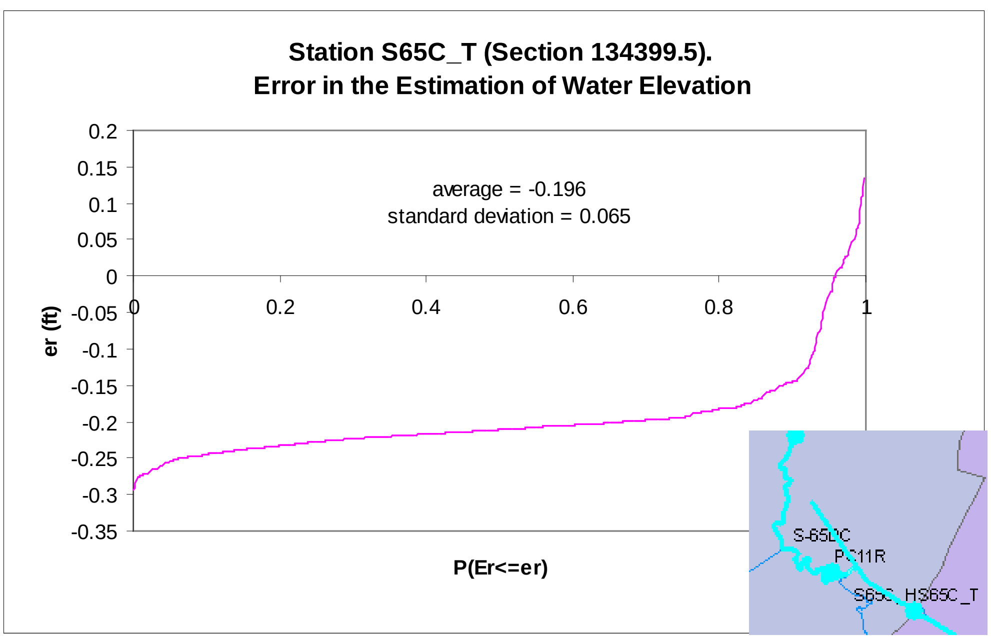


Figure 6.14: Station S65C_T’s error distribution.

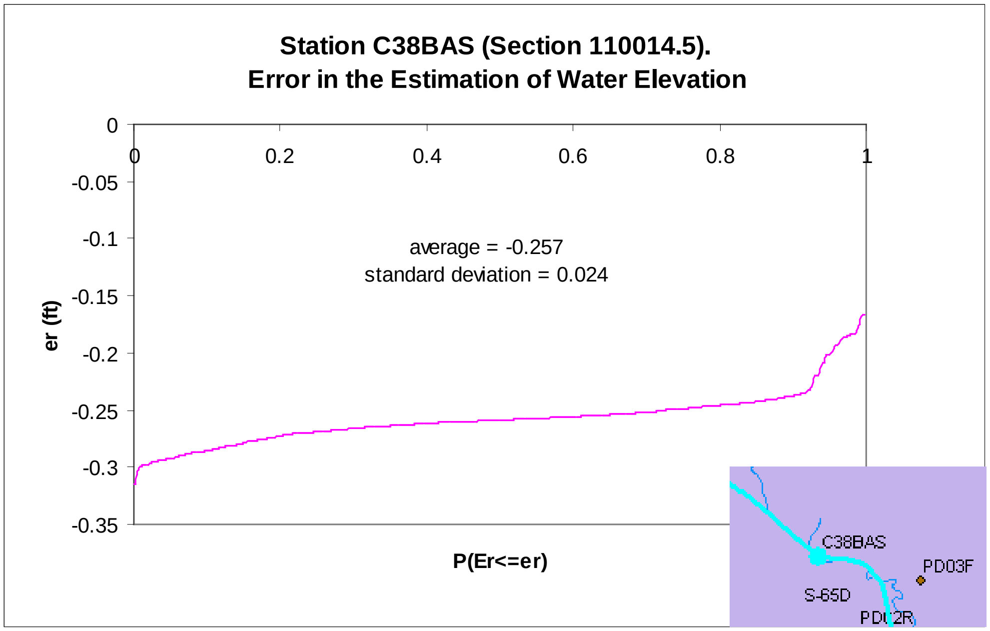


Figure 6.15: Station C38BAS's error distribution.

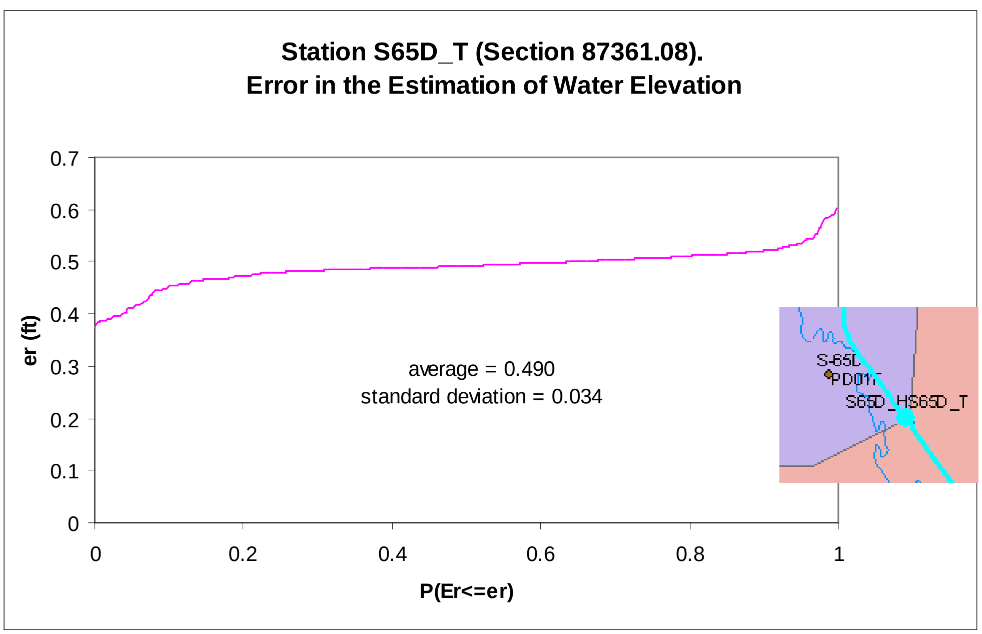


Figure 6.16: Station S65D_T's error distribution.

6.3 Second Optimization of the Stage-Monitoring Network in Pool AE Main Stream

An improvement of the results of the study of the case study 5, main stream of Pool AE, is presented here. This is achieved using the tools, developed for the second methodology, to optimize the network along streams. This description begins from the optimization step because all previous steps were already executed during the first optimization study.

The optimization of the network in streams was carried out using the NetOptStreams tool. The user interface of the NetOptStreams tool is shown in Figure 6.17. It is invoked by clicking on the NetOptStreams button of the Network Optimization in Streams toolbar. Details of the use of this tool are in the user manual (Appendix A).

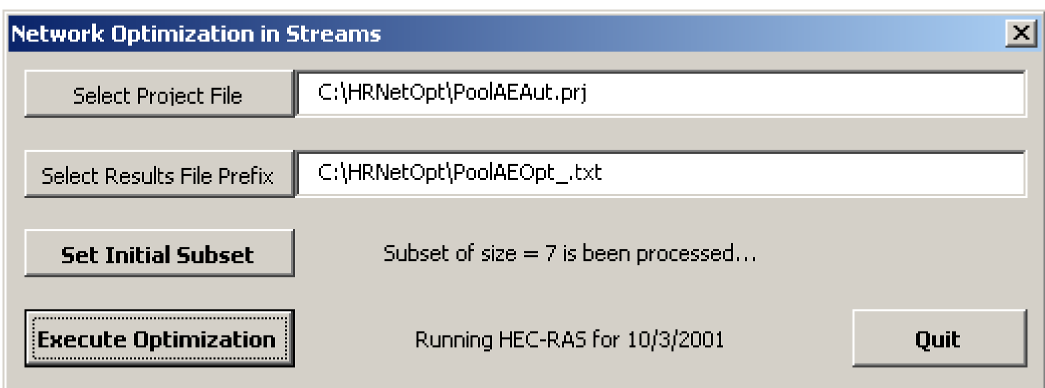


Figure 6.17: Network optimization in streams user interface.

Applying this tool to the stream of interest, the following general results were found. Table 6.6 shows the evolution of the maximum *RMSESt* for all the possible sizes of the base subset. The maximum *RMSESt* value signals the station that will take part in the base subset for the next cycle.

Number of stations in Subset	Station entering base subset (corresponding section)	Maximum *RMSESt* (ft)
7	WEIR1_T (211720.5)	1.240
8	KRDRS (188833.2)	0.901
9	PC33 (149107.5)	0.717
10	KRBNS (171777.7)	0.613
11	S65D_T (87361.08)	0.485
12	C38BAS (110014.5)	0.263
13	WEIR2_T (225184.9)	0.223
14	S65A_T (263634.2)	0.207
15	PC11R (139678.6)	0.152
16	KRFNS (314076.8)	0.080
17	S65C_T (134399.5)	0.071
18	S65_T (320105.8)	0.060
19	WEIR3_T (232164)	0.048

Table 6.6: Maximum *RMSESt* for base subsets' sizes from 7 to 19 stations.

Figure 6.18 shows the variation of *RMSESt* in each station for the base subsets considered. As expected, the *RMSESt* values are larger when the size of the base subset is small than when it is large.

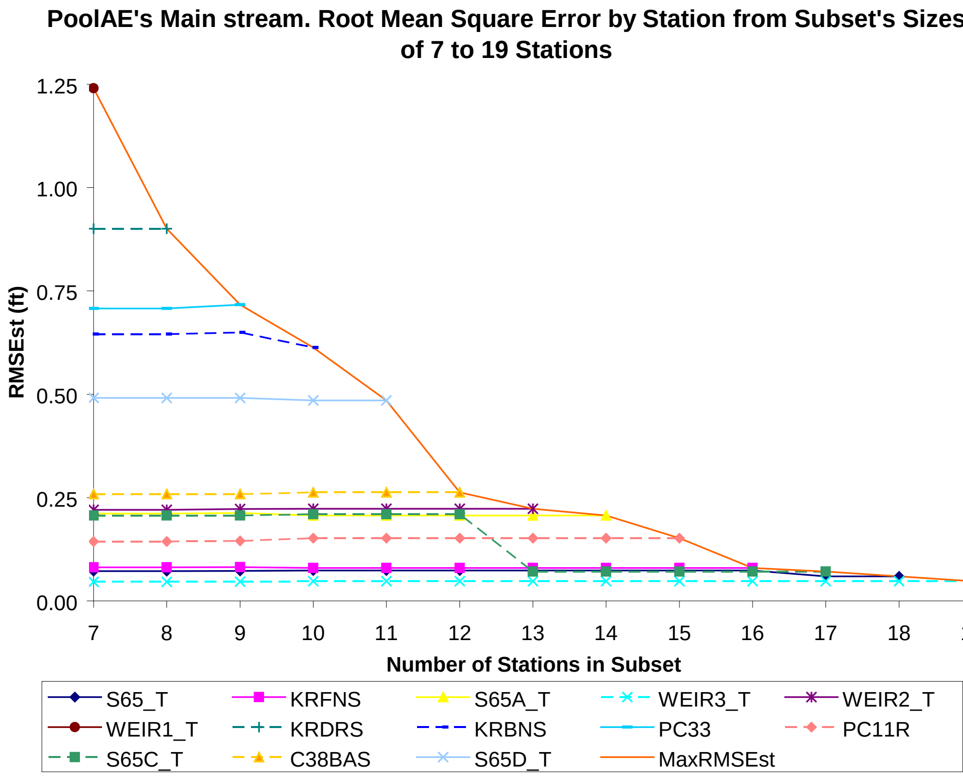


Figure 6.18: *RMSESt*'s evolution for the stations outside the base subset.

The orange line marks the maximum *RMSESt* for each size of the base subset. Another aspect of the results is the evolution of the main statistical parameters of *RMSE* for the thirteen base subsets considered. In Table 6.7, the average and standard deviation of *RMSE* are shown. The average values decrease consistently while the standard deviation values first decrease, then increase, and finally decrease again.

Size of base subset	Average *RMSE* (ft)	Standard Deviation of *RMSE* (ft)
7	0.434	0.306
8	0.356	0.224

9	0.311	0.157
10	0.280	0.084
11	0.230	0.020
12	0.171	0.032
13	0.130	0.047
14	0.109	0.048
15	0.075	0.053
16	0.058	0.038
17	0.053	0.030
18	0.045	0.030
19	0.038	0.029

Table 6.7: Average and standard deviation of *RMSE* for the base subsets considered.

Figure 6.19 provides a graphical view of the *RMSE's* empirical probability distributions obtained in the optimization process. One can see that as the size of the base subset grows the values of *RMSE* falls closer and closer to zero. While the maximum value of *RMSE* for subsets of size 7 to 10 is greater than 0.60 ft, the same value for subsets with at last 11 stations is less than 0.30 ft. The minimum *RMSE* goes from 0.21 ft (subset of 7 stations) to 0.00 ft (subset of 19 stations). As in the initial optimization study a set of graphs of the distribution of errors at given stations (and size of base subsets) could also be plotted. They would have the same aspect as those presented in the previous section.

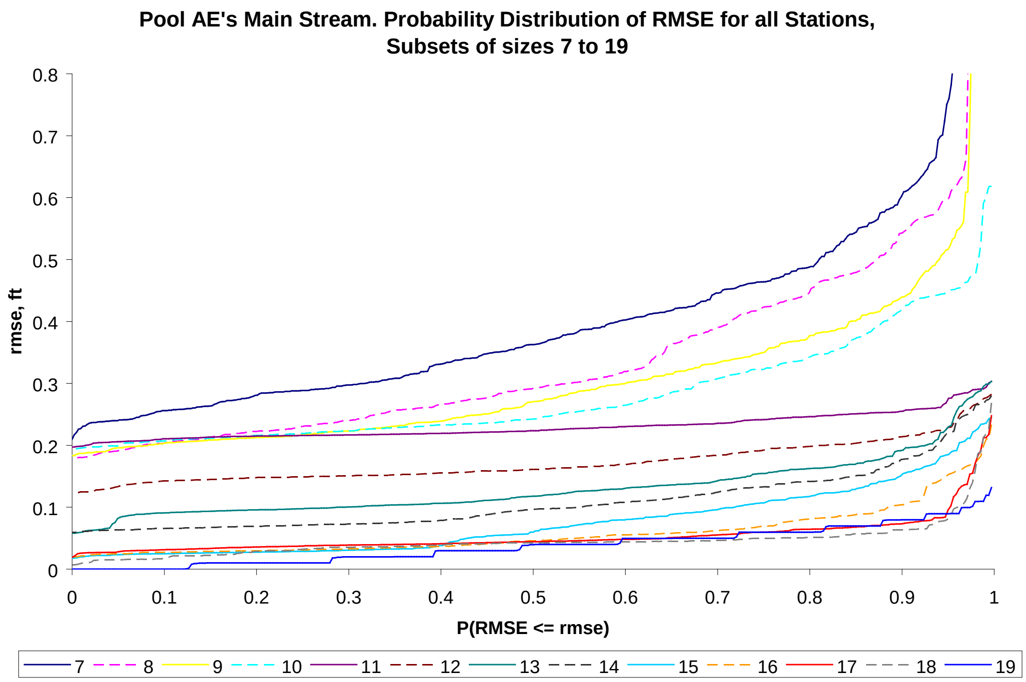


Figure 6.19: Probability distribution of *RMSE* for the base subsets considered.

It was found that running HEC-RAS on a channel geometry analytically obtained gives adequate assessment of water surface elevation in this system. Therefore, the application of the toolset for optimizing stations in streams for the main stream of Pool AE was satisfactory. The base subsets of stations for different sizes were identified; specifically the first base subset can be set with the rule, valid for subcritical state, “as a first base subset, use all the headwater gages at the hydraulic structures inside the system”. The base subset (S65A_H, WEIR3_H, WEIR2_H, WEIR1_H, S65C_H, S65D_H, S65E_H, WEIR1_T, KRDRS, PC33, KRBNS, and S65D_T) of 12 stations with average *RMSE* equal to 0.171 ft, standard deviation of *RMSE* equal to 0.032 ft and *RMSESt* maximum equal to 0.263 ft seems to be a good candidate for consideration as the optimal base subset. The empirical probability *P*(*RMSE* ≤ admissible *RMSE*) shown in Figure 6.19 for the base subset of 12 stations implies the need of an admissible *RMSE* larger than 0.12 ft. For example, the admissible *RMSE* for a *P*(*RMSE* ≤ admissible *RMSE*) = 0.5 is 0.16 ft. If an admissible *RMSE* of 0.2 ft is adopted, the *P*(*RMSE* ≤ admissible *RMSE*) is equal to 0.82. That is, the 82 % of the time the base subset of 12 stations should be have a *RMSE* smaller than or equal to 0.2 ft. A graph of *RMSESt* in each station for the case of a base subset of 12 stations produced by the NetOptStreamsGr tool is shown in Figure 6.20.

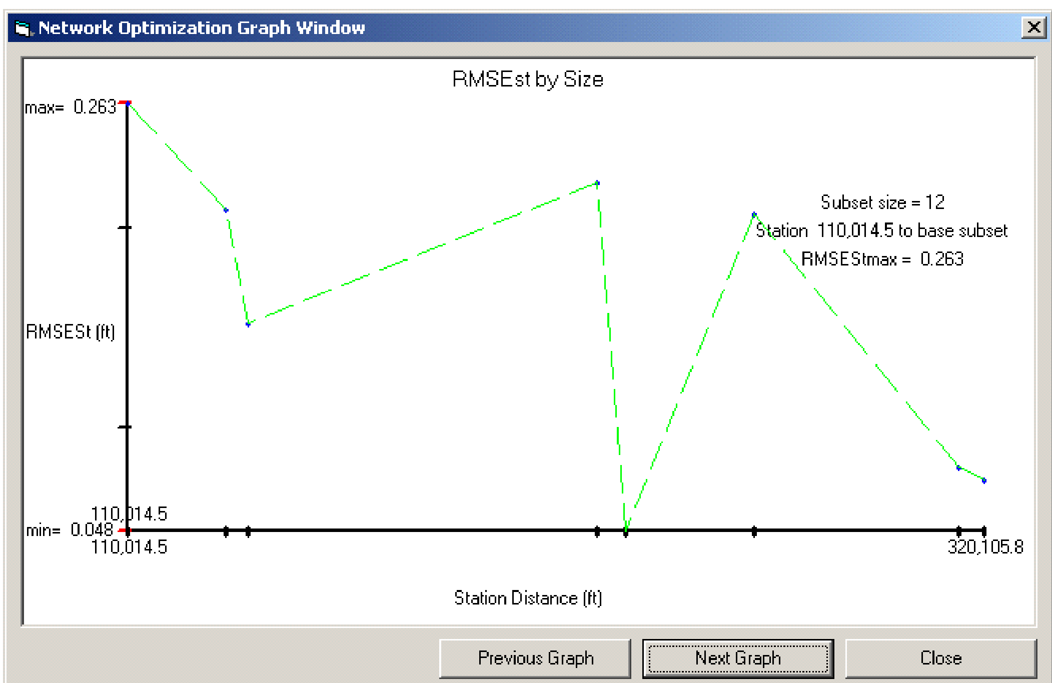


Figure 6.20: *RMSESt* for the base subset of 12 stations.

The only check to be done is the study of the distribution of errors in each of the stations left outside this subset. At the end, the decision to select one base subset depends on the criteria that the water management district (in this case, the SFWMD) might adopt.

In summary, this study demonstrates that by using a combination of HEC-RAS modeling and stage gages, it is possible to reduce the current number of gages on the Pool AE's main stream substantially from 20 to 12 stations and still obtain good estimates of the water surface elevation.

6.4 CONCLUDING REMARKS

The successive steps of the methodology explained in Chapter 3 lead to the refining of the optimization methodology for stations in streams with steady flow. An initial calibration study was necessary before attempting the optimization of Pool AE's main stream stations. Then, a first optimization methodology was applied to three sizes of base subset (7 to 9 stations). Results from those optimizations lead to the formulation of the second methodology. The application of the second methodology reported good results in Kissimmee River, the main stream of Pool AE (case study 5), the existing network of 20 stations could be reduced to 12 stations without loosing too much information.

Chapter 7: Optimization of Stage-Monitoring Networks in Streams (Unsteady Flow)

7.1 INTRODUCTION

In this chapter, the main results of the development of the methodology for optimizing stage-monitoring networks in streams considering unsteady flow are presented. First, the preliminary calibration of the studied stream is presented. Then, the results of the application of the first methodology for streams with unsteady flow to the optimization of the main stream of Pool AE's monitoring network with one-day duration data (case study 6) are shown. Finally, the results of application of the second methodology to the same stream, but considering a longer time period (case study 7), are presented.

7.2 PRELIMINARY CALIBRATION OF POOL AE MAIN STREAM'S HEC-RAS MODEL

Case study 6, described in Chapter 3 is used in this preliminary calibration. Beginning from the calibrated model developed for steady flow, a series of unsteady models was studied. As the study was unfolding, successive modifications were performed, with the following results:

First model. As indicated in Chapter 3, only two days of the calibrated model developed for the steady flow case were used in the first model. Results were discouraging; the water level increases steadily in the upper part of the naturalized river. The extreme cross sections of the naturalized river are 203112.3 and 145061.9, its upstream section is close to station WEIR1 (211720.5) and its downstream section is close to station PC11R (139678.6). The relationship between code of a station and main channel distance, in feet, is equal to code of station minus 47,964.26. For example, the main channel distance of station Weir 1 is = 211,720.5 ft – 47,964.26 ft = 163,756.24 ft. Figure 7.1 shows the beginning of the simulation while Figure 7.2 shows its end. The vertical green lines represent inline structures.

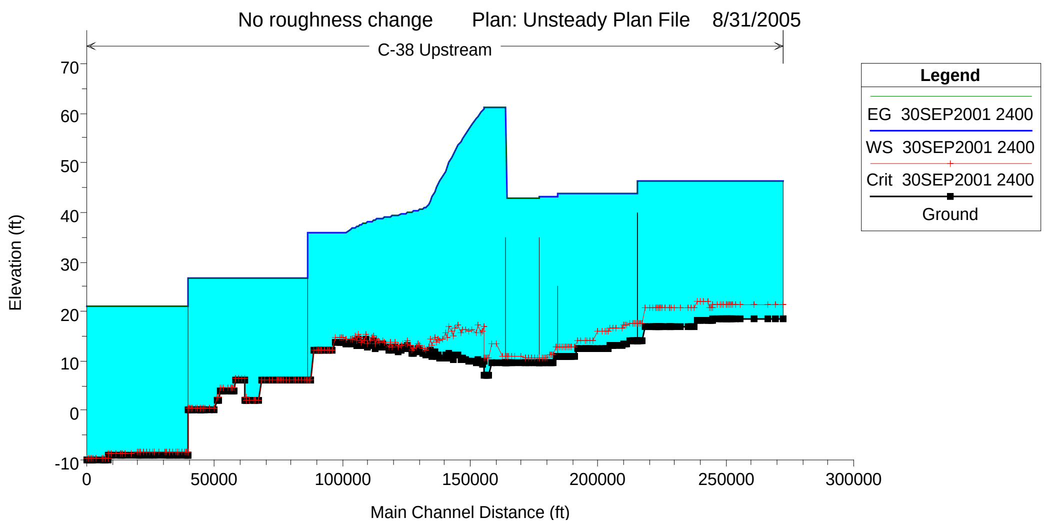


Figure 7.1: First model. Initial time profile.

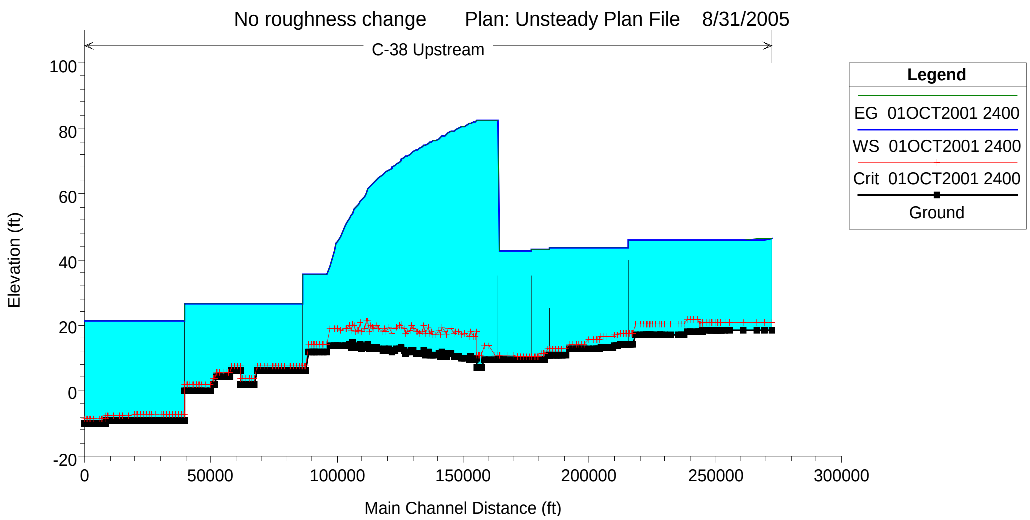


Figure 7.2: First model. Final time profile.

One reason for the presence of the high water elevations between stations WEIR1 and PC11R was the absence of the table of roughness factors that are used by the HEC-RAS unsteady flow module. This table is additional to that used by the steady flow module. The big jump shown in Figures 7.1 and 7.2 in the WEIR1_H at around the main channel distance of 164,000 ft appears because the water elevation in WEIR1_H is set to compute the flow profile from that station to the next upstream station.

Second model. After the model was supplemented with the required table of roughness factors, a new simulation was run. There was a noticeable change; the huge increase of water elevation along the naturalized river reach present in the previous model is now a column close to Weir 1 at the initial time. As time runs, this column grows and extends downstream to station PC11R. Profiles of this model and the previous model have similar maximum water elevations. Figures 7.3 and 7.4 show, respectively, the initial and final profiles.

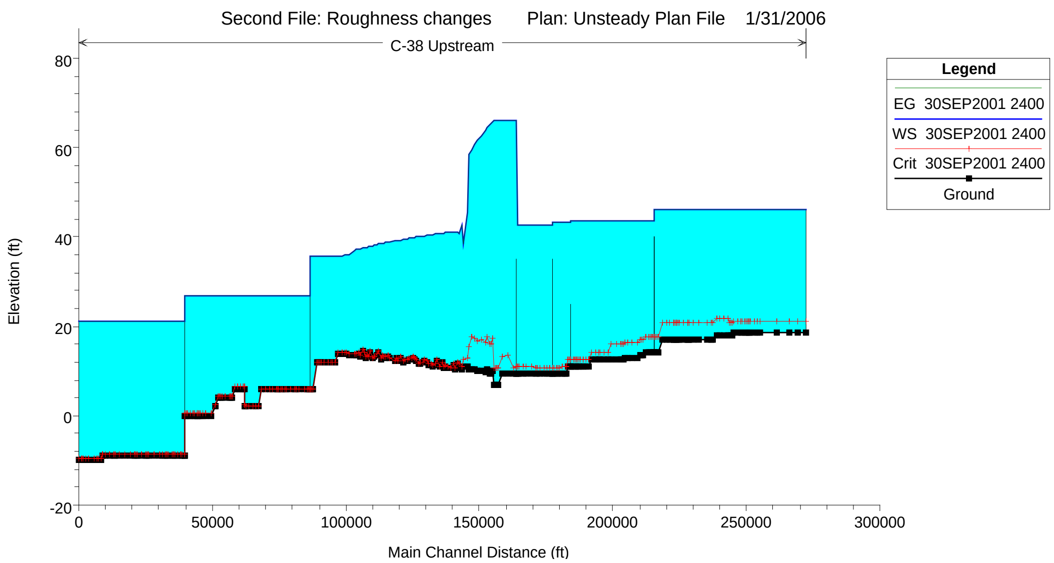


Figure 7.3: Second model. Initial time profile.

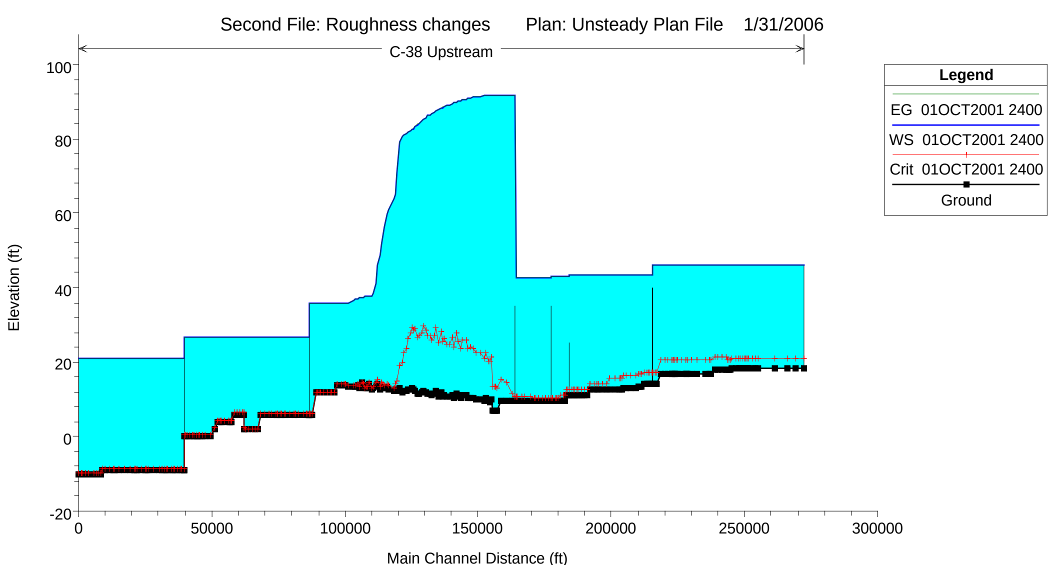


Figure 7.4: Second model. Final time profile.

Third model. After creating several new cross sections along the stream, a new simulation was run. Running the unsteady flow simulation again did not improve results. There is another reason for the accumulation of water downstream stations WEIR1_H. Figures 7.5 and 7.6 show the initial and final profiles.

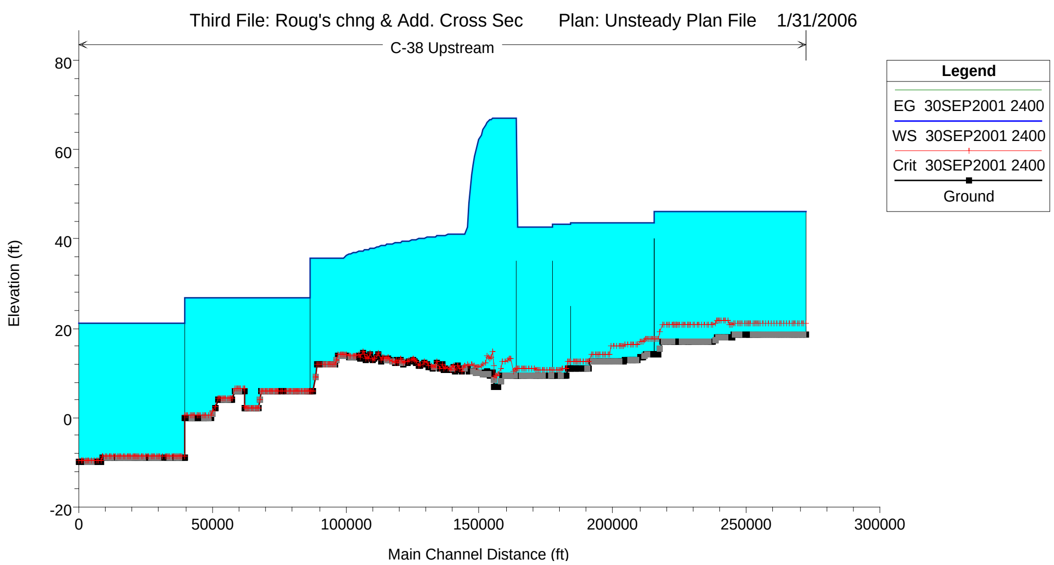


Figure 7.5: Third model. Initial time profile.

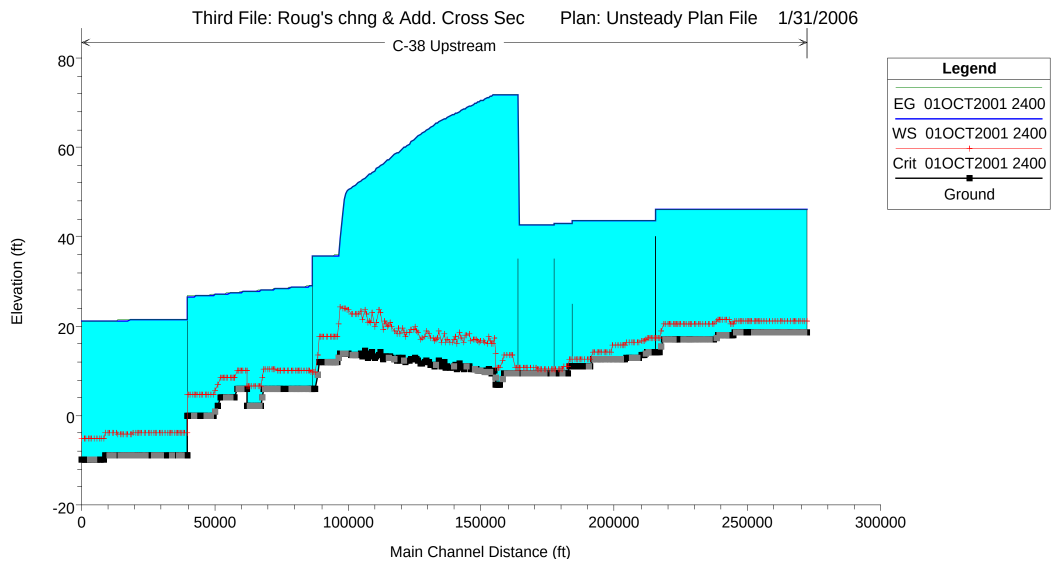


Figure 7.6: Third model. Final time profile.

Fourth model. Taking into account the results of the previous model, an adjustment of the roughness factors of the naturalized river was made. It was found that the naturalized river's Manning's *n* values needed to be reduced from 0.065 to values to 0.0065 to have more or less credible profiles. Manning's *n* values of 0.0065 are unrealistically small, although to obtain credible profiles they were adopted. Taking into account the roughness factors, the value of *n* for discharges close to 0.0 cfs is 0.0631, and 0.0033 for discharges of the order of 13,000.0 cfs. Additionally, it was found that an oscillating behavior is present in the system. The bigger oscillations in water level elevation and flows are in the Pool BC area. For the case of n = 0.0065, Figures 7.7 to 7.11 show the evolution of the water profile in increments of 6 hours.

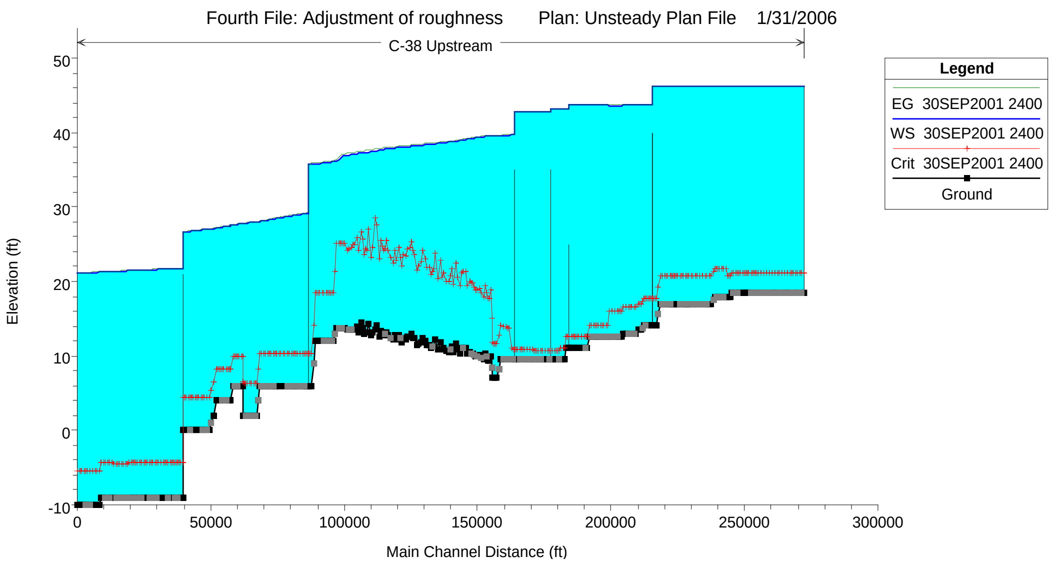


Figure 7.7: Fourth model. Initial time profile.

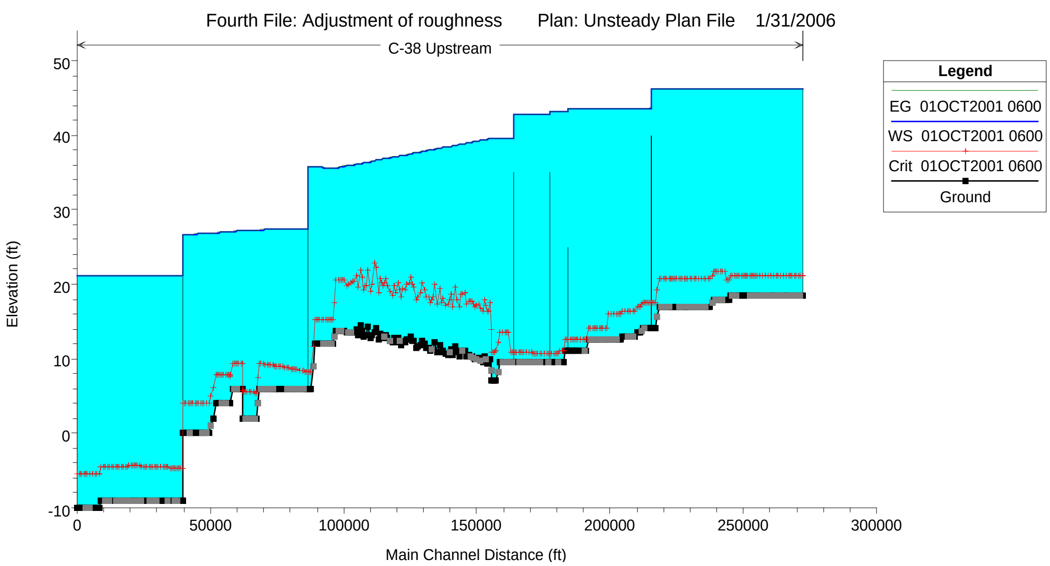


Figure 7.8: Fourth model. Profile at 6 hr of simulation.

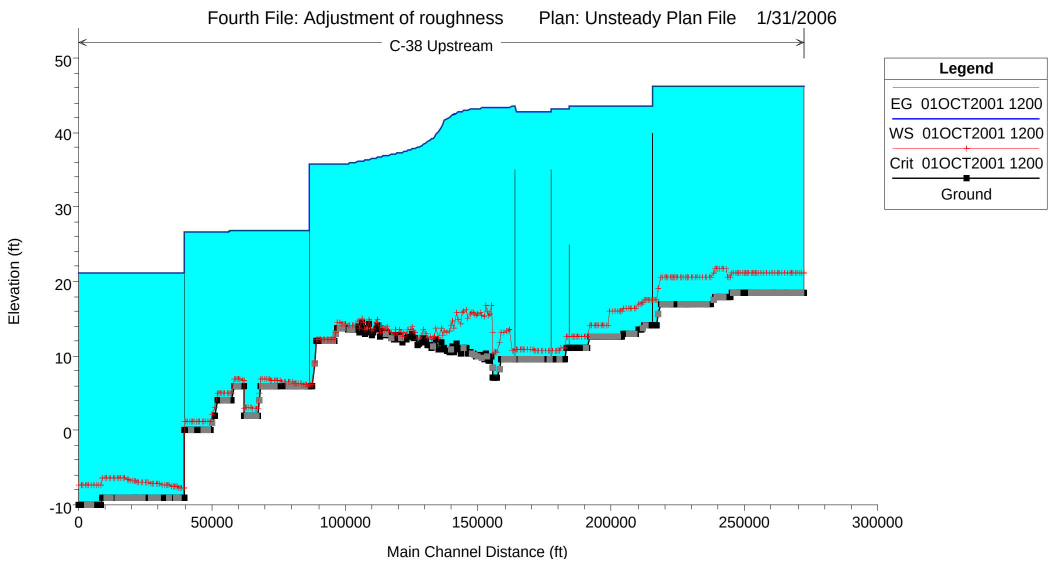


Figure 7.9: Fourth model. Profile at 12 hr of simulation.

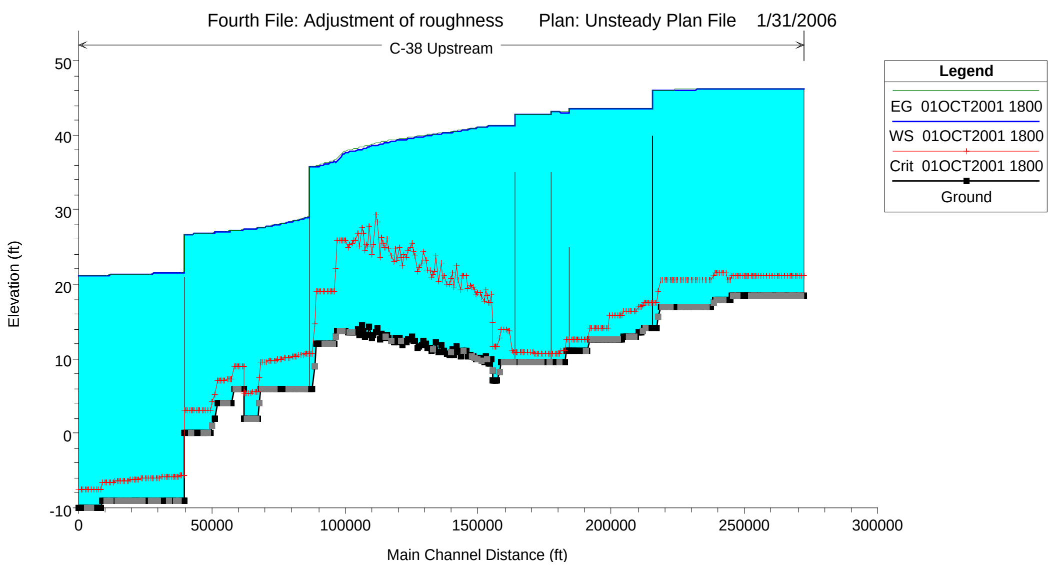


Figure 7.10: Fourth model. Profile at 18 hr of simulation.

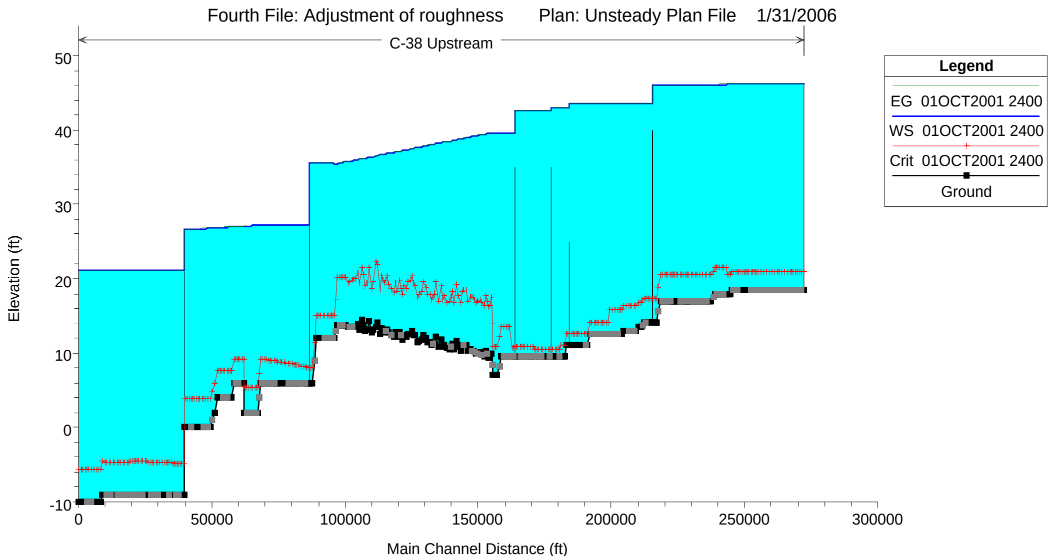


Figure 7.11: Fourth model. Final time profile.

Fifth model. In the previous models, a physically inadmissible feature is present. It is a sudden decrement of water level when water should go from the tail side to the head side of Weir 1. There was a need to change the former model from using internal boundaries in the inline structures to using properties of these structures, namely, length of crest, elevation of crest and discharge coefficient (data taken from SFWMD's DBHydro). In all cases, it was considered that the flow is uncontrolled. The new results, for n = 0.0065 in the naturalized river, are exemplified in Figures 7.12 to 7.16.

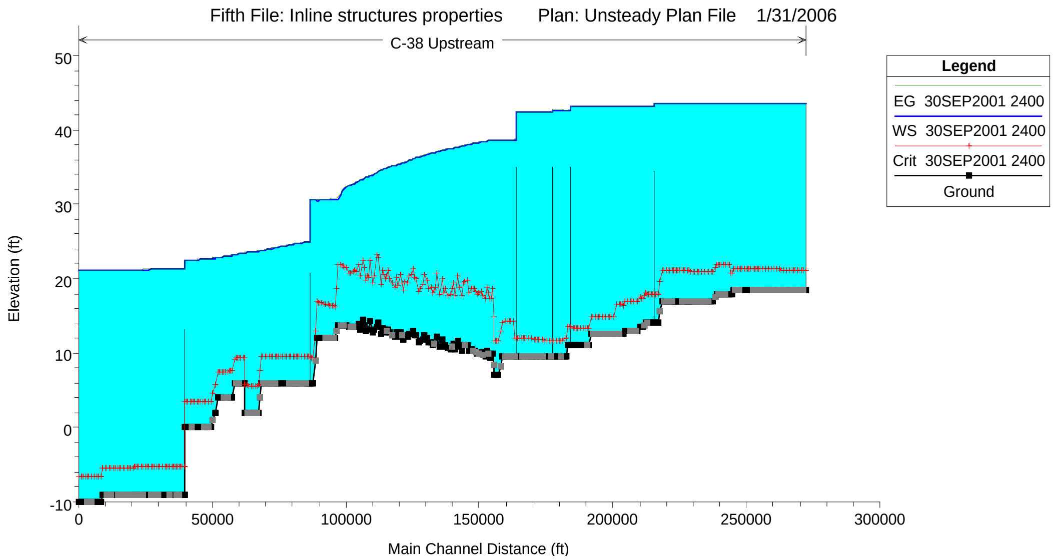


Figure 7.12: Fifth model. Initial time profile.

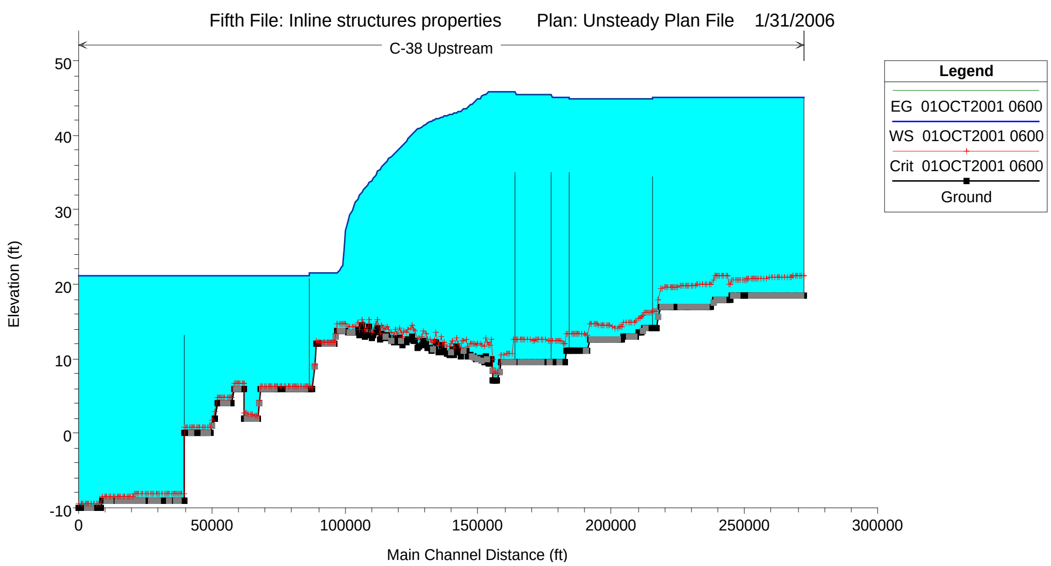


Figure 7.13: Fifth model. Profile at 6 hr of simulation.

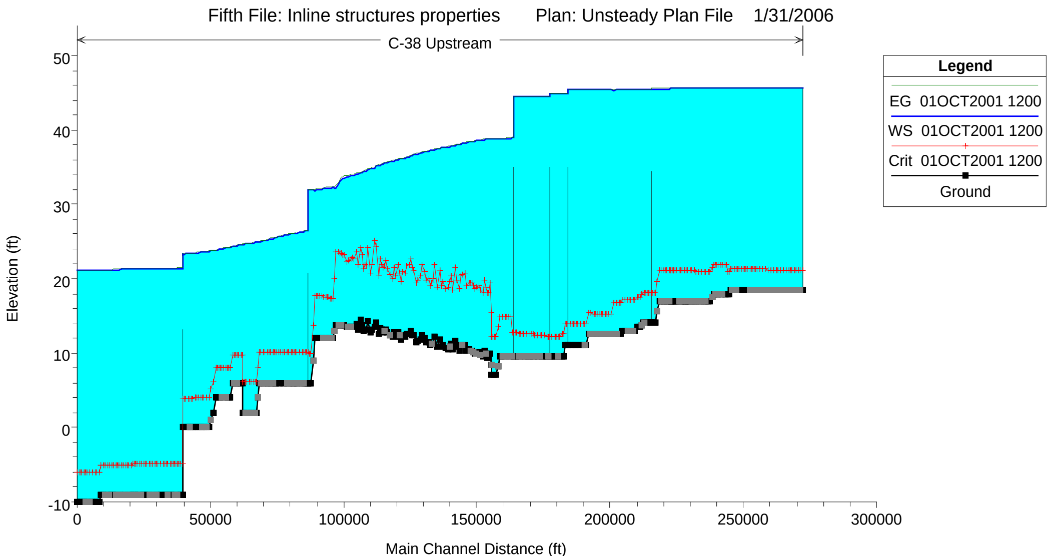


Figure 7.14: Fifth model. Profile at 12 hr of simulation.

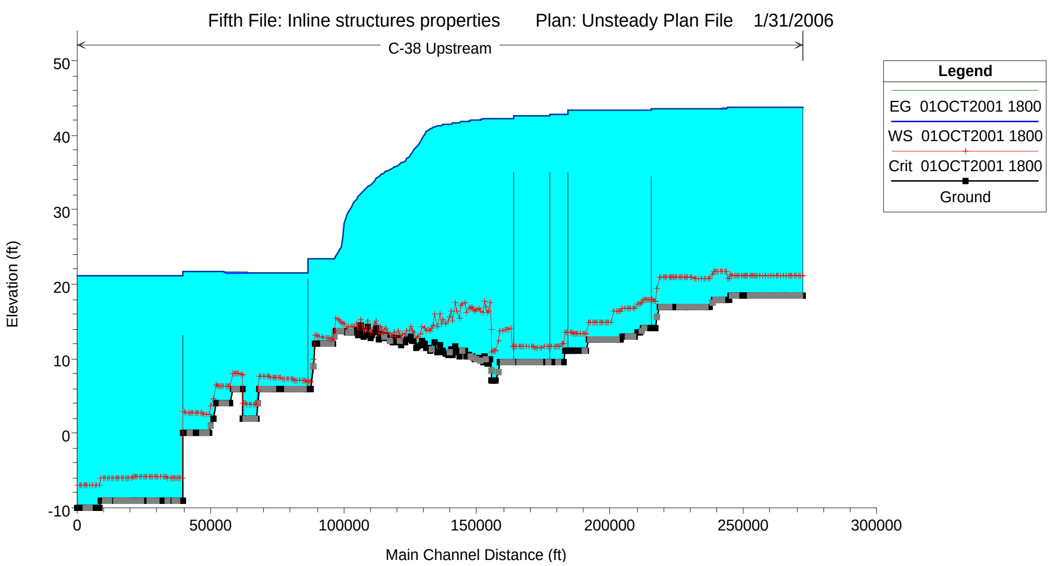


Figure 7.15: Fifth model. Profile at 18 hr of simulation.

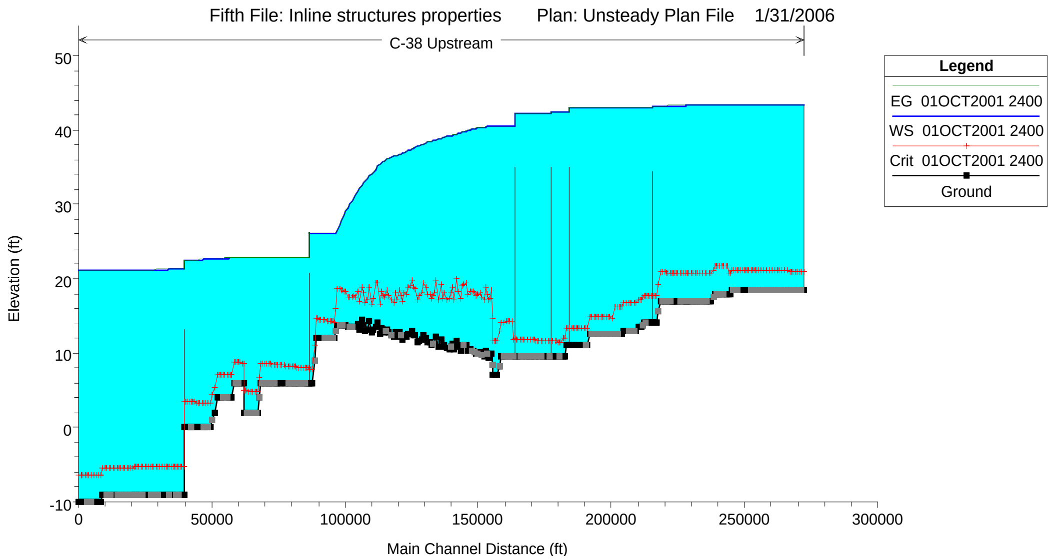


Figure 7.16: Fifth model. Final time profile.

With the last modifications to the model, the wall of water downstream Weir 1 disappeared and backwards flow appeared. This model, like the previous one, shows noticeable oscillations in the water elevations and flows along the naturalized river.

Sixth model. Several attempts to produce steady flow were made. It was not possible to produce it even in an extreme case that considered 15 days of simulation. In that case, the input for Sep 30, 2001, at 24:00 was extended for three days, then one day of transition from the data of Sep 30, 2001, at 24:00 to the data of Oct 1, 2001, at 24:00, and finally the data of the second date was extended for another eleven days. During all simulations, there was an oscillating flow.

Influence of the computation interval's size. Using the fifth model a series of runs was performed. Figures 7.17 to 7.21 show the profiles for a computation interval equal to 12 minutes.

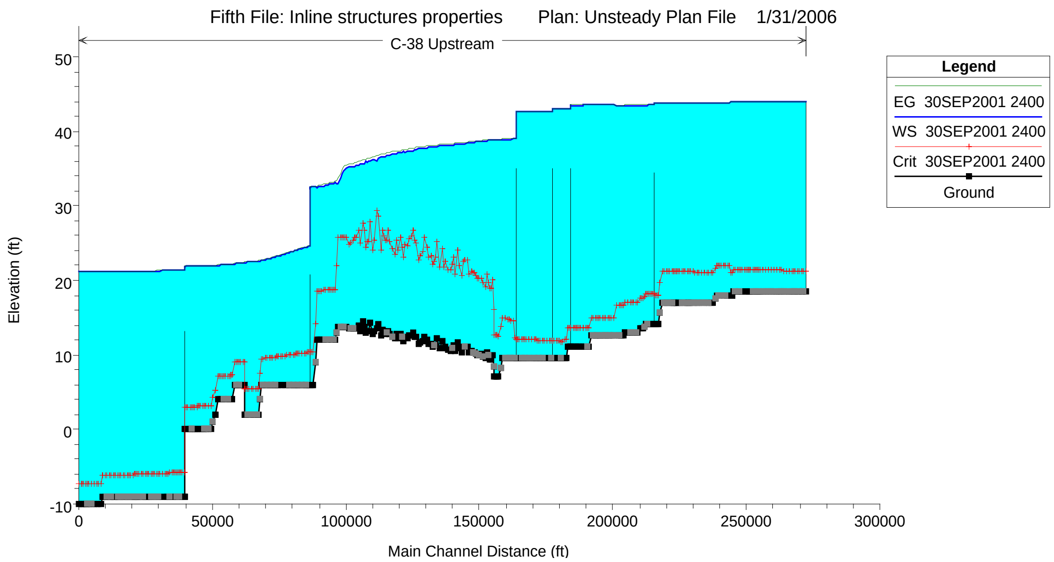


Figure 7.17: Fifth model with Δt = 12 min. Initial time profile.

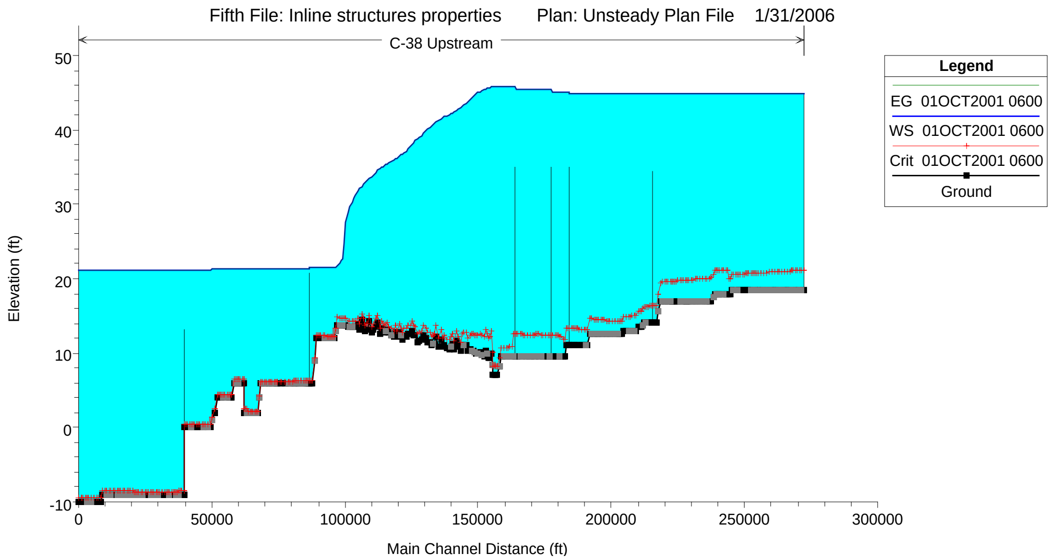


Figure 7.18: Fifth model with Δt = 12 min. Profile at 6 hr of simulation.

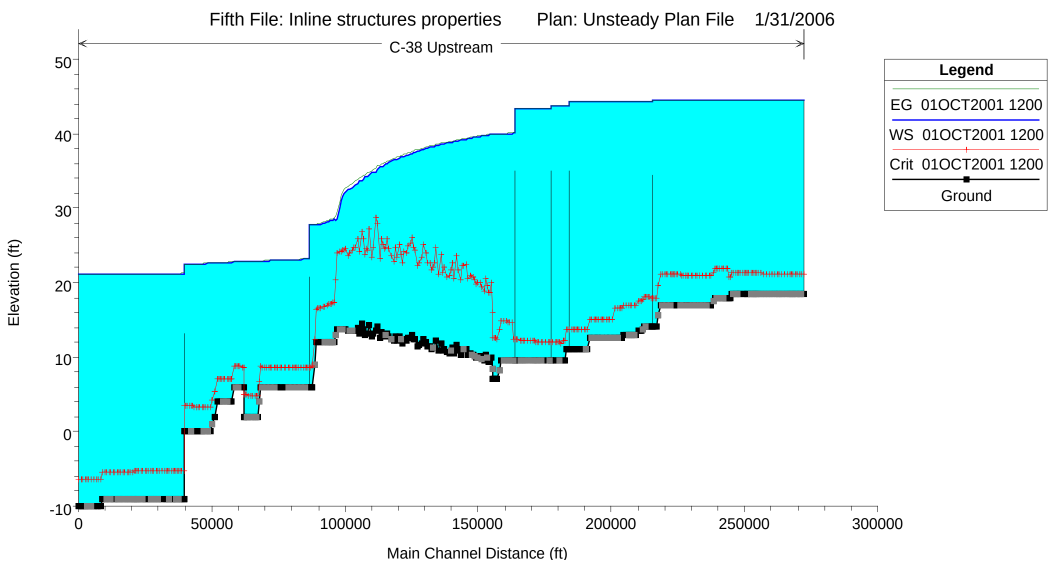


Figure 7.19: Fifth model with Δt = 12 min. Profile at 12 hr of simulation.

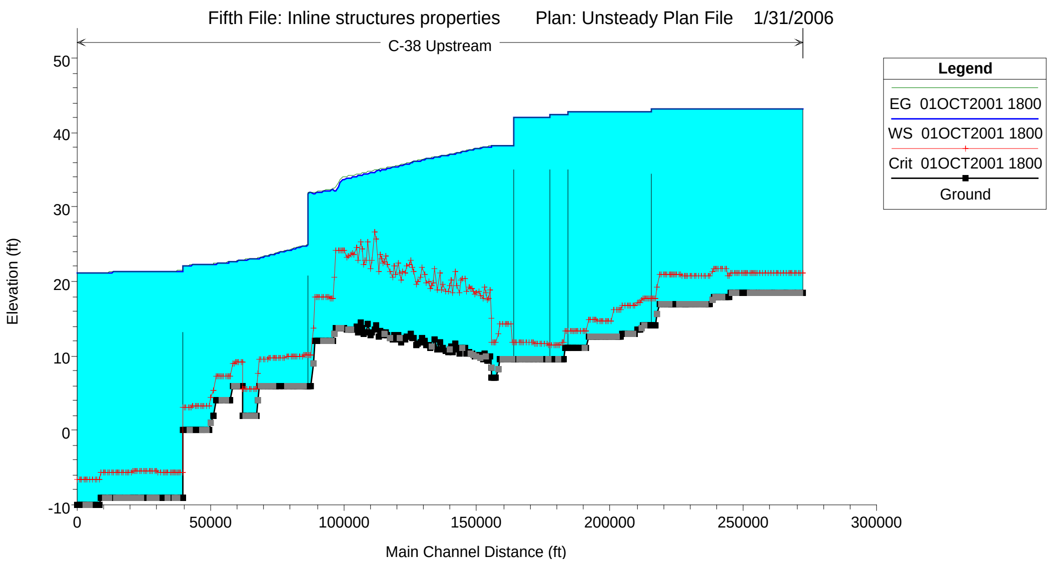


Figure 7.20: Fifth model with Δt = 12 min. Profile at 18 hr of simulation.

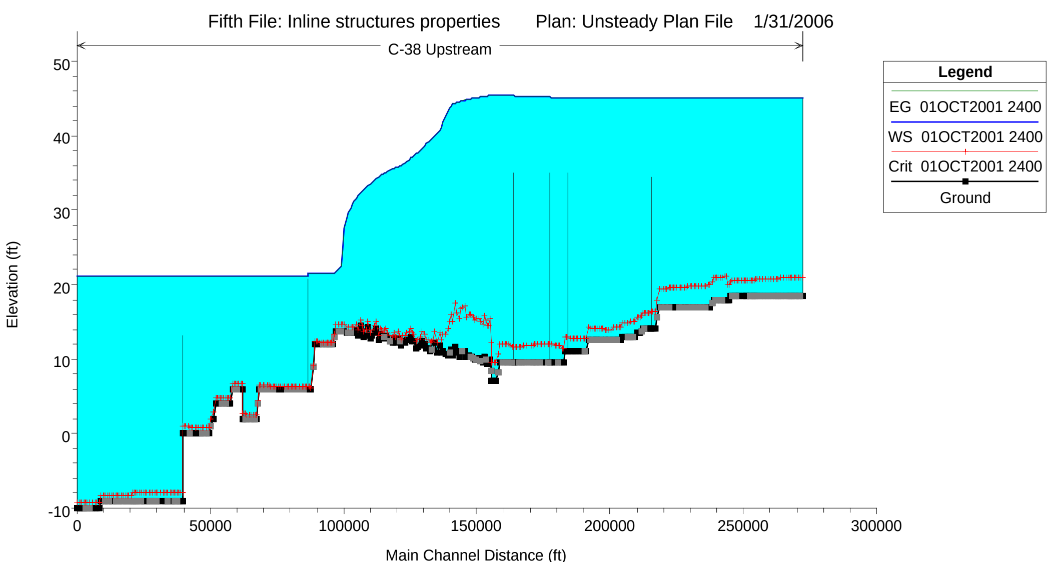


Figure 7.21: Fifth model with Δt = 12 min. Final time profile.

Running the model with Δt = 10, 6, and 5 min, similar profiles were computed. If the water profiles for different values of computational interval are compared, it seems that they are becoming more similar as the computation interval decreases. However, water profiles for different Δt are not in the same phase. It seems that the settling of the initial profile depends on the computation interval used. Moreover, for effectively comparing profiles obtained with different computation intervals one may begin the simulation of unsteady flow until the flow at the begging of the simulation has reached a steady state.

A graph that can be used to judge whether a computational time is sufficiently small is the maximum water elevation profile. Figures 7.22 to 7.27 show that profile for the considered computational times.

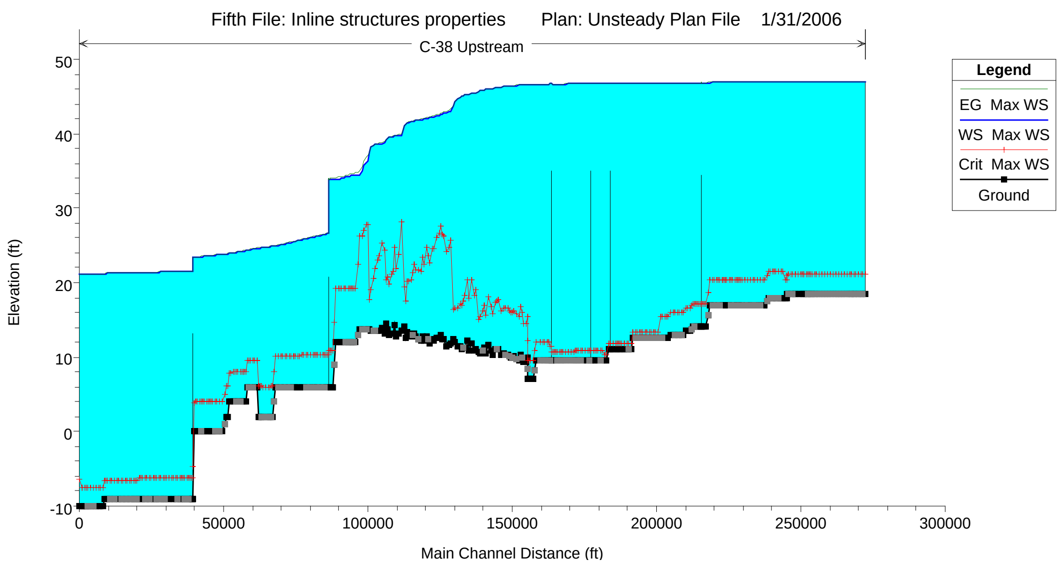


Figure 7.22: Fifth model with Δt = 15 min. Maximum water elevation profile.

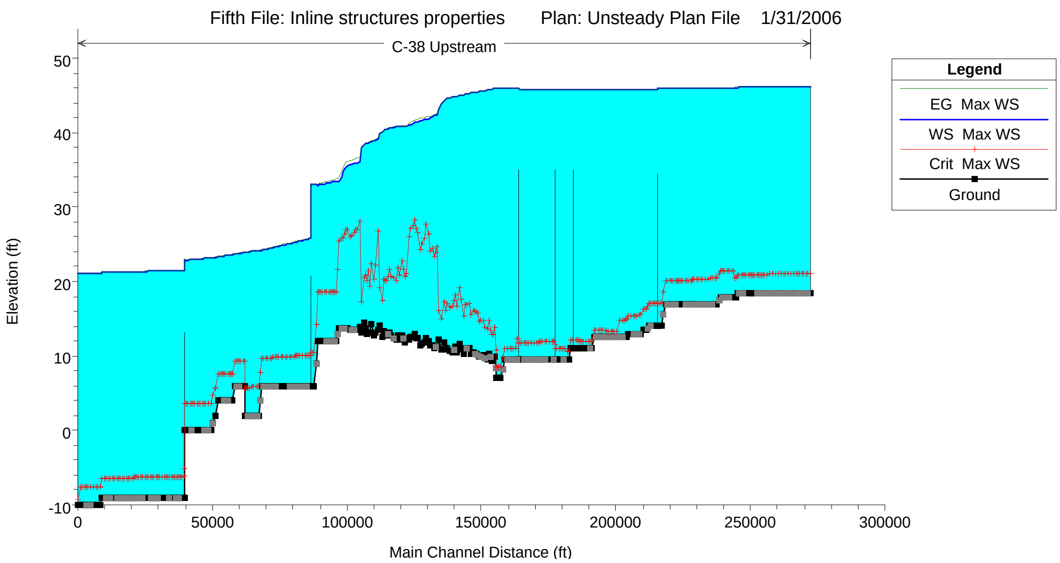


Figure 7.23: Fifth model with Δt = 12 min. Maximum water elevation profile.

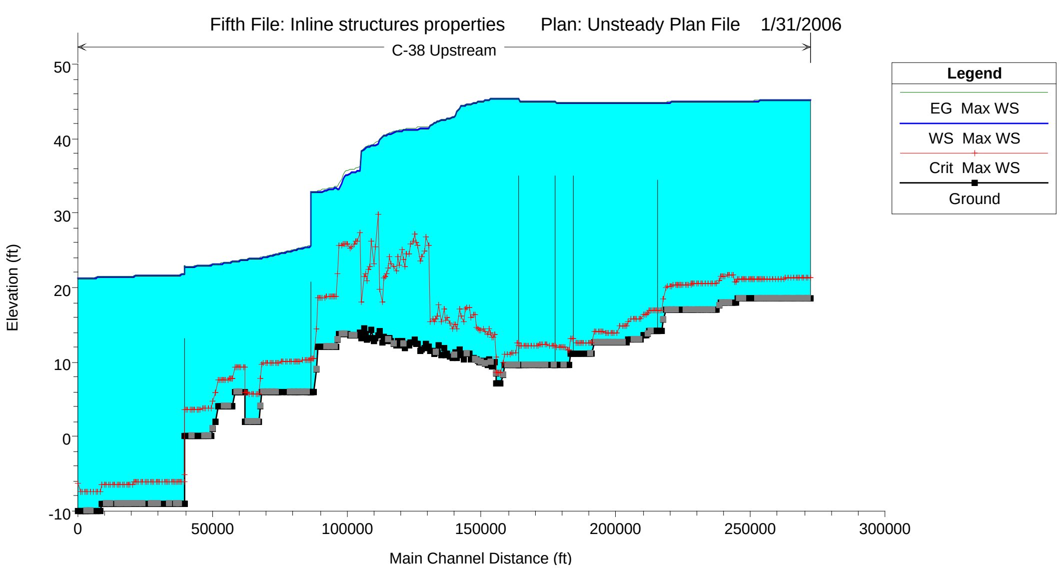


Figure 7.24: Fifth model with Δt = 10 min. Maximum water elevation profile.

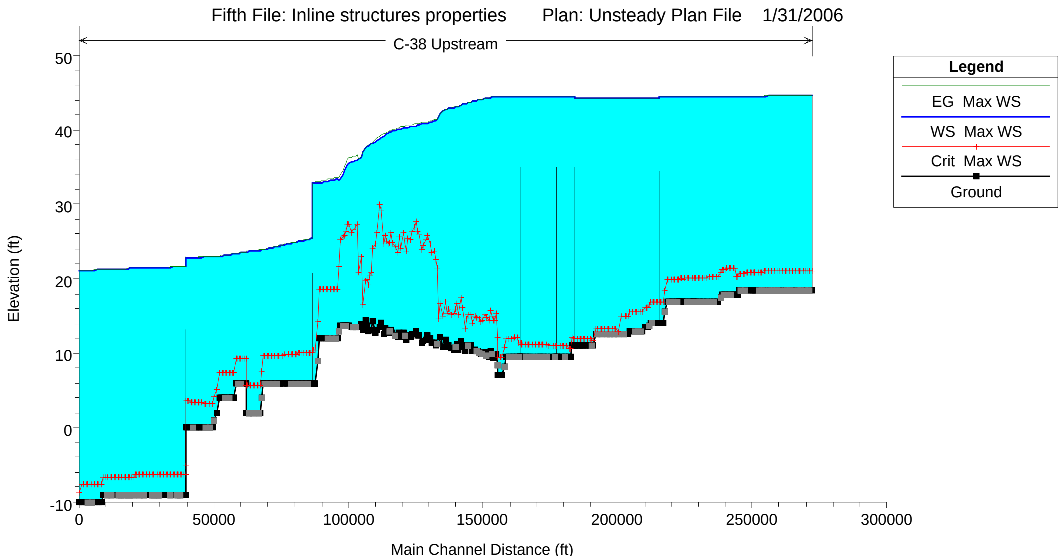


Figure 7.25: Fifth model with Δt = 6 min. Maximum water elevation profile.

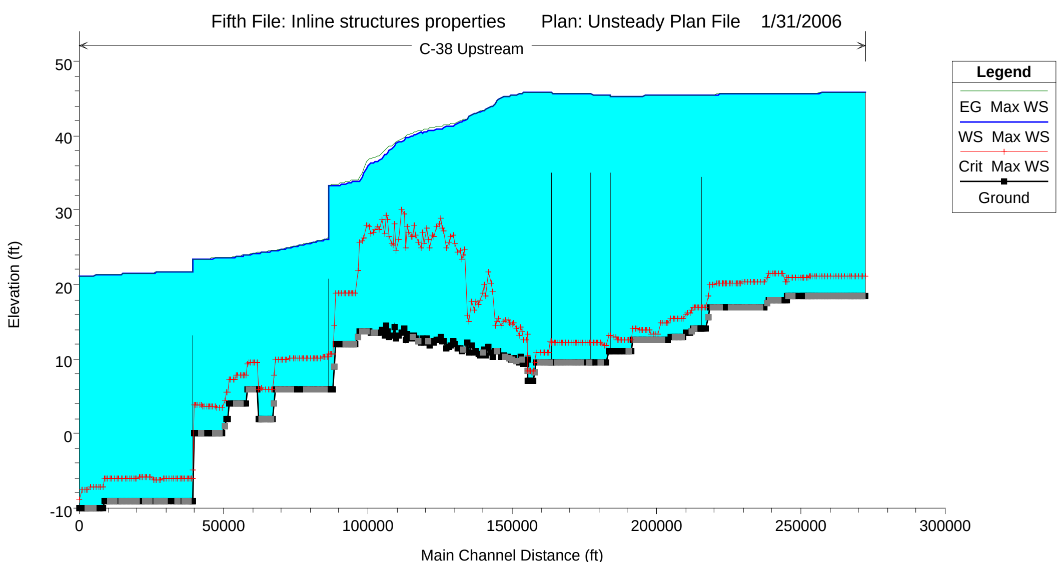


Figure 7.26: Fifth model with Δt = 5 min. Maximum water elevation profile.

The evolution of the maximum elevation profile for computational times of 15 to 5 minutes shows a tendency to have similar profiles for small consecutive computational times; leading to the identification of a computational time sufficiently small, this value could be 5 minutes.

According to the results obtained so far, the following statements can be formulated:

- Simulation of unsteady flow in the main stream of Pool AE is much more complex than that of steady flow.
- Steady flow cannot be attained, at least, for several days for the given input and boundary data.
- An important issue to the possible use of unsteady flow modeling in the optimization of the stage-monitoring network is that the introduction of internal boundaries (needed to set water elevations) makes the computed profiles physically inadmissible.
- Probably, a revision of the naturalized river's cross sections could be useful to gain a better understanding of this modeling problem.

In summary, according to the results obtained to this point, a use of unsteady flow modeling may be not recommended to be applied in the stage measuring network optimization problem, at least for the studied stream.

7.3 FIRST OPTIMIZATION OF POOL AE MAIN STREAM'S MONITORING NETWORK

Using the modified model of the stream with one-day data and varying the roughness value of the naturalized river reach, the results in Table 7.1 were found.

Manning's *n*	Average *RMSE*	Standard Deviation of *RMSE*
0.0063	1.02	0.50
0.0064	1.01	0.49
0.0065	1.01	0.49
0.0066	1.01	0.49
0.0067	1.09	0.52

Table 7.1: Results of the calibration of the one-day duration Pool AE main stream HEC-RAS model.

The optimal value of n can be considered as that which minimizes the average of *RMSE*; therefore, from Table 7.1, n = 0.0065. Once the best n value is known, the optimization process can begin. Running the optimization tool for unsteady flow developed to solve this problem and considering base subset sizes from 7 to 19 stations, the results shown in Table 7.2 and Figure 7.27 were found. The initial base subset of seven stations is the same as the steady flow case is: S65A_H, WEIR3_H, WEIR2_H, WEIR1_H, S65C_H, S65D_H and S65E_H. As in the steady flow case, the station that goes into the base subset for the next cycle is that which has the maximum *RMSESt*.

Number of stations in base subset	Station entering base subset	Maximum *RMSESt* (ft)
7	WEIR1_T (211720.5)	2.35
8	KRDRS (188833.2)	1.84
9	S65C_T (134399.5)	2.08
10	PC33 (149107.5)	1.38
11	KRBNS (171777.7)	1.26
12	S65D_T (87361.08)	1.07
13	PC11R (139678.6)	0.79
14	C38BAS (110014.5)	0.83
15	S65A_T (263634.2)	0.28
16	WEIR2_T (225184.9)	0.22
17	KRFNS (314076.8)	0.14
18	WEIR3_T (232164)	0.05
19	S65_T (320105.8)	0.01

Table 7.2: Results of the optimization of Pool AE main stream network considering a one-day duration dataset (case study 6).

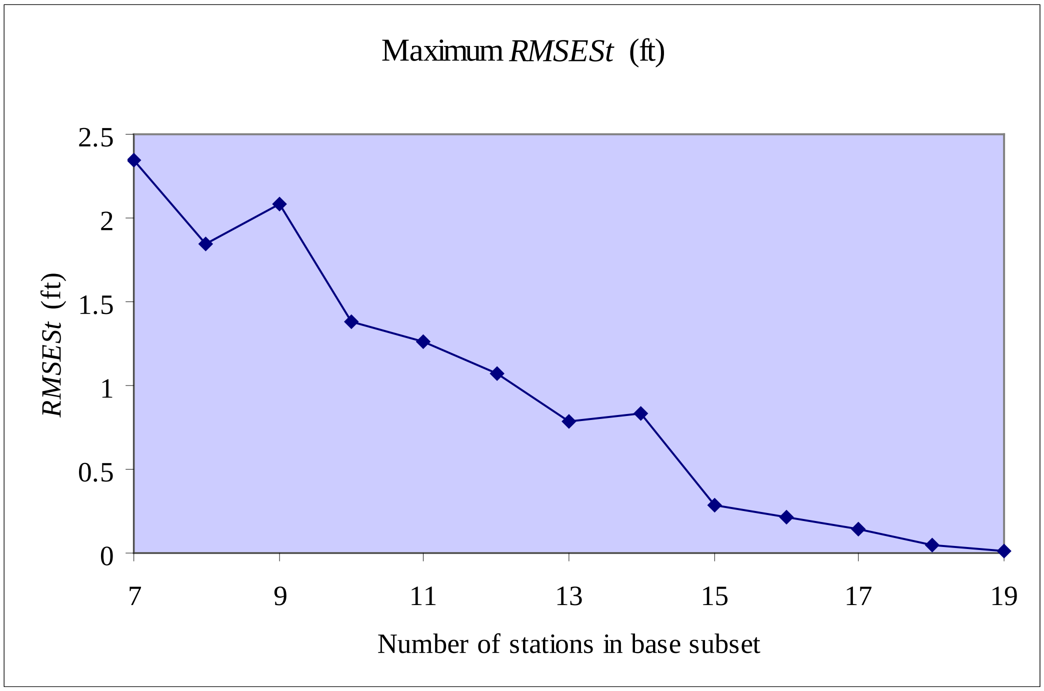


Figure 7.27: Optimization of Pool AE main stream's network considering a one-day duration dataset.

In Figure 7.27, it is evident that something unusual happened with the maximum *RMSESt* of the base subsets of 9 and 14 stations: instead of it being smaller that the previous value, it is larger. This behavior might be the result of a limited time of simulation in which the errors were computed. Using longer times of simulation might lead to more robust estimations of the maximum *RMSESt* values.

7.4 SECOND OPTIMIZATION OF POOL AE MAIN STREAM'S MONITORING NETWORK

From the results of the previous optimization, a new seven-day period dataset (case study 7) is considered here, from 30SEP2001 2400 to 07OCT2001

2400. Calibrating the value of Manning's *n* for this period it was found that *n* = 0.0065 was not the optimal value, the optimal value was now *n* = 0.0057. Once the calibration was finished, the optimizing process was run for all the possible sizes of base subset. The results shown in Table 7.3 and Figure 7.28 were found.

Number of Stations in base subset	Station entering base subset	Maximum *RMSESt* (ft)
7	WEIR1_T (211720.5)	1.890
8	KRDRS (188833.2)	1.604
9	S65C_T (134399.5)	1.831
10	S65D_T (87361.08)	1.613
11	KRBNS (171777.7)	1.472
12	PC33 (149107.5)	1.634
13	C38BAS (110014.5)	0.928
14	PC11R (139678.6)	0.723
15	WEIR2_T (225184.9)	0.269
16	S65A_T (263634.2)	0.185
17	KRFNS (314076.8)	0.090
18	S65_T (320105.8)	0.048
19	WEIR3_T (232164)	0.045

Table 7.3: Results of the optimization of the Pool AE main stream's network considering a seven-day duration dataset (case study 7).

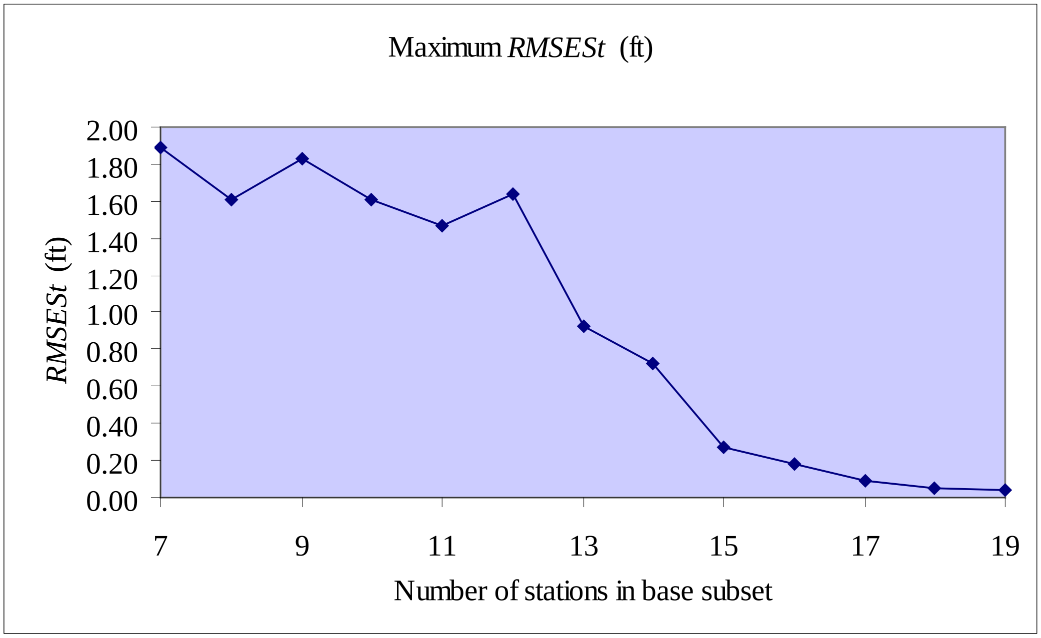


Figure 7.28: Optimization of Pool AE main stream's network considering a seven-day duration dataset.

Table 7.3 and Figure 7.28 present two increases of *RMSESt* when the size of the base subset passes from 8 to 9 stations and when the size passes from 11 to 12 stations. As was mentioned before, *RMSESt* should decrease with an increase in the number of stations in the subset. An increase in *RMSESt* could be caused by some degree of numerical instability using the new subset and/or the need for a longer period of simulation. A typical water elevation profile for $r = 12$ is shown in Figure 7.29. A map illustrating the results of the optimization process at the end of the iteration with nine stations in the base subset is shown in Figure 7.30.

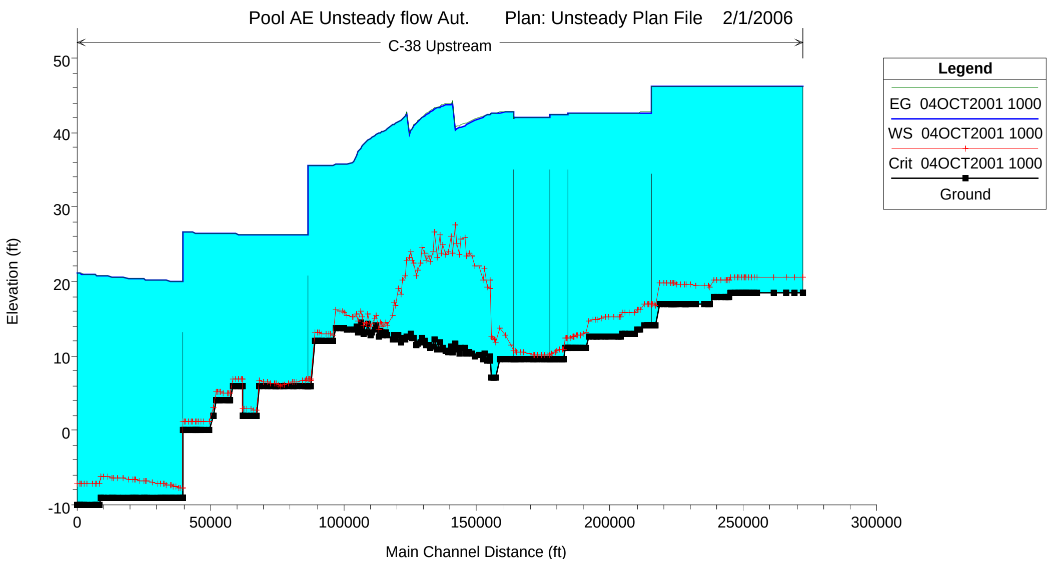


Figure 7.29: Computed surface water profile along Pool AE main stream on October 4, 2001, at 10 hours. Base subset with 12 stations.

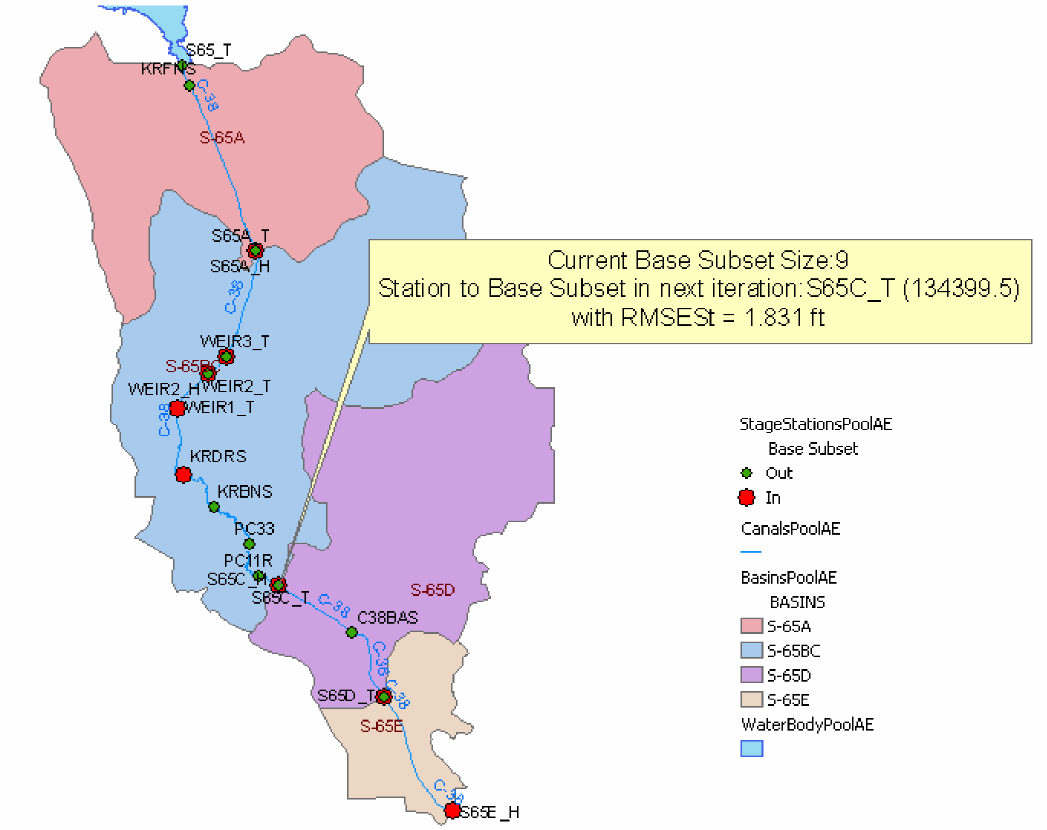


Figure 7.30: Results of nine stations in base subset of the optimization of Pool AE main stream's network (seven-day duration dataset).

The last results show that with a combination of HEC-RAS unsteady flow modeling and stage-monitoring network data it is possible to study the reduction of the current number of gages on the Kissimmee River. Nevertheless, the difficulties added to the problem using unsteady flow make it harder to optimize the network. In addition, measurement data should be instantaneous data, not daily or hourly data. The hydraulic geometry of the system should also be more precise than that needed for a steady flow-based optimization.

7.5 CONCLUDING REMARKS

The successive steps, concerning unsteady flow, of the methodology explained in Chapter 3 lead to the refining of the optimization methodology for stations in streams with unsteady flow. An extensive initial calibration study was necessary before attempting the optimization of Pool AE's main stream stations. Then, an optimization using the first methodology was performed using one-day data (case study 6). Results were discouraging. Another attempt of optimization using the second methodology was performed, this time using a seven-day dataset (case study 7). Results again were discouraging. It seems that better flow and stage data and a more precise geometry of the stream should be required to obtain results that could be considered adequate.

Chapter 8: Improved Optimization of the Stage-Monitoring Networks in a Lake and in a Stream with Steady Flow

This chapter contains the results of the application of the VBA macros that implement tabu search to solve the problems of the network optimization in lakes and in streams with steady flow. These macros support the third and four methodologies for lakes, and the third methodology for streams with steady flow. Many details are not presented because they are very similar to details explained in previous sections.

8.1 Lake Okeechobee's Results of the Third Methodology for Lakes

The application of the VBA macro that implements the third methodology for lakes, to find the optimal subset of the stage-measuring network of Lake Okeechobee (case study 8) with an *admissible RMSE* of 0.1 ft and a *reliability* of 0.8 reported the following results. In 20 runs with $r_i = 1$, $r_f = 13$, $itmax = 48$, $itBtOpmax = 10$ and $itBtChgRmax = 5$, the average number of function evaluations to reach the optimum was 568 and the average total number of function evaluations was 809. The average time to find the optimum was 150 seconds (2.5 min) and the average total time was 233 seconds (3.9 min). Computation time is the same computation time previously mentioned for a Dell Workstation PWS360. The average percentage of function evaluations computed with respect to the cardinality of the problem ($2^n - 2 = 2^{14} - 2 = 16{,}832$) is 3.5 % for the optimum, and 4.9 % for the total function evaluations. In all the runs, the true optimum was found. It is $F_r = 0.7723 < R_e = 0.8$. The $r = 5$ stations of the optimal subset are L001, CULV10A_H, CULV5A_H, L_OKEE, and S3_T. To aid in the graphic interpretation of this result, the relative frequencies for all possible sizes of optimal subset are shown in Figure 8.1

r	*Fr*
13	0.9971
12	0.9942
11	09697
10	0.9538
9	0.9201
8	0.8989
7	0.8618
6	0.8319
5	0.7723
4	0.7174
3	0.6255
2	0.5278
1	0.4260

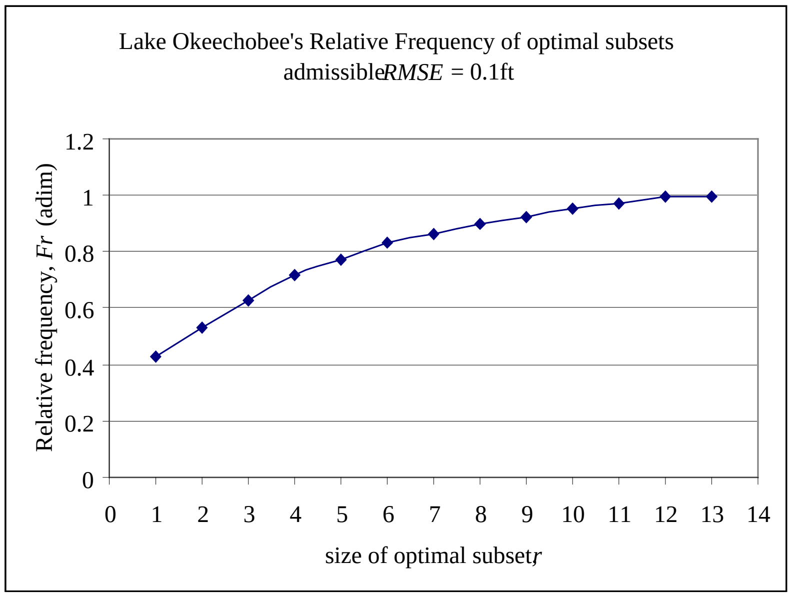


Figure 8.1: Lake Okeechobee’s relative frequency of the optimal subsets for admissible *RMSE* = 0.1 ft.

A screening of the effect of the use of different periods of data to solve the same problem (find the optimal subset for *admRMSE* = 0.1 ft and R_e = 0.8) results in three different optimal subsets. From 10/01/2001 to 9/30/2002, the optimal subset has seven stations; they are S127_T, S191_T, CULV10A_H, CULV5A_H, L_OKEE, L_OKEE.M_G and S3_T. From 10/01/2002 to 9/30/2003, the optimal subset has only the stations S127_T and L005. The optimal subset from 10/01/2001 to 9/20/2003 is already known. Figure 8.2 shows the three subsets. It can be noted that the optimal subset for the entire period comprises stations that are evenly distributed and include a central station (L_OKEE). It seems to be the best of the three solutions.

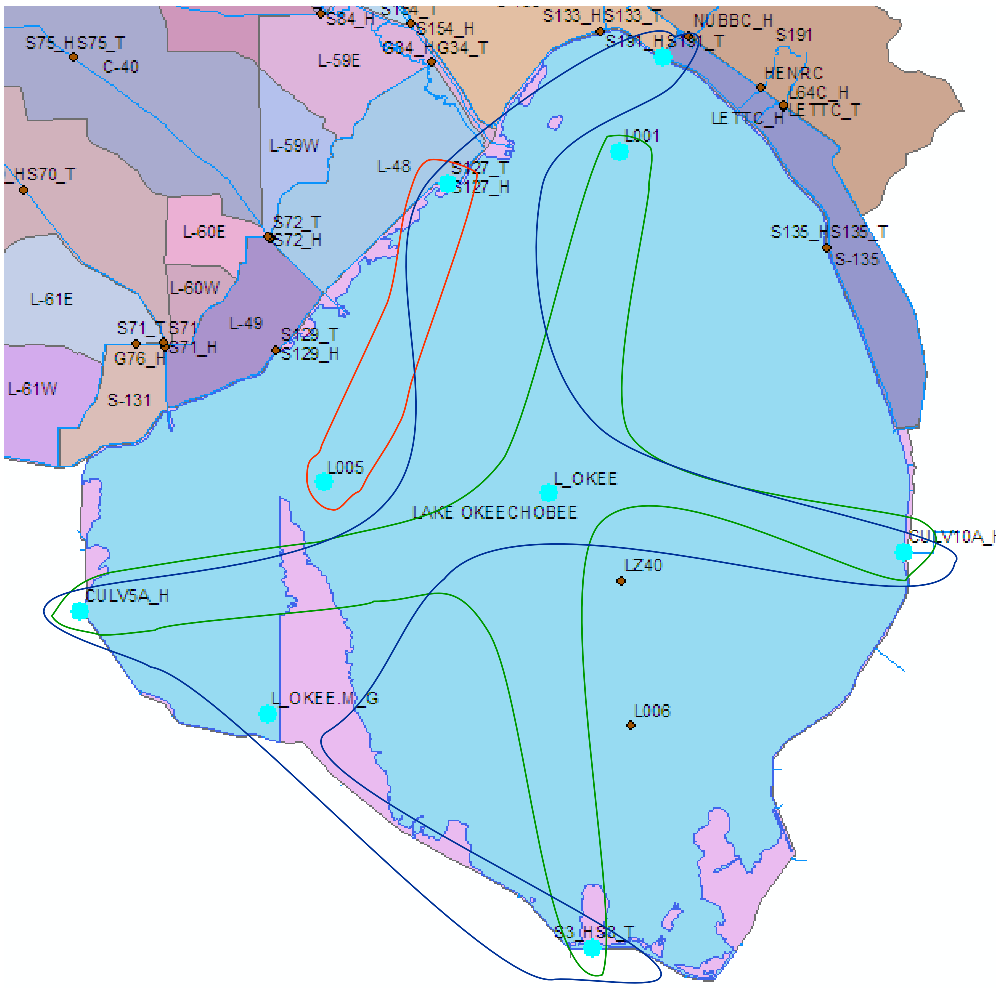


Figure 8.2: Closed curves indicating the optimal subsets in Lake Okeechobee for the three different periods. The blue curve refers to the optimal subset of the first year, the red to that of the second year, and the green to that of the entire period.

A summary of results is shown in Table 8.1 and Figure 8.3. In Table 8.1, the relative frequency of the required optimal subsets for the three periods is highlighted. In Figure 8.3 and Table 8.1, it can be observed that, for a given size, the relative frequency of the two-year period is more or less the average of the relative frequencies of the two annual periods.

r	10/01/2001 to 9/30/2002	10/01/2002 to 9/30/2003	10/01/2001 to 9/30/2003
13	0.9944	1.0000	0.9971
12	0.9888	1.0000	0.9942
11	0.9638	1.0000	0.9697
10	0.9387	1.0000	0.9538
9	0.8774	1.0000	0.9201
8	0.8384	0.9917	0.8989
7	0.7671	0.9805	0.8618
6	0.7245	0.9555	0.8319
5	0.6547	0.9258	0.7723
4	0.5918	0.9021	0.7174
3	0.4530	0.8368	0.6255
2	0.3445	0.7267	0.5278
1	0.2527	0.6000	0.4260

Table 8.1: Lake Okeechobee. Relative frequency of the optimal subsets for all sizes and three different periods of data.

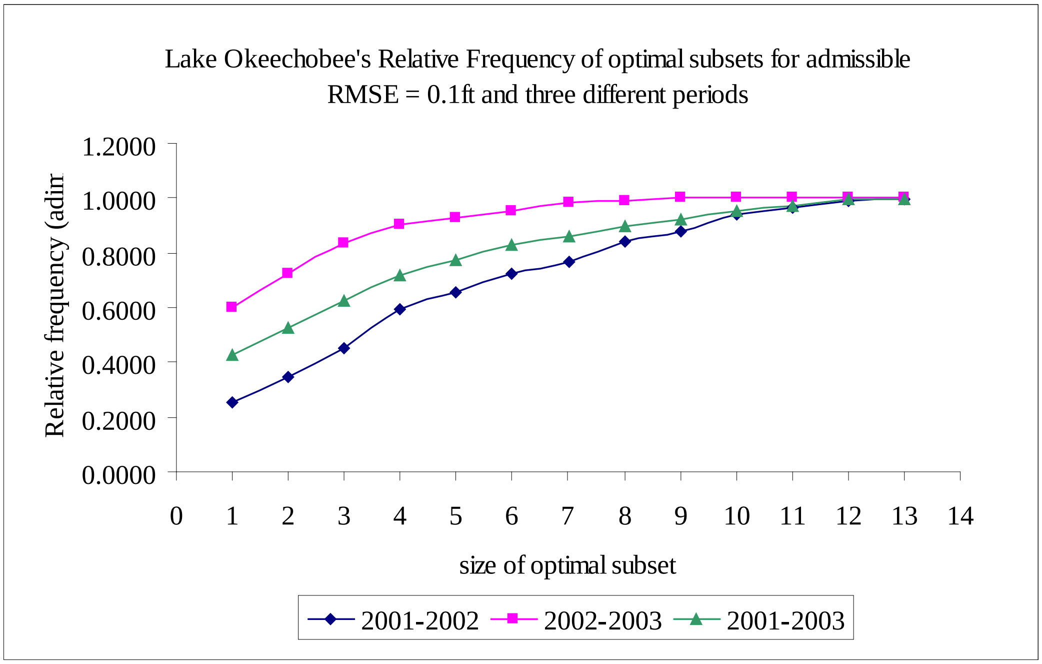


Figure 8.3: Lake Okeechobee's relative frequency of the optimal subsets for admissible *RMSE* = 0.1 ft and three different periods.

The identification of the optimal subsets of all sizes of Lake Okeechobee using the second methodology (based on genetic algorithm) and the third methodology (based on tabu search) allow for comparison of the results of both optimization procedures. Computational time was chosen to be used to compare both methods. Although the results could be obtained in only one run, each result (of size and method) was found in one separated run to be able to differentiate the time the computer spent computing each particular result. Even though the computational times of both methods are probabilistic, (the genetic algorithm execution is in essence random and the tabu search method begins with a random solution), the difference of times of each pair of corresponding results is sufficient to draw conclusions. Running the respective computational tools, the results of Table 8.2 were found.

Size	Genetic Algorithm		Tabu Search	
	T (minutes)	F_r (adim)	T (minutes)	F_r (adim)
1	1.18	0.4260	0.03	0.4260
2	8.47	0.5278	0.20	0.5278
3	18.33	0.6255	0.45	0.6255
4	25.00	0.7174	0.82	0.7174
5	31.97	0.7521	1.12	0.7723
6	36.98	0.8250	1.25	0.8319
7	39.60	0.8565	1.58	0.8618
8	39.00	0.8883	1.30	0.8989
9	34.27	0.9201	0.88	0.9201
10	26.32	0.9462	0.68	0.9538
11	16.77	0.9697	0.42	0.9697
12	7.23	0.9942	0.13	0.9942
13	2.38	0.9971	0.02	0.9971
Total	287.5		8.88	

Table 8.2: Lake Okeechobee. Computational Time (T) to find the optimal subsets for all sizes and corresponding relative frequencies (F_r) using genetic algorithm and tabu search.

In sizes 5, 6, 7, 8 and 10, the F_r values obtained with genetic algorithm are smaller than the values obtained with tabu search, because the true optimal subset of the period was never a daily optimal subset. However, both sets of results are essentially the same. The difference in computational time, taking as reference the time corresponding to the process using a genetic algorithm is –278.62 minutes or –96.9 %. This difference is an important improvement when using tabu search instead of genetic algorithm in the case of lakes.

8.2 Lake Okeechobee's Results of the Fourth Methodology for Lakes

The modified VBA macro that supports the fourth methodology and which considers the objective function as the *average RMSE* was applied to the network of Lake Okeechobee (case study 8). The results of this optimization are shown in Tables 8.3 and 8.4. These results correspond to one run that specified 10 iterations after change of optimum and 9 iterations after change of size. The run made 137 iterations to find the last optimum value and 147 to end the process.

Size of subset	Accumulated number of function evaluations to optimum	Accumulated time to optimum (sec)	*avgRMSE* (ft)
1	15	1.9	0.157
2	52	5.8	0.141
3	217	30.3	0.125
4	375	56.4	0.107
5	650	106.1	0.095
6	936	160.3	0.082
7	1193	209.6	0.072
8	1481	264.2	0.059
9	1742	312.1	0.049
10	2061	367.5	0.041
11	2278	419.2	0.035
12	2429	422.4	0.026
13	2505	430.1	0.019
Total	2515	430.2	1.008

Table 8.3: Lake Okeechobee. Accumulated number of function evaluations to optimum, accumulated time to optimum and average RMSE for all sizes.

Table 8.3 contains the accumulated number of function evaluations to reach the optimum of each size, the accumulated computational time to the optimum and the respective optimum value (*avgRMSE*). The last row shows the total number of function evaluations, the total computational time and the sum of the *average RMSE* of all sizes. The total number of function evaluations constitutes 15.4 % of the cardinality (16382 subsets) of the problem. The time spent (430.2 sec = 7.17 min) in all the processes is of the same order as the time spent solving the same problem but considering relative frequency (Table 8.2). In a similar run but with four iterations after change of optimum and three iterations after change of size, the same optimal subsets were obtained in 62 iterations to the last optimum and 66 iterations to end the processes. The number of function evaluations to optimum was 1327 and 1331 to end the process. The total number of function evaluations constitutes 8.1 % of the cardinality of the problem. Finally, the total time was 214.5 sec = 3.58 min. Table 8.4 shows the stations that constitute all the optimal subsets.

	Size of subset												
STATIONS	1	2	3	4	5	6	7	8	9	10	11	12	13
S127_T		*						*	*	*	*	*	*
LZ40			*				*	*	*	*	*	*	*
CULV10A_H			*	*	*	*	*	*	*	*	*	*	*
S3_T				*	*	*	*	*	*	*	*	*	*
L_OKEE.M_G					*	*	*	*	*	*	*	*	*
S191_T								*	*	*	*	*	*
CULV5A_H						*	*	*	*	*	*	*	*
L006							*	*	*	*	*	*	*
S135_T									*	*	*	*	*
L005		*								*	*	*	*
S129_T											*	*	*
L_OKEE	*			*	*	*						*	*
S133_T													*
L001			*	*	*	*	*						

Table 8.4: Lake Okeechobee. Stations of the optimal subsets of all sizes. If a station is a member of a subset, an asterisk is shown in the intersection of the station's row and subset's column.

In addition, the results of the application to the Lake Okeechobee's network of a tabu search based program written to identify the optimal ordered list of stations are shown in Tables 8.5 and 8.6. Table 8.5 shows the triangular pattern that the selection of stations forms when the stations are ordered in selection order. The same order was used in Table 8.4 to compare the optimal subsets and the optimal ordered list of stations. Comparing the patterns of asterisks in Tables 8.4 and 8.5 one can conclude that the subsets of the optimal ordered list of stations with size = 1 to 7 are suboptimal, while the remaining subsets are equal to the corresponding optimal subsets and, therefore, are optimal.

	Size of subset												
STATIONS	1	2	3	4	5	6	7	8	9	10	11	12	13
S127_T	*	*	*	*	*	*	*	*	*	*	*	*	*
LZ40		*	*	*	*	*	*	*	*	*	*	*	*
CULV10A_H			*	*	*	*	*	*	*	*	*	*	*
S3_T				*	*	*	*	*	*	*	*	*	*
L_OKEE.M_G					*	*	*	*	*	*	*	*	*
S191_T						*	*	*	*	*	*	*	*
CULV5A_H							*	*	*	*	*	*	*
L006								*	*	*	*	*	*
S135_T									*	*	*	*	*
L005										*	*	*	*
S129_T											*	*	*
L_OKEE												*	*
S133_T													*
L001													

Table 8.5: Lake Okeechobee. Subsets members of the optimal ordered list of stations. If a station is a member of a subset, an asterisk is shown in the intersection of the station's row and subset's column.

The corresponding *average RMSE* of the subsets members of the optimal ordered list of stations are indicated in Table 8.6. The first row corresponds to station S127_T, the second row corresponds to the subset constituted by stations S127_T and LZ40, the third row corresponds to the subset formed by the three stations, and so on. A comparison can be established between the last columns of Tables 8.3 and 8.6. The ordered list's members 1 to 7 have a greater *average RMSE* than the optimal subsets of the corresponding size have. The values of average *RMSE* of members 1 to 7 are in average 4.9 % greater than those of the corresponding optimal subsets are. The Sum of the *average RMSE* of the ordered list is 3.5 % greater than that of the subsets. In conclusion, in Lake Okeechobee, the optimal ordered list of stations is not optimal if one takes into account the optimality of each one of its members.

Member	Stations	*avgRMSE* (ft)
1	S127_T	0.164
2	LZ40	0.145
3	CULV10A_H	0.126
4	S3_T	0.112
5	L_OKEE.M_G	0.102
6	S191_T	0.090
7	CULV5A_H	0.075
8	L006	0.059
9	S135_T	0.049
10	L005	0.041
11	S129_T	0.035
12	L_OKEE	0.026
13	S133_T	0.019
14	L001	0
	Sum avgRMSE (ft)	1.043

Table 8.6: Lake Okeechobee. Stations ordered according to how they were selected. The *average RMSE* corresponds to the subset members of the ordered list of stations.

8.3 KISSIMMEE RIVER'S RESULTS OF THE THIRD METHODOLOGY FOR STREAMS WITH STEADY FLOW

Case study 9, described in Chapter 3, was selected to be used in the development and testing of the third methodology for streams with steady flow. Many of the steps that were explained in Chapters 3 and 6 are still applicable to the optimization process of a stream using tabu search. Particularly, the calibration performed for the entire period (from 10/91/2001 to 12/31/2003), and shown in Chapter 6, was considered valid for this additional study. In the case of the optimization of the stage-monitoring network of streams with steady flow, the time to evaluate the objective function depends on the computation of water surface profiles. This computation depends on the hydraulic complexity of the studied stream. In this problem, it is of the greatest importance to limit the number of function evaluations performed when optimizing the stage-monitoring network. Tabu search performs a limited number of function evaluations. Using the VBA macro developed for this purpose, considering four iterations after change of optimum and three iterations after change of size, a series of results were found for the three periods of data presented in Chapter 3. Two periods are of 30 days and the third period comprises the 60 days of the first two periods. Tables 8.7, 8.8 and 8.9 show a summary of the results found. In these tables, the accumulated number of function evaluations to optimum, accumulated number of function evaluations to change size (or end process), *minimum RMSE*, *maximum RMSE*, *average RMSE*, *standard deviation* of *RMSE* and the number of valid days are indicated for the optimal subsets of all sizes. The total number of function evaluations (second number on row 19 of Tables 8.7 to 8.9) in the three different periods considered (901, 941 and 922) could be used to estimate which fraction of the cardinality of the problem they represent. The cardinality can be computed with $C = 2^{n-r_b} - 2$, where r_b is the number of stations that are required to hydraulically model the stream. In the stream of the case study n = 20 and r_b = 7. Therefore, the fraction of the cardinality can be computed with the *average of total number of function evaluations divided by the cardinality* = ((901 + 941 + 922) / 3) / (2^{20-7} – 2) = 921 / 8190 = 0.112 or 11.2 %. To estimate the computational time of a function evaluation for the stream of interest, it is necessary to apply the empirical equation $t_c = 7.3 \times$ valid days (seconds), valid for the Dell Workstation PWS360 used in this study.

Size	Num. Function Eval.	*minRMSE* (ft)	*maxRMSE* (ft)	*avgRMSE* (ft)	*StDvRMSE* (ft)	valid days
7	1	0.210	0.489	0.305	0.072	30
8	14, 18	0.196	0.338	0.262	0.041	30
9	43, 73	0.129	0.314	0.222	0.053	30
10	85, 142	0.136	0.279	0.205	0.037	30
11	153, 228	0.127	0.187	0.153	0.016	28
12	271, 346	0.104	0.173	0.129	0.019	28
13	391, 469	0.065	0.163	0.115	0.023	28
14	514, 589	0.060	0.143	0.099	0.022	28
15	596, 683	0.051	0.102	0.079	0.018	28
16	722, 773	0.012	0.102	0.064	0.029	28
17	778, 835	0.011	0.088	0.032	0.019	28
18	860, 882	0.001	0.094	0.030	0.024	28
19	897, 901	0.000	0.060	0.017	0.019	28

Table 8.7: Kissimmee River. Summary of solution of the available 30 days of daily data between 10/1/2001 and 6/15/2002.

As was mentioned in Chapter 6, all subsets have as elements the seven stations, S65A_H, WEIR3_H, WEIR2_H, WEIR1_H, S65C_H, S65D_H, and S65E_H, which are required to hydraulically model the studied reach of Kissimmee River. In Table 8.7, the main results of the first period of 30 days are shown. These results needed 52 iterations to find the last optimum and 56 to end the process. The estimated computational time of this process is $t_c = 142 \times 7.3 \times 30 + (901 - 142) \times 7.3 \times 28 = 186{,}238$ sec = 3,104 min = 51.73 hr = 2.16 days. The analysis of the results of this period showed that the identified optimal subsets constitute an ordered list of stations. The subset of size = 8 is {S65A_H, WEIR3_H, WEIR2_H, WEIR1_H, S65C_H, S65D_H, S65E_H and WEIR1_T}.

The subset of size = 9 is {S65A_H, WEIR3_H, WEIR2_H, WEIR1_H, S65C_H, S65D_H, S65E_H, WEIR1_T and S65D_T}. The remaining subsets are formed by successively adding the stations PC33, KRDRS, C38BAS, S65A_T, KRBNS, WEIR2_T, S65C_T, PC11R, S65_T, KRFNS and WEIR3_T.

Size	Num. Function Eval.	*minRMSE* (ft)	*maxRMSE* (ft)	*avgRMSE* (ft)	*StDvRMSE* (ft)	valid days
7	1	0.269	2.579	1.074	0.812	30
8	14, 18	0.260	1.197	0.643	0.301	30
9	31, 64	0.217	0.798	0.483	0.160	30
10	116, 167	0.200	0.491	0.299	0.073	30
11	178, 253	0.210	0.259	0.239	0.016	30
12	296, 371	0.130	0.208	0.178	0.025	30
13	416, 494	0.062	0.180	0.140	0.035	30
14	539, 614	0.067	0.154	0.123	0.021	30
15	621, 708	0.068	0.113	0.089	0.012	30
16	714, 789	0.040	0.062	0.053	0.006	30
17	822, 868	0.019	0.044	0.030	0.006	30
18	893, 923	0.006	0.041	0.022	0.010	30
19	926, 941	0.001	0.057	0.022	0.016	30

Table 8.8: Kissimmee River. Summary of solution of the 30 days of daily data between 6/16/2002 and 7/15/2002.

In Table 8.8, the results of the second period of 30 days are shown. These results also needed 52 iterations to find the last optimum and 56 to end the process. The estimated computational time of this process is $t_c = 941 \times 7.3 \times 30 =$ 306,079 sec = 3,435 min = 57.24 hr = 2.39 days. The optimal subsets of this period also form an ordered list of stations. The ordered list is S65A_H, WEIR3_H, WEIR2_H, WEIR1_H, S65C_H, S65D_H, S65E_H, KRBNS, WEIR1_T, PC33, KRDRS, S65D_T, C38BAS, PC11R, WEIR2_T, S65A_T, S65C_T, WEIR3_T, S65_T, and KRFNS.

Size	Num. Function Eval.	*minRMSE* (ft)	*maxRMSE* (ft)	*avgRMSE* (ft)	*StDvRMSE* (ft)	valid days
7	1	0.210	2.579	0.690	0.693	60
8	14, 18	0.213	1.197	0.471	0.277	60
9	31, 55	0.197	0.798	0.376	0.158	60
10	88, 139	0.200	0.491	0.266	0.062	60
11	178, 242	0.127	0.488	0.218	0.080	60
12	285, 361	0.128	0.208	0.164	0.025	58
13	406, 484	0.062	0.180	0.130	0.029	58
14	529, 604	0.043	0.179	0.117	0.036	58
15	611, 698	0.046	0.157	0.093	0.032	58
16	704, 779	0.040	0.094	0.059	0.010	58
17	812, 858	0.011	0.088	0.031	0.014	58
18	862, 904	0.006	0.102	0.029	0.018	58
19	907, 922	0.000	0.060	0.026	0.021	58

Table 8.9: Kissimmee River. Summary of solution of the available 60 days of daily data between 10/01/2001 and 7/15/2002.

In Table 8.9, the results of the period of 60 days are shown. These results needed 51 iterations to find the last optimum and 55 to end the process. $t_c = 242 \times 7.3 \times 60 + (922 - 242) \times 7.3 \times 58$ = 393,908 sec = 6,565 min = 109.42 hr = 4.56 days. The optimal subsets obtained from this period do not form a complete ordered list of stations. The truncated ordered list is S65A_H. WEIR3_H, WEIR2_H, WEIR1_H, S65C_H, S65D_H, S65E_H, KRBNS, WEIR1_T, PC33, S65D_T, KRDRS, C38BAS, S65A_T, WEIR2_T, PC11R, S65C_T and WEIR3_T. The subset of size = 19, drops station WEIR3_T and adds stations S65_T and KRFNS. In Table 8.10, the ordered lists of stations for the three periods are shown. The assigned colors make easer to see the differences, for example, station S65D_T is the ninth element of the first list, the twelfth of the second list, and the eleventh of the third list.

Member	First period of 30 days	Second period of 30 days	Period of 60 Days
1	S65A_H	S65A_H	S65A_H
2	WEIR3_H	WEIR3_H	WEIR3_H
3	WEIR2_H	WEIR2_H	WEIR2_H
4	WEIR1_H	WEIR1_H	WEIR1_H
5	S65C_H	S65C_H	S65C_H
6	S65D_H	S65D_H	S65D_H
7	S65E_H	S65E_H	S65E_H
8	WEIR1_T	KRBNS	KRBNS
9	S65D_T	WEIR1_T	WEIR1_T
10	PC33	PC33	PC33
11	KRDRS	KRDRS	S65D_T
12	C38BAS	S65D_T	KRDRS
13	S65A_T	C38BAS	C38BAS
14	KRBNS	PC11R	S65A_T
15	WEIR2_T	WEIR2_T	WEIR2_T
16	S65C_T	S65A_T	PC11R
17	PC11R	S65C_T	S65C_T
18	S65_T	WEIR3_T	WEIR3_T
19	KRFNS	S65_T	
20	WEIR3_T	KRFNS	

Table 8.10: Kissimmee River. Summary of the ordered list of stations of the available 60 days of daily data between 10/01/2001 and 7/15/2002.

Figure 8.4 shows the *average RMSE* of the optimal subsets found for the three different periods studied. In the figure, one can observe that the errors of the period of 60 days are between the errors of the other two periods.

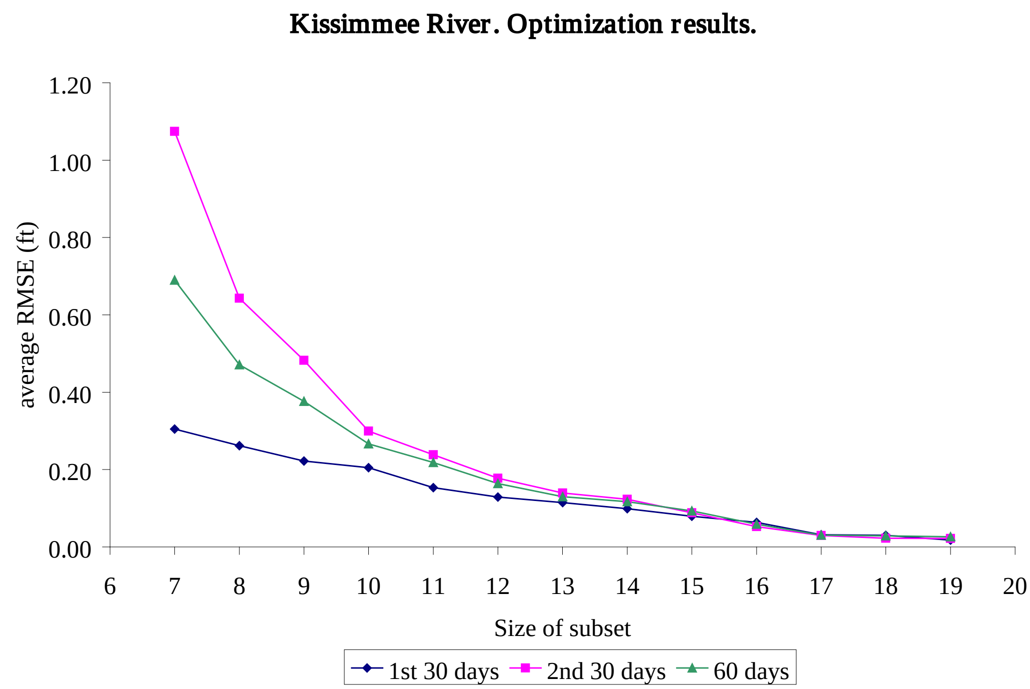


Figure 8.4: Kissimmee River. Network optimization using three different periods of data.

8.4 Concluding Remarks

The results of the application to the Lake Okeechobee's network (case study 8) of tabu search based procedures demonstrate that tabu search had a better performance than genetic algorithm. The problems solved were the identification of the optimal subsets of all sizes given *admissible RMSE* and *reliability* using the third methodology for lakes; the optimal subsets of all sizes given *admissible RMSE* using the second methodology for lakes; the subsets of minimum *average RMSE* using the fourth methodology for lakes; and the optimal ordered list of stations. The third methodology for streams with steady flow, based on tabu search, was applied in the optimization of the stage-measuring network of Kissimmee River in Pool AE (case study 9). The results of this optimization allow to conclude that the limited number of function evaluations can make feasible the optimization of the stage-measurement network of Kissimmee River using the available daily data of 428 days. Additionally, the results show that for longer periods, the optimal subsets may tend to form ordered list of stations. In any case, these results also show that forward selection may give results no too distant from the optimal results defined by tabu search.

At this point of the research, based on the performance in the solution of their respective problems, it is possible to adopt three of the studied methodologies. For lakes, the fourth methodology can be adopted, and to distinguish it from the other proposed methodologies can be called the Lake Okeechobee Stage-monitoring Network Optimization methodology or the LOSNO methodology. For streams with steady flow, the third methodology can be adopted and called the Kissimmee River Stage-monitoring Network Optimization methodology or, using an acronym, the KRSNO methodology. The second methodology for streams with unsteady flow offers, before creating a third methodology based on tabu search, a provisional approach to solve its problem; for this methodology, no name is proposed.

Chapter 9: Conclusions and Recommendations

9.1 INTRODUCTION

The results of this dissertation address the three research questions posed in Chapter 1. Which are the most important stations of the network or which of the stations can be closed without serious loss of information? Which is the best methodology to answer any of the alternative forms of the previous question? Where should be the stations located? Several optimization methodologies that answer the first question were proposed and three of them were adopted. The adopted methodologies are the Lake Okeechobee Stage-monitoring Network Optimization (LOSNO) methodology (the fourth methodology for lakes), the Kissimmee River Stage-monitoring Network Optimization (KRSNO) methodology (the third methodology for streams with steady flow), and the unnamed second methodology for stage-monitoring networks in streams with unsteady flow. The second question is answered with the study of methodologies based in three different optimization methods. The third question is answered with the optimization guidelines proposed in Section 9.5. The three adopted methodologies can be viewed as components that may be applied jointly with the optimization guidelines, especially for the case of existing networks. Conclusions that concern the research questions are discussed in Section 9.2. Conclusions drawn about the methodology for lakes are discussed in Section 9.3. Conclusions about the methodologies for streams are discussed in Section 9.4. The optimization guidelines are discussed in Section 9.5. Finally, recommendations for future work are in Section 9.6.

9.2 RESEARCH QUESTIONS

The three research questions considered in this research can be grouped in two groups, one that considers an existing network:

- Which are the most important stations of the network or which of the stations can be closed without serious loss of information?
- Which is the best methodology to answer the previous question?

And other that considers a new network:

- Where should be the stations located?

Five assumptions were proposed on which to base the answer to the research questions:

- A subset of the measuring stations of a lake or stream and a model are adequate to estimate the water elevation in the remaining station locations.
- Historical data, measured in the same period in all the stations of a lake or stream can be used to optimize its measuring network.
- The application of an interpolation method and of an optimization method provides the necessary and sufficient tools to optimize the stage-measuring network of a given lake or stream.
- In the case of a lake, a geointerpolation method can be used to estimate water surface elevations.
- In streams, the computation of the water profile constitutes an adequate interpolation method.

The entire set of assumptions were fulfilled; upon them, several optimization methodologies were built and three of them were adopted. The adopted methodologies are the fourth methodology for stage-monitoring networks in lakes, the third methodology for stage-monitoring networks in streams with steady flow, and the second methodology for stage-monitoring networks in streams with unsteady flow. The methodologies quantitatively answer the first question. That is, in general, from the set of n stations, S_1, S_2, S_3, S_4, S_5, …, S_n, a subset of r stations is chosen in a manner that the loss of information is minimal. This subset is called the optimal or base subset. In all the proposed methodologies, the quantification of information loss is based on a function of the root mean square error (*RMSE*) of estimation of water elevation. The most important stations are those that remain in the optimal subset. The location of both the most important stations and the stations that can be closed without a serious loss of information are implicitly obtained with the methodologies devised. Originally, two versions of the root mean square error were needed: one defined by measurements in all the stations in a given day or instant (*RMSE*) and the other defined by measurements in one station for the entire period of optimization (*RMSESt*). The use of *RMSESt* was adopted in the first and second methodology for streams and was suggested by the results of the second methodology for lakes, where statistical properties of the *RMSE* of all the days of the period of optimization were needed to identify the optimal subset of stations.

The second question is answered with the study of three different optimization methods as part of the methodologies. In the early stage of this work, simulated annealing and genetic algorithm were tested. They were compared according to several criteria in which the computational time and the ability to arrive to the true optimum were considered. The performance in two initial study cases (case studies 1 and 2) of the genetic algorithm based program was better than that of the simulated annealing program; consequently, the genetic algorithm was adopted to be used in the completion of the first methodology for lakes. Posterior testing of the genetic algorithm implementation in a case study that involved the consideration of time series of daily stage data (case study 3) made necessary, to obtain always true optimal values, a modification of the optimization routine. The adopted modification was to repeat the optimization of each size of subset and day until a result was repeated three times. That result was considered the true optimum. The price to pay for having a more robust routine was the excessive growth of the number of function evaluations, making the routine inefficient. The modified routine was adopted in the first and second methodologies for lakes, because it was thought that was still useful for that case. The inefficiency made the genetic algorithm prohibitive to be implemented in the case of streams, where each function evaluation can consume several seconds of computing time. In the case of streams with steady or unsteady flow, the optimization method adopted was forward selection. In summary, the originally proposed first and second methodologies use, for lakes, genetic algorithm, and for streams forward selection. An additional study performed after the proposal of the first and second methodologies shown the advantage of using tabu search in both, lakes and streams network optimization, because its robustness and efficiency.

The additional study also demonstrated the convenience of using as objective function the *average RMSE* instead of *probability of no exceedence* or *relative frequency* in the case of lakes, obviating the use of different *admissible RMSE* values and making possible the optimization of the network in only one run. The third and the LOSNO methodologies for lakes are based on tabu search. The third uses as its objective function the probability of no exceedence of *RMSE* and the LOSNO uses the *average RMSE*. The adoption of tabu search also allowed the use of the *average RMSE* as objective function instead of *RMSESt* in the case of streams given origin to the KRSNO methodology for streams with steady flow.

The third question is answered with the optimization guidelines. Although, the guidelines can be used to optimize existing networks their main emphasis is on the optimization of new networks, in which the location of new stations is proposed according to the hydrologic or hydraulic component that must be monitored. So far, numerical criteria for establishing new stations have not been obtained; the guidelines are qualitative not quantitative. Experience with more optimizations of stage-monitoring networks is needed to convert the guidelines from qualitative to quantitative.

9.3 Optimization of Stage-Monitoring Networks in Lakes

Four methodologies for lakes are proposed in this work. The first and second are based on the genetic algorithm method and the third and fourth (LOSNO) are based on the tabu search method. Only the second methodology for lakes (an improvement of the first methodology for lakes) is thoroughly explained and even a manual of the accompanying tools is presented in the appendix.

Neither explicit implementation nor tools' manual are presented for the third and LOSNO methodologies, because these methodologies share many features with the second and only the details that differ to those of the second methodology are explained.

Based on the corresponding assumptions, the core of the second methodology for lakes is the identification of daily optimal subsets of stations. Each day a genetic algorithm is launched for finding the least root mean square error in the estimation of water surface elevations in those stations that are outside the subset that is being tested. The estimation of water elevations is computed using inverse square distance weighting interpolation, a simple but adequate method for the generally smooth variation of water surface in lakes. The optimization is repeated for all the days of a given period. It is a prerequisite to have data of a stable set of stations for at least two years. In this context, stable means that the set should have the chosen stations operating in unchanged conditions for most of the analyzed period. At the end, only daily optimal subsets that have root mean square errors smaller than or equal to an admissible root mean square error (i.e. 01. ft) previously given are considered to be ranked. Then, the daily optimal subset that most frequently has in the period a root mean square error smaller than or equal to the admissible value, is considered the optimal subset. In essence, the optimization results should have a probabilistic or statistical interpretation because even the optimal subset of the period can have root mean square errors greater than the admissible *RMSE* in several days of the period. The application of the second optimization methodology for lakes to the stage-monitoring network of Lake Kissimmee (case 4) determined that three stations (LKISS9, LKISS7, and S65_H) of a set of four stations (LKISS9, LKISS7, KISS5B, and S65_H) form the optimal subset of the analyzed period (10/01/2001 to 9/30/2003). The admissible *RMSE* is considered equal to 0.1 ft. The empirical frequency of *RMSE* ≤ admissible *RMSE* of this subset is 0.8890 and the corresponding probability *P*(*RMSE* ≤ admissible *RMSE*), assuming an *RMSE* log-normally distributed, is equal to 0.8648. That is, 86.5 % of the time *RMSE* is smaller than or equal to 0.1 ft. If the value of the admissible *RMSE* increases, the *P*(*RMSE* ≤ admissible *RMSE*) value increases. Furthermore, if the admissible *RMSE* value is sufficiently increased, new optimal subsets with fewer stations should appear. However, the increase of admissible *RMSE* must be limited because a large *RMSE* could render useless the estimation of water elevation obtained with the remaining operating stations.

The LOSNO methodology for lakes uses tabu search to identify the subsets of stations of all sizes that have a minimum *average RMSE* computed with daily values of the entire period in study. The application of this methodology to the network of Lake Okeechobee, with 14 stations (case 8), which is larger than that of Lake Kissimmee, allowed the identification of the optimal subsets of all sizes in only one run and having to compute a very limited number of *average RMSE* values to find the minimum value for each size. With the data of Lake Okeechobee, the optimization method only needed to compute 8.1 % of the total number of solutions. Additionally, it was demonstrated for the case of Lake Okeechobee, and in general, that an optimal ordered list of stations could not be defined because, to be optimal, the list needs to have as members the optimal subsets of all sizes. Moreover, the optimal subset of a given size not necessarily has the same stations than the optimal subset of the previous size and an additional station.

The implementation of these procedures in a series of VBA macros inside ArcMap, implies the linkage of attributes of geographic features and associated time series with an optimization process, and demonstrates the power of the integration of GIS and customized routines. The VBA network optimization macros have made the process of optimizing the stage-monitoring networks in lakes a direct and user-friendly task. The time consuming part is the retrieval, preparation and loading of the data in the Arc Hydro geodatabase used in the toolset created.

9.4 Optimization of Stage-Monitoring Networks in Streams

Based on the appropriate assumptions, the nucleus of the second methodology for streams with steady flow and the second for streams with unsteady flow is the identification of the station that has the maximum *RMSESt* for the specified times of a given period. The station with the largest *RMSESt* is the worst modeled station, i.e. its stage measurements are the least explained by the model and for that reason this station is a good candidate to enter into the base subset of stations to be retained. Steady and unsteady flows are considered in separate, but closely related optimization methodologies. The main difference between the second methodologies for streams, for the sake of the posterior computation of errors, is that in the case of steady flow, water elevations are computed for valid days of a given period of at least two years. While in the case of unsteady state, water elevations are computed for specific instants of a much shorter time of simulation, on the order of several days. A valid day is a day that has data for all the stations that are part of the base subset. The interpolated values are provided by HEC-RAS and the optimization method is *forward selection*, a simple sequential method. The interpolation provided by a mathematical formula was not adequate to represent water surface profiles along streams because the water surface in a stream is not a smooth surface without discontinuities; it is a complex surface governed by the laws of hydrodynamics and affected by changes of cross-sections, the presence of control structures and the variation of circulating flows. A hydrodynamic model was needed to provide the interpolated values. The genetic algorithm approach was abandoned because the time needed to compute water elevations and then their root mean square error (the objective function) is several orders of magnitude greater than the time needed for the same purpose in the case of lakes. The optimization method was reduced to a method that searches for the station with maximum *RMSESt*, puts it into the base subset and launches the next cycle of optimization. This method is a sort of empirical search of the optimum of a combinatorial problem.

Complications emerge because a process of calibration of the HEC-RAS model of the stream must be performed, first with a limited dataset, representative of the stage and flow variations encountered in the considered period, and later with all the data of the period. The calibration seeks to match measured with computed water elevations. Once the calibration is finished, the optimization can begin. The stations that set the stages that hydraulically model the stream are considered the initial base subset of stations. In the case of the optimization of the network of the main stream of Pool AE (case studies 5, 6, and 7), the initial base subset of (seven) stations that are required to hydraulically model the stream is S65A_H, WEIR3_H, WEIR2_H, WEIR1_H, S65C_H, S65D_H, and S65E_H. This base subset is equally valid for steady and for unsteady flow. In natural coastal streams or, in general, in streams with mild bottom slope, water flows in a subcritical state. In steady flow, a subcritical state requires a downstream boundary condition in each reach defined between hydraulic structures. In unsteady flow, two boundary conditions are needed for each reach. In a subcritical state, a stage downstream condition is valid while a discharge condition can be set at the upstream end of the reach. Commonly a stage-measuring station is installed in the head side of each structure. Head side stations are in consequence the stations that set the stages for the hydraulic model of the stream and thus form the initial base subset. With the first base subset of stations and after running the HEC-RAS procedure that computes water elevations for all the valid days (or specified instants, in the unsteady flow case) of the period, the root mean square error in each station outside the base subset is computed. The station that has the largest root mean error of its estimated stages is entered in the base subset. The new base subset is used to estimate the water elevations of the remaining outside stations. The root mean square error for each station is then computed. Again, the station that has the largest root mean square error goes into the base subset. This iterative process is continued until a certain number of stations in the base subset is reached or an admissible root mean square error is reached. For steady flow (case study 5), 12 of 20 stations are selected to form the optimal base subset; they are S65A_H, WEIR3_H, WEIR2_H, WEIR1_H, S65C_H, S65D_H, S65E_H, WEIR1_T, KRDRS, PC33, KRBNS, and S65D_T. The average *RMSE* is equal to 0.171 ft, standard deviation of *RMSE* equal to 0.032 ft and *RMSESt* maximum equal to 0.263 ft. If an admissible *RMSE* of 0.2 ft is adopted, 82 % of the time the base subset of 12 stations should estimate appropriate *RMSE* values.

Both second methodologies for streams work well, but there are some issues in the case of unsteady flow. In the case of steady flow, having data with a sufficient number of valid days, it is possible to identify the best base subset of each size, even with analytically obtained channel geometries. However, in the case of unsteady flow it is more difficult to identify the best subset of each size because there can be many uncertainties in the data and parameters used. In principle, simulating unsteady flow is much more difficult than simulating steady flow. In the specific case studied, the hydrodynamic behavior of water in the naturalized reach of the stream may be considered not physically sound; water elevations always oscillate even after long periods of constant inflow. Several causes might be responsible for this behavior. One possible cause arises from the setting of internal boundary conditions using measured stages and flows that themselves may have errors and that in combination do not satisfy the equations of motion. Other causes might be that the analytic cross sections of this reach are not completely adequate for modeling unsteady flow. Another possible cause of these results is the use of daily values of stage and flow as instantaneous values at specific times and the assumption of their linear variation between those "instantaneous values". There is no real unsteady variation of measured stages and flows. In summary, it is possible and practical to optimize the stage-monitoring network along streams using steady flow. Optimization using unsteady flow is much more difficult and its validity is, so far, tentative.

The KRSNO methodology, based on tabu search, is proposed for the case of streams with steady flow. The number of objective function evaluations required is reduced to a fraction of the cardinality of the problem, in the Kissimmee River (case study 9), is 0.112, and allows the rejection of the forward selection method. In this methodology, the *average RMSE* is adopted as objective function. Results of the application of this new methodology to the Kissimmee River reach located in the Pool AE basin and for three limited periods (two of 30 days and one of 60 days) encourage the application of the methodology using data of the entire available period of 428 days. The computational time of the optimizations using 30 days was of the order of 55 hours, while the time of the optimization using 60 days was around 110 hours. Since this process is not a real-time process, it is possible to run it using the entire period of data. Additionally, the existence of optimal ordered lists of stations was confirmed for the two periods of 30 days and was not present for the period of 60 days. It seems that the existence of ordered list of stations is more likely to appear in streams than in lakes, it may be caused by the separated flow segments that isolate the stations along the streams, compared to the relative liberty that all stations have in a lake.

As in the case of lakes, the implementation of each one of these procedures in a VBA macro and a VB procedure inside ArcMap have made the process of optimizing the stage-monitoring networks in streams with steady and unsteady flow, a direct and user-friendly task. The retrieval, preparation and loading of the data in the tables used in the toolsets created is not time consuming. The time consuming aspect is the calibration process needed before performing the optimization. A remarkable aspect of the created procedures is the coupling of the customized VBA code with HEC-RAS. A continuous flow of data moves from ArcMap to HEC-RAS and then from HEC-RAS to ArcMap. This data includes, apart from the specifications of changing base subsets, the time dependent values of water elevations and flows stored in/read from customized time series tables accessed by the GIS software to store in/read from text files accessed by HEC-RAS. This coupling obviated the need to write or adapt complex algorithms present in this model or in similar models. A theoretical-practical consequence of the creation of this methodology is that it corroborates the possibility of establishing links between hydrodynamic models and GIS and modestly advances the creation of Hydrologic Information Systems (HIS).

9.5 OPTIMIZATION GUIDELINES

The set of proposed guidelines may be used in the process of optimizing an existing network or in designing a new network. In the case of an existing network, the rules serve to eliminate certain combinations of stations; and in the case of a new network, they serve to locate probable sites to establish stations. These rules are different for streams and lakes:

Lakes, specifically large ones, such as Lake Okeechobee, can have several stage-monitoring stations and one or more of them can be in excess. Try to keep pairs of stations located at the shore of a lake in opposite sides. Keep at least two pairs –with an imaginary line of connection that passes as close as possible by the center of the lake, – one in the north-south direction and another in the east-west direction; or in general, in the direction of predominant winds and in its transversal direction. If the lake is big enough, also keep a station as close as possible to its center. A preferable minimum number of stage stations in a lake should be two, one in an inflow structure and one in an outflow structure. If there are not flow structures, a station at the center of a small lake should be sufficient.

In the case of ***streams***, the presence of control structures, which substantially modify water elevations, must be taken into account. The base subset of stations used to start the optimization process should consider the head station of all the control structures inside the studied reach. In pool level stream reaches, the head station of the reach is sufficient to obtain water levels. In rapidly changing slope and/or cross-section reaches, one station may be needed for each sub-reach with different slope/cross-section combination.

9.6 RECOMMENDATIONS FOR FUTURE WORK

More experience in the use of the optimization of stage-monitoring networks in lakes and in streams methodologies is needed to see what modifications or refinements can be done on them. The same can be said of the optimization guidelines proposed because a limited set of results make difficult the recommendation of quantitative guidelines such as stations per square mile of lake's water surface, or stations per mile of stream, et cetera. In the case of optimization of stage-monitoring networks in streams with unsteady flow, what is needed is to further study the methodology with better geometric and hydraulic data. Adequate topographic surveys of streams and real instantaneous values of stage and flows are needed. Additionally, the use of tabu search in a new methodology should be considered.

With respect to the computational tools, what is already identified is the need of tools to automate the management of data, from retrieval to loading the data structures used in the tools. Other improvements that can be done in the case of the tools for lakes are related with the file format of results; currently they are stored in easily corruptible text files that, additionally, lack the possibility to be used directly in relational databases. Furthermore, the creation of compiled files, such as DLL files, to implement the toolsets is an improvement that should increase the tools' portability, so that programming code no longer needs to be stored in Arc Map documents in order to be distributed.

In the case of the optimization of stations in lakes, a possibility that could be studied is the use of more than one spatial interpolation method. Even when it is not expected that their consideration could change the optimization results, it can improve the confidence on results. An improvement that may be necessary to the toolsets for the optimization of stations along streams depends on the study of dendritic streams. So far, the toolsets have been tested only in a one-branch stream. Another improvement of the tools for streams can be the automation or at least the semi-automation of the calibration process. Automation of any of the processes could be of great help for the analyst, speeding up the whole optimization process.

The applicability of the methodologies for optimization of stage-monitoring networks in lakes and in streams with steady flow resides on the applicability of the interpolation methods. It means in the case of lakes that if a fairly smooth plane water surface is present, or if currents do not spatially change the water elevations, the tools are applicable. Of course, any lake that is suspected of having too many stage stations could benefit from this methodology. In the case of streams, the limit is the applicability of HEC-RAS. Until now, these methodologies only have been applied to a small region of SFWMD.

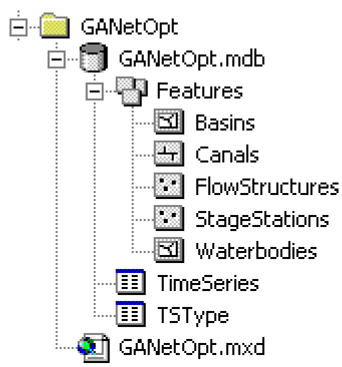
To know the water stages in the hydrologic and hydraulic components of a district is the first step to knowing the volume stored in each one of these components and even more, to knowing how much water is moving from one component to the next. Stage can be used to compute storage in components and flows between components. Storage in components and flows between components can be used to compute water balances. Water balances can help to know how much water is moving inside the district or how much water is entering or leaving the district. An even better use of water balances is to help the district to control the movement of water in a manner consistent with its own objectives. This constitutes the outline of a much wider scope of work in the future.

Appendix A

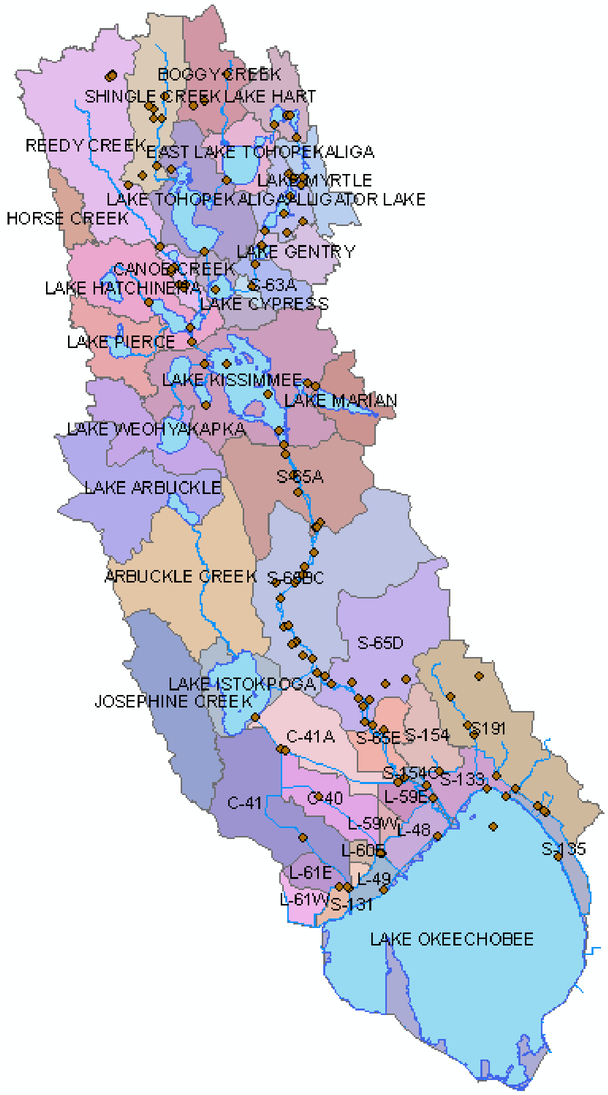
NETWORK OPTIMIZATION TOOLS USERS MANUAL

Network Optimization in Lakes Toolset

The purpose of this section is to show how to use the Network Optimization in Lakes toolset. The optimization of the stations of Lake Kissimmee is used as an example. The files necessary to use the NetOpt Toolbar are in file GANetOpt.zip located at ftp://ftp.crwr.utexas.edu/pub/outgoing/smartinez/, and are shown below:

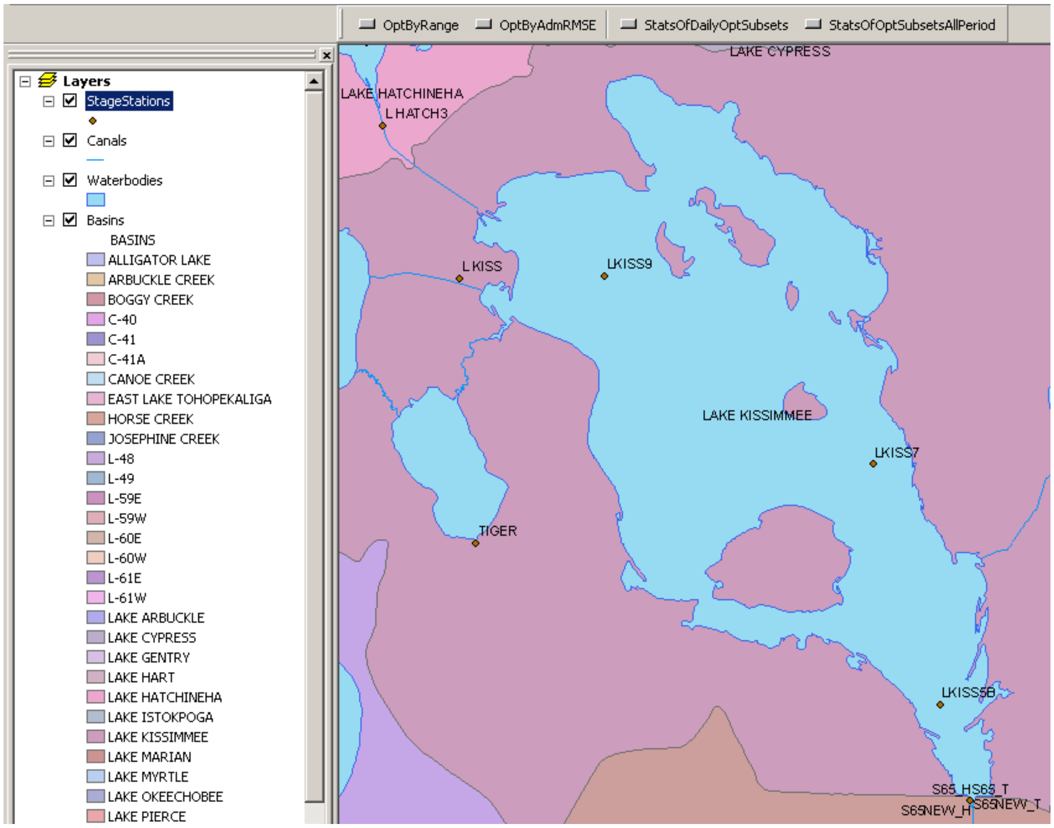


GANetOpt.mdb is a geodatabase with several feature classes stored in the Features dataset. The TimeSeries table contains daily water elevations measured for active stage stations in the study area from 10/01/2001 to 9/30/2004. GANetOpt.mxd is an ArcMap document with the NetOptLakes toolbar, its associated VBA code and geographic features from GANetOpt.mdb. Additionally, several files with the results of the exercise are included. The following figure shows the data from GANetOpt.mdb.

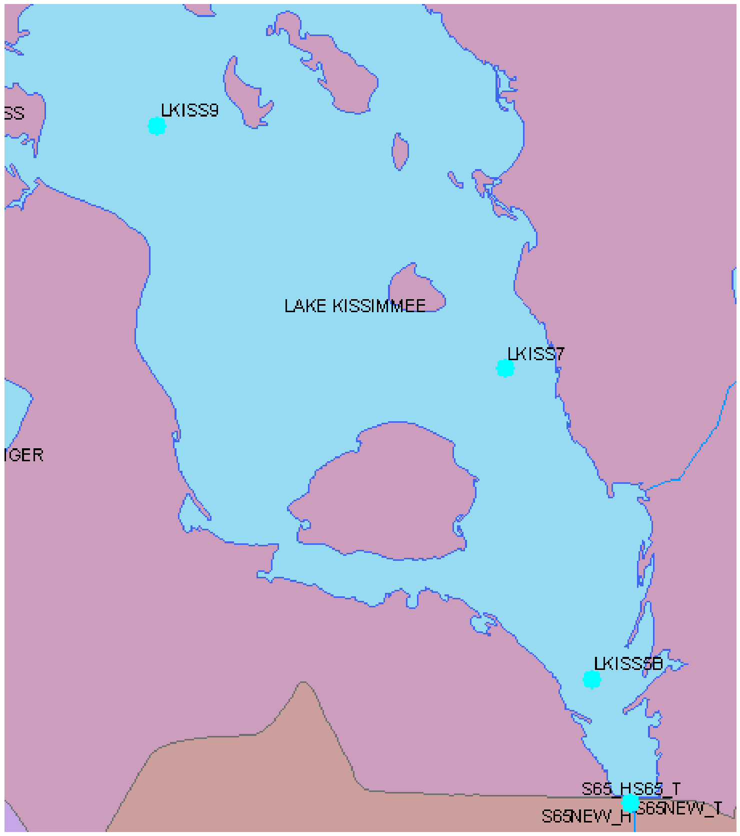


Getting Started

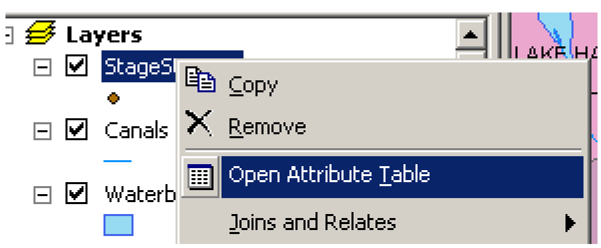
To use the Network Optimization in Lakes toolbar, unzip GANetOpt.zip and copy both NetOpt.mdb and GANetOpt.mxd in one folder on your local hard drive (For example, c:\GANetOpt). Open ArcMap, and open GANetOpt.mxd.



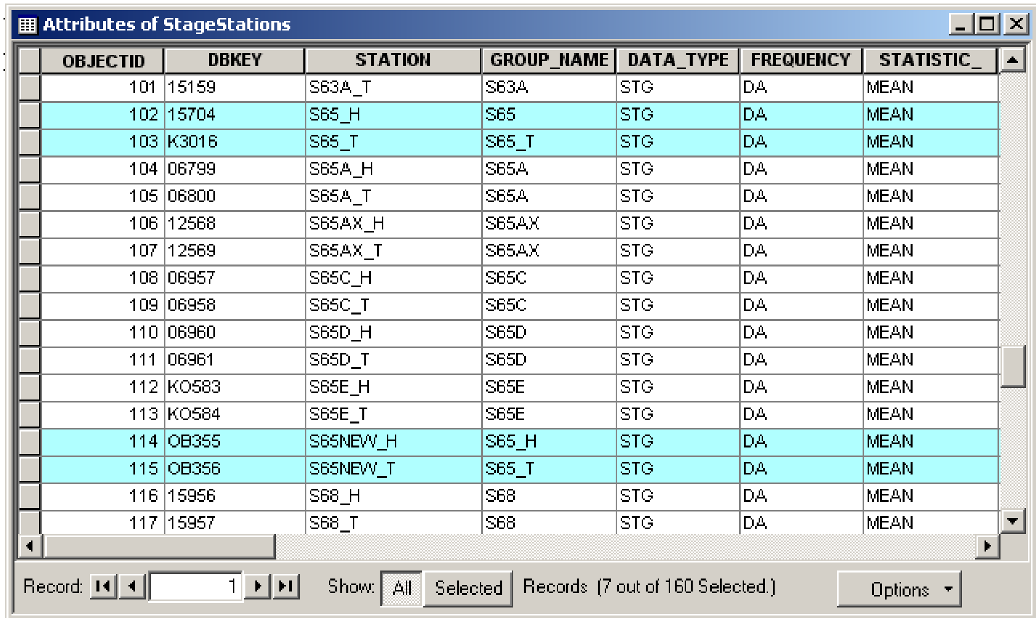
Click on the Select Features tool , and select LKISS7 station. Now press the **Shift** button, and select LKISS5B, LKISS9 and S65_H stations as shown below:



Since S65_H, S65_T, S65NEW_H and S65NEW_T lie at the same location; four features are selected when S65_H is selected. Therefore, we need to unselect stations S65_T, S65NEW_H and S65NEW_T. To unselect the undesired stations, right click on the StageStations layer, and then select **Open Attribute Table** as shown below:



Scroll the table to see the selected features as shown below:

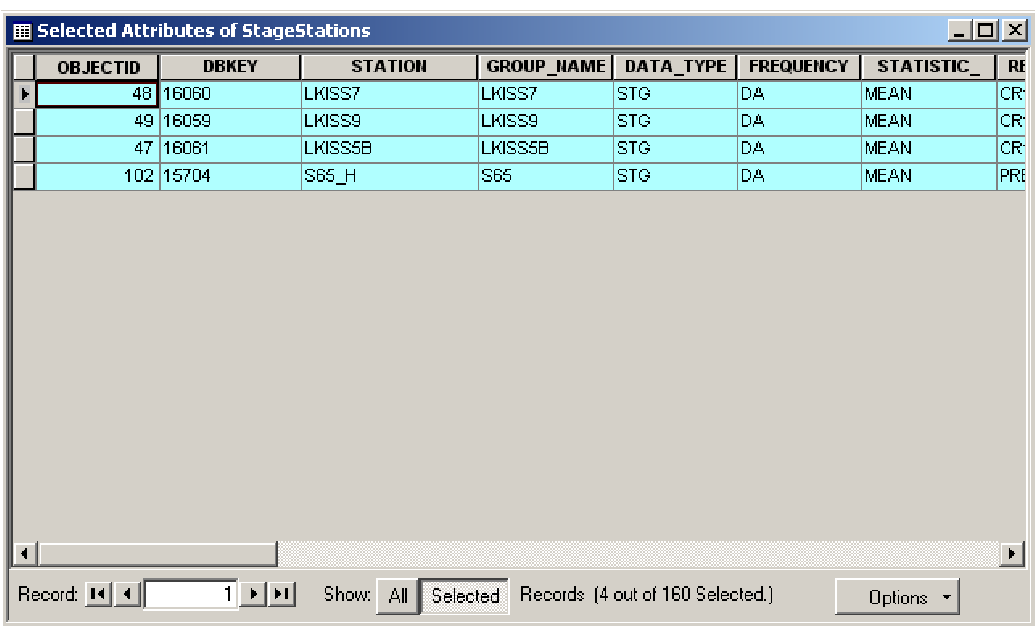


Attributes of StageStations

OBJECTID	DBKEY	STATION	GROUP_NAME	DATA_TYPE	FREQUENCY	STATISTIC_
101	15159	S63A_T	S63A	STG	DA	MEAN
102	15704	S65_H	S65	STG	DA	MEAN
103	K3016	S65_T	S65_T	STG	DA	MEAN
104	06799	S65A_H	S65A	STG	DA	MEAN
105	06800	S65A_T	S65A	STG	DA	MEAN
106	12568	S65AX_H	S65AX	STG	DA	MEAN
107	12569	S65AX_T	S65AX	STG	DA	MEAN
108	06957	S65C_H	S65C	STG	DA	MEAN
109	06958	S65C_T	S65C	STG	DA	MEAN
110	06960	S65D_H	S65D	STG	DA	MEAN
111	06961	S65D_T	S65D	STG	DA	MEAN
112	KO583	S65E_H	S65E	STG	DA	MEAN
113	KO584	S65E_T	S65E	STG	DA	MEAN
114	OB355	S65NEW_H	S65_H	STG	DA	MEAN
115	OB356	S65NEW_T	S65_T	STG	DA	MEAN
116	15956	S68_H	S68	STG	DA	MEAN
117	15957	S68_T	S68	STG	DA	MEAN

Record: 1 Show: All Selected Records (7 out of 160 Selected.) Options

To unselect S65_T, press Ctrl, and click on the row containing S65_T. Repeat the same to unselect S65NEW_H and S65NEW_T. After unselecting these three stations, you can click on the **Selected** button to change the view of the table to show only the selected features as shown below:



Selected Attributes of StageStations

OBJECTID	DBKEY	STATION	GROUP_NAME	DATA_TYPE	FREQUENCY	STATISTIC_	RE
48	16060	LKISS7	LKISS7	STG	DA	MEAN	CR
49	16059	LKISS9	LKISS9	STG	DA	MEAN	CR
47	16061	LKISS5B	LKISS5B	STG	DA	MEAN	CR
102	15704	S65_H	S65	STG	DA	MEAN	PRE

Record: 1 Show: All Selected Records (4 out of 160 Selected.) Options

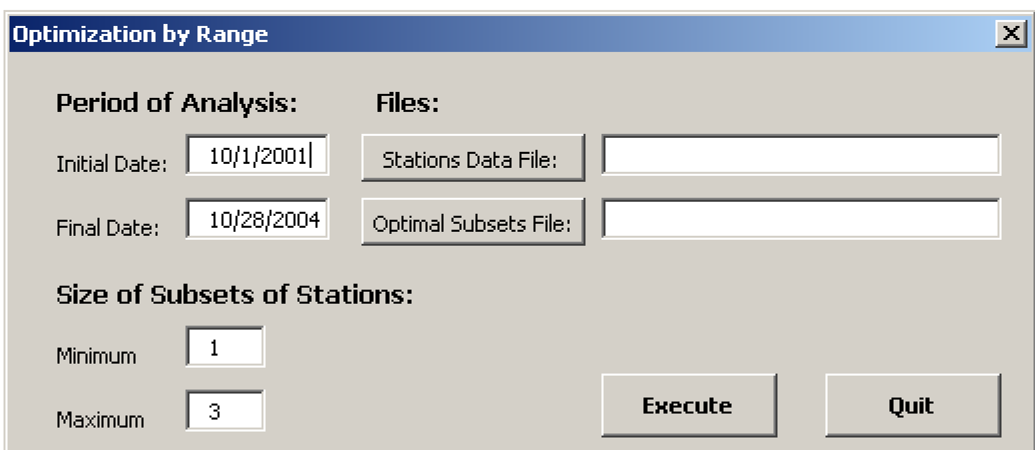
Close the attribute table for StageStations. Now we are ready to use the NetOpt toolbar.

Network Optimization in Lakes Toolbar

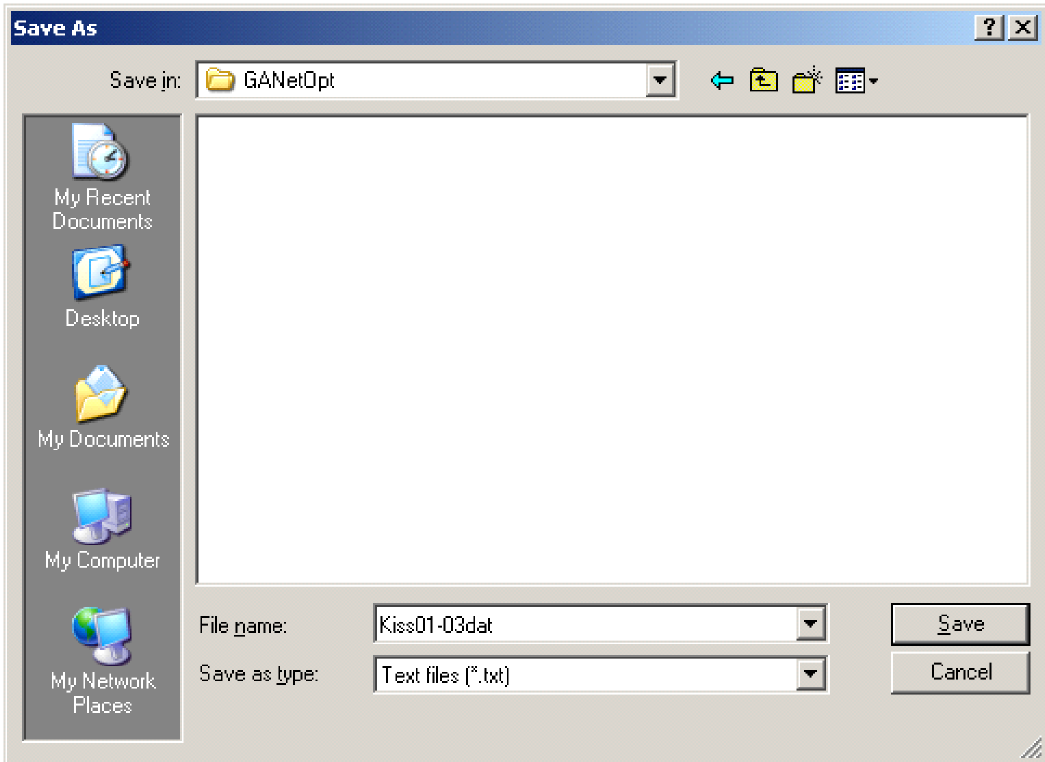
The Network Optimization in Lakes toolbar has four tools. The first two tools (OptByRange and OptByAdmRMSE) identify optimal subsets of stations and the other two tools (StatsOfDailyOptASubsets and StatsOfOptSubsetsAllPeriod) compute statistics of the optimal subsets identified.

Optimization Tools

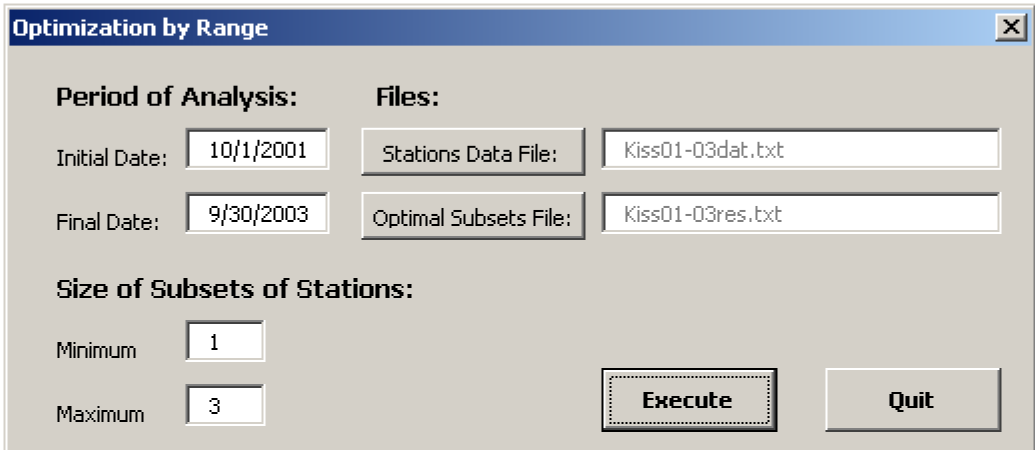
Click on the **OptByRange** button to activate the user interface for the optimization by range tool as shown below:



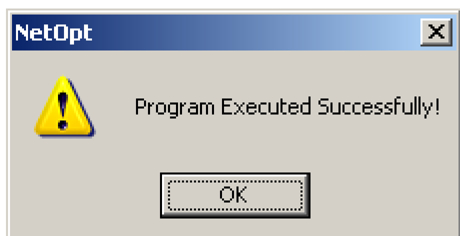
Enter the period of analysis as 10/1/2001 for Initial Date and 9/30/2003 for Final Date. The suggested maximum size of subsets is the number of selected stations minus one. Accept the values suggested. Click on **Stations Data File** button, and browse to the folder where GANetOpt.mxd and GANetOpt.mdb files are stored (c:\GANetOpt). Name the stations data file as Kiss01-03dat.



Click the **Save** button. The Kiss01-03dat.txt file is used to temporarily save the stations data for each day. Once the data are processed, the data for the entire period are stored in the Kiss01-03datAll.txt file. Now click on the **Optimal Subsets File** button, save the results file as Kiss01-03res.txt in the same folder where the data file is saved. You must take into account that the stations data and the optimization results are stored sequentially in your files, that is, they are not erased when a new optimization is run. It is, therefore, advisable that you change the names of both files if you are running a new optimization session; otherwise the results of the summary tools are difficult to interpret.

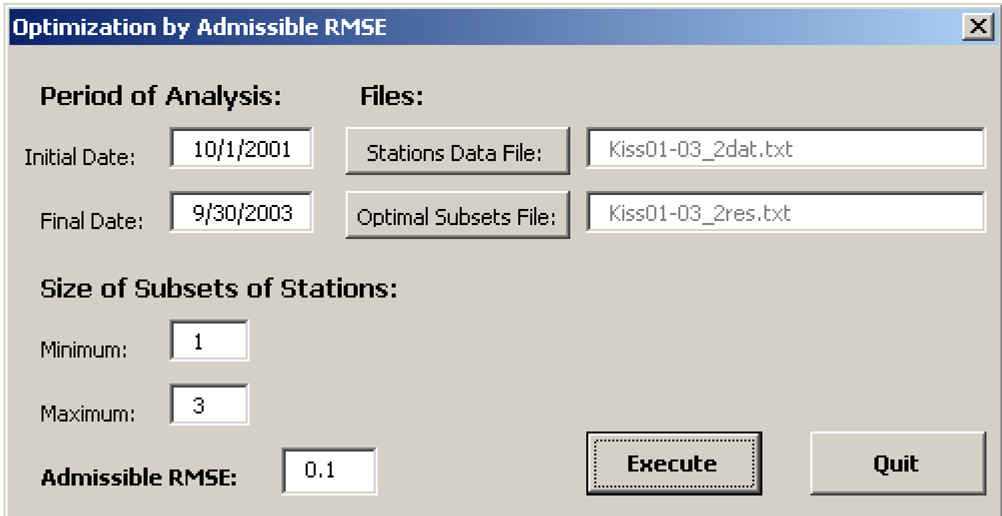


Click on the **Execute** button. In the bottom left of the screen, a progress bar will show you the progress of the optimization process. When the process ends, the following message box appears:



Click **OK** on the message box, and **Quit** on the optimization interface.

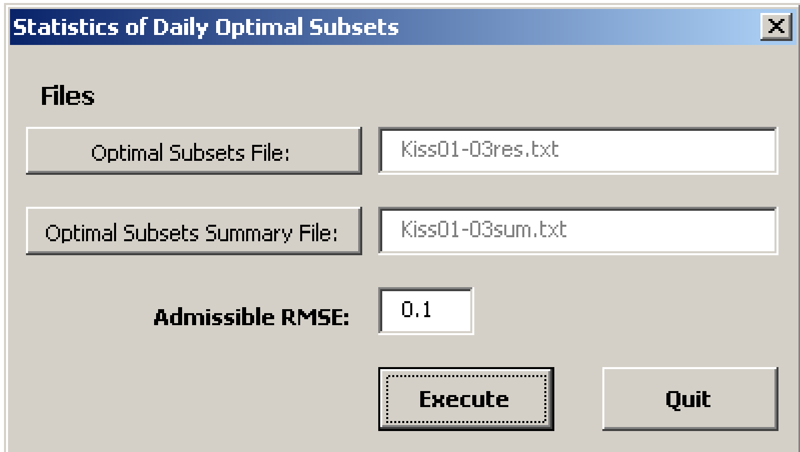
Now let us check the optimization by admissible *RMSE* tool. This tool also requires selection of stations in ArcMap. Since the stations were selected in the previous step, click on the **OptByAdmRMSE** button. Provide the inputs as shown below:



The main difference between Optimization by Admissible RMSE tool and Optimization by Range tool is that the user specifies the admissible RMSE, and the program chooses optimal subsets of stations with RMSE smaller or equal to the admissible RMSE. After the tool completes its execution, dismiss the message box by clicking **OK**. Click **Quit** on the Optimization by Admissible RMSE interface.

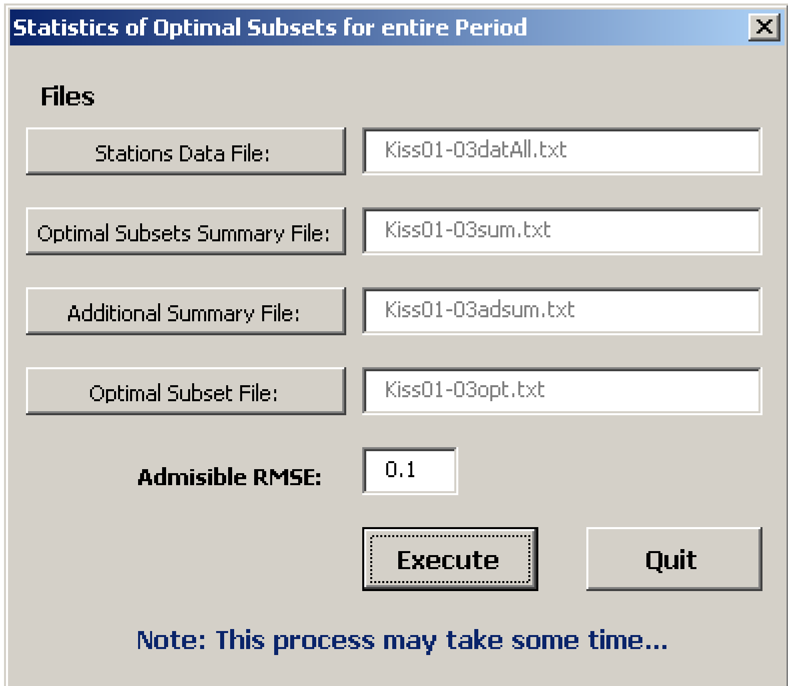
Statistics Tools

Both statistics tools use the results obtained with the optimizations tools, and, therefore, it is not necessary to choose stations or provide periods of analysis. The Statistics of Daily Optimal Subsets tool creates a text file where statistics of the identified optimal subsets are summarized. The Statistics of Daily Optimal Subsets tool can be invoked by clicking on the **StatsOfDailyOptSubsets** button of the NetOpt tool bar. Once you click on that button, you will see the Statistics of Daily Optimal Subsets interface. In this interface, provide the Optimal Subsets File, Optimal Subsets Summary File and admissible RMSE value as shown below:

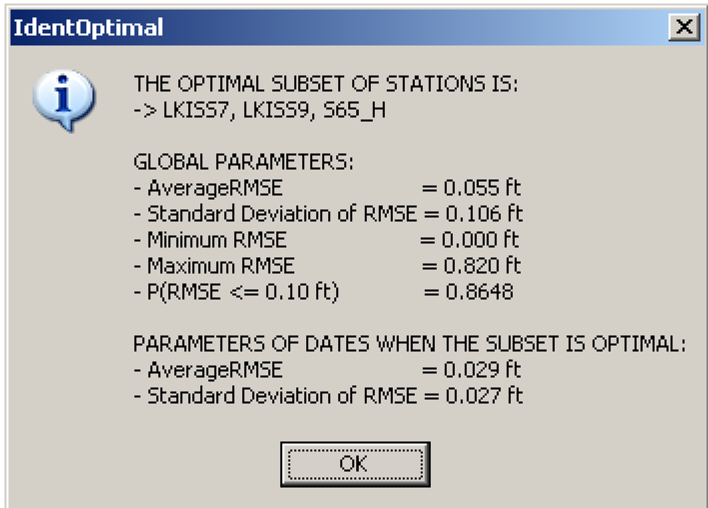


The statistical results saved in the Optimal Subsets Summary File comprise several parameters of the identified optimal subsets for the dates they were optimal. To know the statistical parameters of the same subsets for all the period of analysis and the optimal subset of the period, you must use the Statistics of Optimal Subsets for entire Period tool.

Click on the **StatsOfOptSubsetsAllPeriod** button to launch the Statistics of Optimal Subsets for entire Period tool. Immediately you will see a user interface that needs four files and one parameter. This user interface must have the following setting before you hit the **Execute** button:

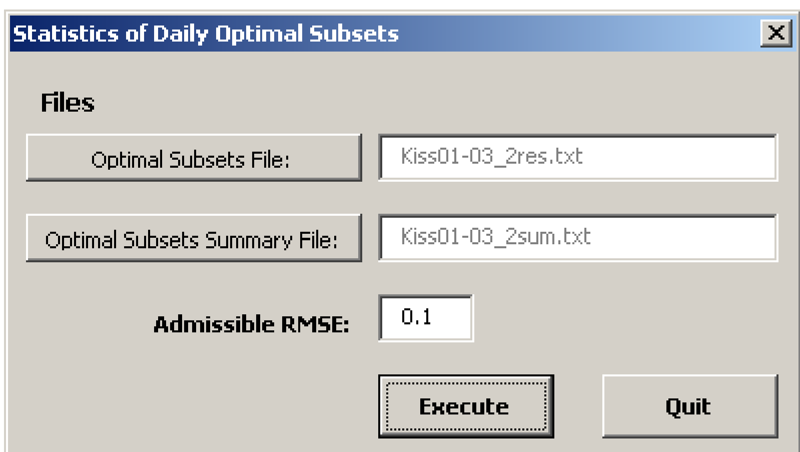


Kiss01-03datAll.txt is going to be used here because it has the stations data, date by date for the entire period. Kiss01-03sum.txt has the summary of the optimal subsets of the period. The third file is the Additional Summary File (for this example, Kiss01-03adsum.txt). In Kiss01-03opt.txt, the last file, this tool will store a summary of the optimal subset of the period along with several statistical parameters of all the identified daily optimal subsets. Click on **Execute**. After a while, the optimal subset is selected in the ArcMap document and a message box shows you the optimal subset and its statistical parameters.

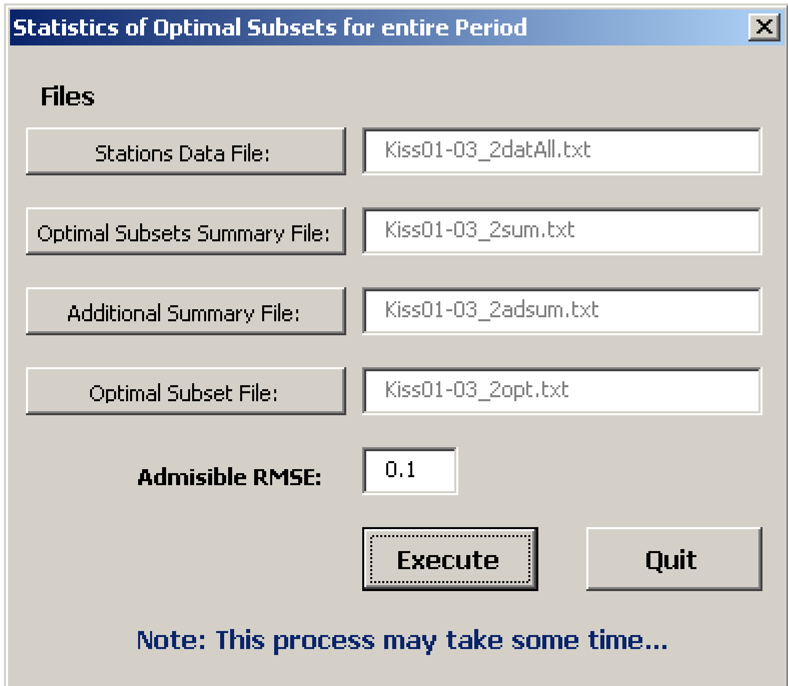


You know the optimal subset! Finally, press twice **OK**, and press **Quit** on the interface.

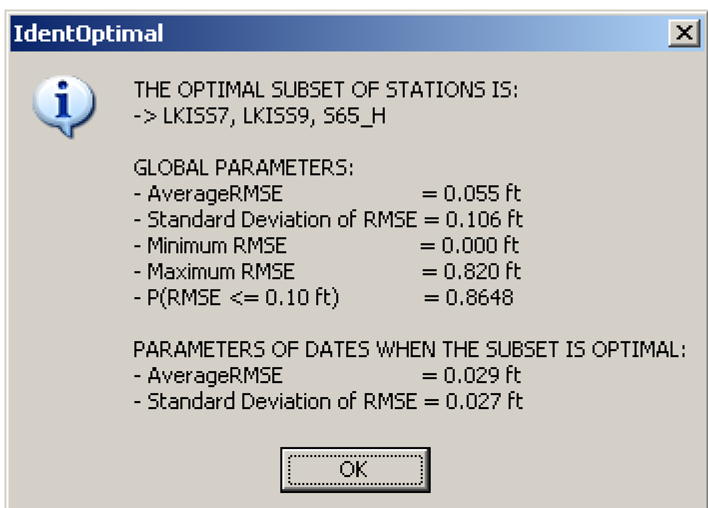
Following exactly the same steps of the statistical processing of the results of the optimization by range tool, obtain the summary files of the optimization by admissible RMSE tool results.



Then obtain the optimal subset of stations with the Statistics of Optimal Subsets for entire Period:



Press **Execute**. Moments later appears the message box:

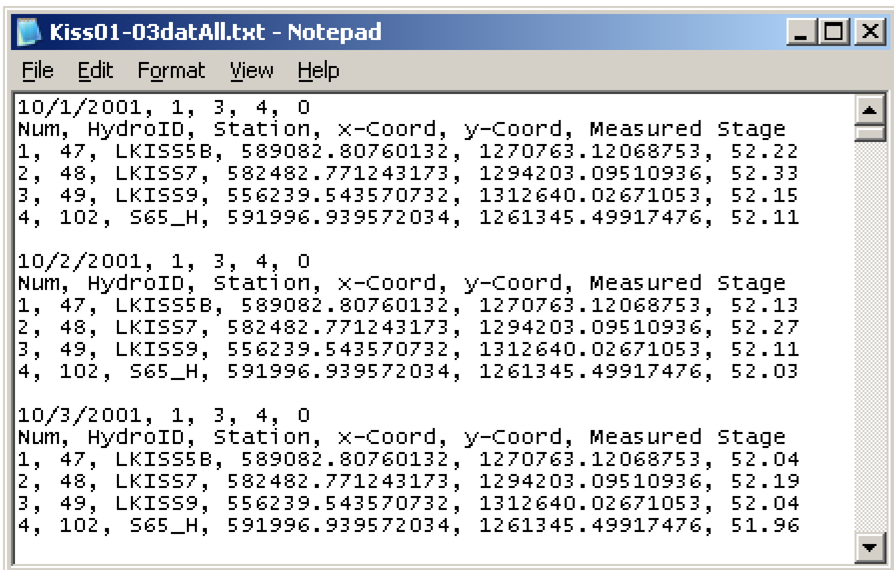


For this example, the results of both optimization options are identical! OK, you are done!

Notes about the Network Optimization in Lakes Toolset

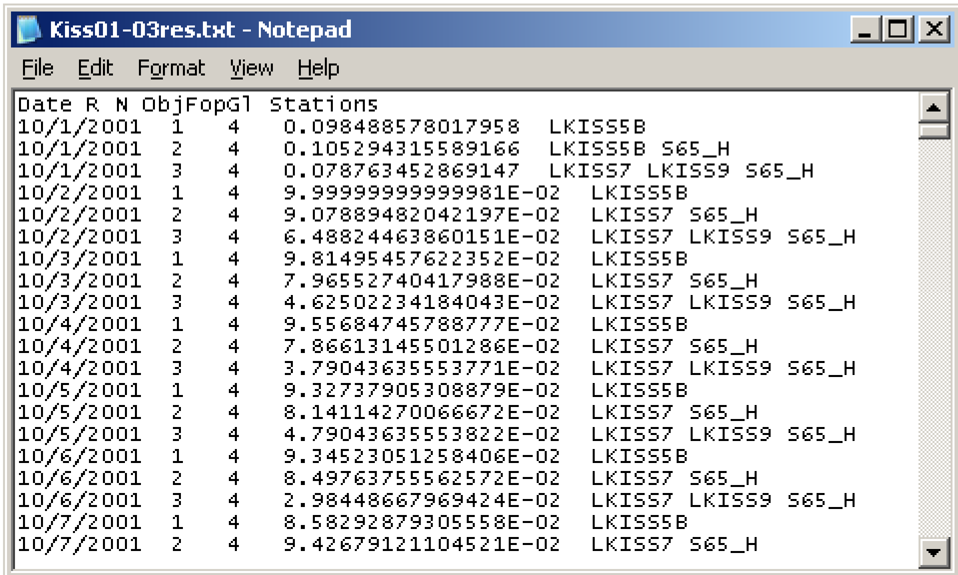
<u>Note 1. Structure of the files created by the Optimization Tools</u>

The Stations Data File (for example, from the manual, Kiss01-03datAll.txt) has for each day, the date, minimum *R*, maximum *R*, number of operating stations, and admissible *RMSE.* The second line has the headings of the data of each station (number of station, HydroID, Name of station, x-coordinate, y-coordinate, and measured stage). From the third line, a line of data is written for each of the operating stations. Next line is a blank line. The structure is repeated until the last day of the period is processed. (See next figure).




```
10/1/2001, 1, 3, 4, 0
Num, HydroID, Station, x-Coord, y-Coord, Measured Stage
1, 47, LKISS5B, 589082.80760132, 1270763.12068753, 52.22
2, 48, LKISS7, 582482.771243173, 1294203.09510936, 52.33
3, 49, LKISS9, 556239.543570732, 1312640.02671053, 52.15
4, 102, S65_H, 591996.939572034, 1261345.49917476, 52.11

10/2/2001, 1, 3, 4, 0
Num, HydroID, Station, x-Coord, y-Coord, Measured Stage
1, 47, LKISS5B, 589082.80760132, 1270763.12068753, 52.13
2, 48, LKISS7, 582482.771243173, 1294203.09510936, 52.27
3, 49, LKISS9, 556239.543570732, 1312640.02671053, 52.11
4, 102, S65_H, 591996.939572034, 1261345.49917476, 52.03

10/3/2001, 1, 3, 4, 0
Num, HydroID, Station, x-Coord, y-Coord, Measured Stage
1, 47, LKISS5B, 589082.80760132, 1270763.12068753, 52.04
2, 48, LKISS7, 582482.771243173, 1294203.09510936, 52.19
3, 49, LKISS9, 556239.543570732, 1312640.02671053, 52.04
4, 102, S65_H, 591996.939572034, 1261345.49917476, 51.96
```

The structure of the Optimal Subsets File (for example, Kiss01-03res.txt) has a heading with date, size of the identified subset of stations (R), size of the set (N), RMSE identified with ObjFopGl, and stations. In rows from row 2 to the final row, each row has the actual values of a given date and the corresponding identified subset. (See next figure).


Kiss01-03res.txt - Notepad

File Edit Format View Help

```
Date R N ObjFopGl Stations
10/1/2001  1   4   0.09848857801 7958  LKISS5B
10/1/2001  2   4   0.105294315589166  LKISS5B S65_H
10/1/2001  3   4   0.07876345286914 7  LKISS7 LKISS9 S65_H
10/2/2001  1   4   9.99999999999981E-02  LKISS5B
10/2/2001  2   4   9.07889482042197E-02  LKISS7 S65_H
10/2/2001  3   4   6.48824463860151E-02  LKISS7 LKISS9 S65_H
10/3/2001  1   4   9.81495457622352E-02  LKISS5B
10/3/2001  2   4   7.96552740417988E-02  LKISS7 S65_H
10/3/2001  3   4   4.62502234184043E-02  LKISS7 LKISS9 S65_H
10/4/2001  1   4   9.55684745788777E-02  LKISS5B
10/4/2001  2   4   7.86613145501286E-02  LKISS7 S65_H
10/4/2001  3   4   3.79043635553771E-02  LKISS7 LKISS9 S65_H
10/5/2001  1   4   9.32737905308879E-02  LKISS5B
10/5/2001  2   4   8.14114270066672E-02  LKISS7 S65_H
10/5/2001  3   4   4.79043635553822E-02  LKISS7 LKISS9 S65_H
10/6/2001  1   4   9.34523051258406E-02  LKISS5B
10/6/2001  2   4   8.49763755562572E-02  LKISS7 S65_H
10/6/2001  3   4   2.98448667969424E-02  LKISS7 LKISS9 S65_H
10/7/2001  1   4   8.58292879305558E-02  LKISS5B
10/7/2001  2   4   9.42679121104521E-02  LKISS7 S65_H
```

Note 2. Optimization by Admissible RMSE warning

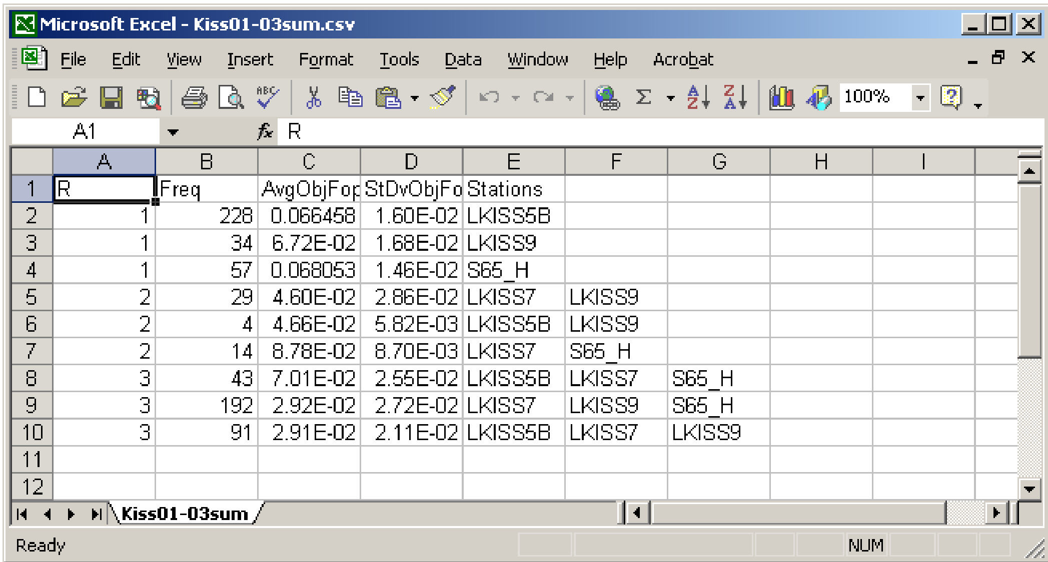
The results of the Optimization by Admissible RMSE tool sometimes miss the optimal subsets with smaller *R* due to an oscillating behavior of the *RMSE* function when *R* changes.

Note 3. Slightly different optimization results warning

It is important to note that you may get slightly different results if you run the optimization tools two or more times with the same data. You may check the files created by these two options. This is because we are applying an optimization method that uses probability to choose optimal subsets. When you apply the statistical tools to those results, those differences should not change the optimal subset of the period.

Note 4. Structure of the file created by the Statistics of Daily Optimal Subsets tool

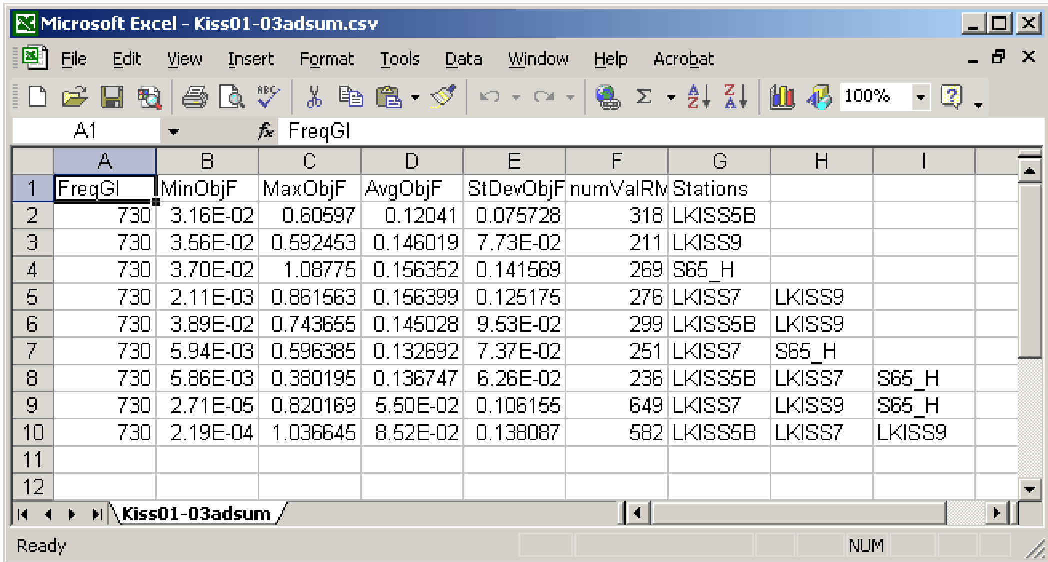
The statistical results saved in the Optimal Subsets Summary File (for example, Kiss01-03sum.txt) have as many rows as the number of identified optimal subsets. One row (or record) for each subset comprises the size of the identified optimal subset of stations, its frequency of occurrence, average of *RMSE*, standard deviation of *RMSE* and the stations that are part of the subset. The Optimal Subsets Summary file can be exported or opened in Excel for further analysis. To open this summary file directly in Excel, change its extension, txt, to csv. Double click on it and you will have the results in a spreadsheet directly in Excel.



	A	B	C	D	E	F	G	H	I
1	R	Freq	AvgObjFor	StDvObjFo	Stations				
2	1	228	0.066458	1.60E-02	LKISS5B				
3	1	34	6.72E-02	1.68E-02	LKISS9				
4	1	57	0.068053	1.46E-02	S65_H				
5	2	29	4.60E-02	2.86E-02	LKISS7	LKISS9			
6	2	4	4.66E-02	5.82E-03	LKISS5B	LKISS9			
7	2	14	8.78E-02	8.70E-03	LKISS7	S65_H			
8	3	43	7.01E-02	2.55E-02	LKISS5B	LKISS7	S65_H		
9	3	192	2.92E-02	2.72E-02	LKISS7	LKISS9	S65_H		
10	3	91	2.91E-02	2.11E-02	LKISS5B	LKISS7	LKISS9		
11									
12									

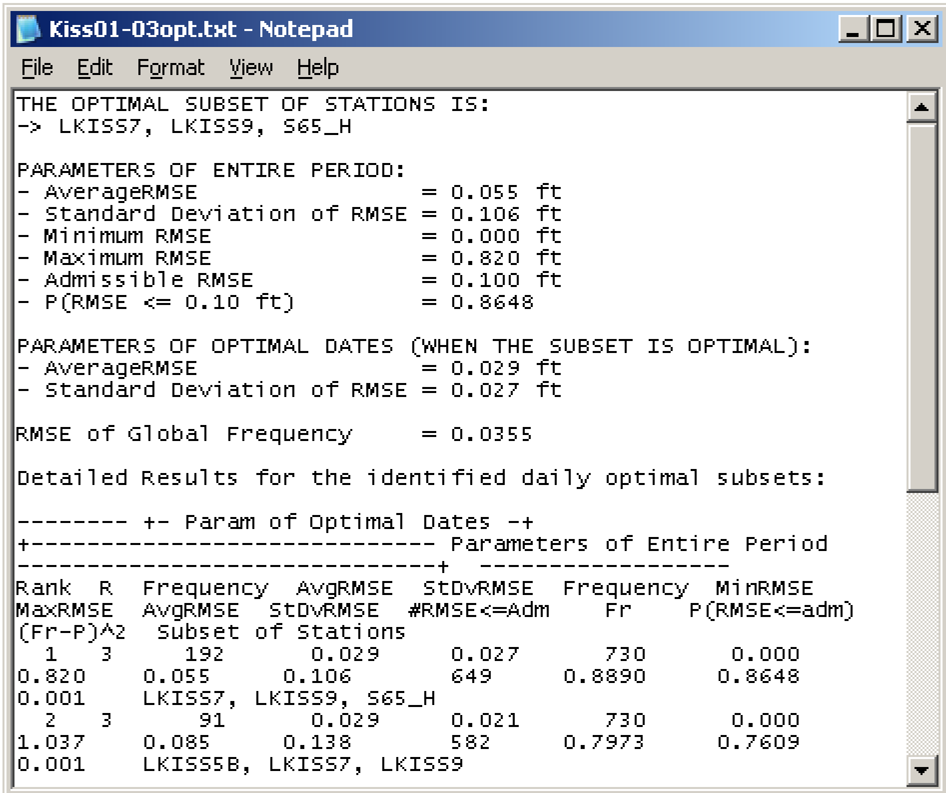
Note 5. Structure of the files created Statistics of Optimal Subsets for entire period

The Additional Summary File (for this example, Kiss01-03adsum.txt) has a row for each optimal subset, its frequency in the period, minimum *RMSE*, maximum *RMSE*, average *RMSE*, standard deviation of *RMSE*, the number of times its *RMSE* is smaller or equal to the admissible *RMSE* and, finally, the stations that are part of the subset. The Additional Summary File can also be exported or opened in Excel for further analysis. To open this additional summary file directly in Excel, change its extension, txt, to csv. Double click on it and you will have the results in a spreadsheet directly in Excel.



	A	B	C	D	E	F	G	H	I
1	FreqGl	MinObjF	MaxObjF	AvgObjF	StDevObjF	numValRM	Stations		
2	730	3.16E-02	0.60597	0.12041	0.075728	318	LKISS5B		
3	730	3.56E-02	0.592453	0.146019	7.73E-02	211	LKISS9		
4	730	3.70E-02	1.08775	0.156352	0.141569	269	S65_H		
5	730	2.11E-03	0.861563	0.156399	0.125175	276	LKISS7	LKISS9	
6	730	3.89E-02	0.743655	0.145028	9.53E-02	299	LKISS5B	LKISS9	
7	730	5.94E-03	0.596385	0.132692	7.37E-02	251	LKISS7	S65_H	
8	730	5.86E-03	0.380195	0.136747	6.26E-02	236	LKISS5B	LKISS7	S65_H
9	730	2.71E-05	0.820169	5.50E-02	0.106155	649	LKISS7	LKISS9	S65_H
10	730	2.19E-04	1.036645	8.52E-02	0.138087	582	LKISS5B	LKISS7	LKISS9

The Optimal Subset File (for example, Kiss01-03opt.txt) has stored the same results shown in the message box and a detailed summary of results of all the daily optimal subsets. The detailed results are organized with rows with three sections, with each row representing each identified optimal subset. The first section has the rank and size of the subset. The second section has the statistics of the subset in its optimal dates, frequency, average RMSE, and standard deviation of RMSE. The final section has the statistics of the subset in the entire period, frequency, minimum RMSE, Maximum RMSE, average RMSE, standard deviation of RMSE, number of days with RMSE smaller or equal than admissible RMSE, relative frequency, probability of no exceedence, square error of relative frequency, and the enumeration of the stations that are part of the subset.



Kiss01-03opt.txt - Notepad

File Edit Format View Help

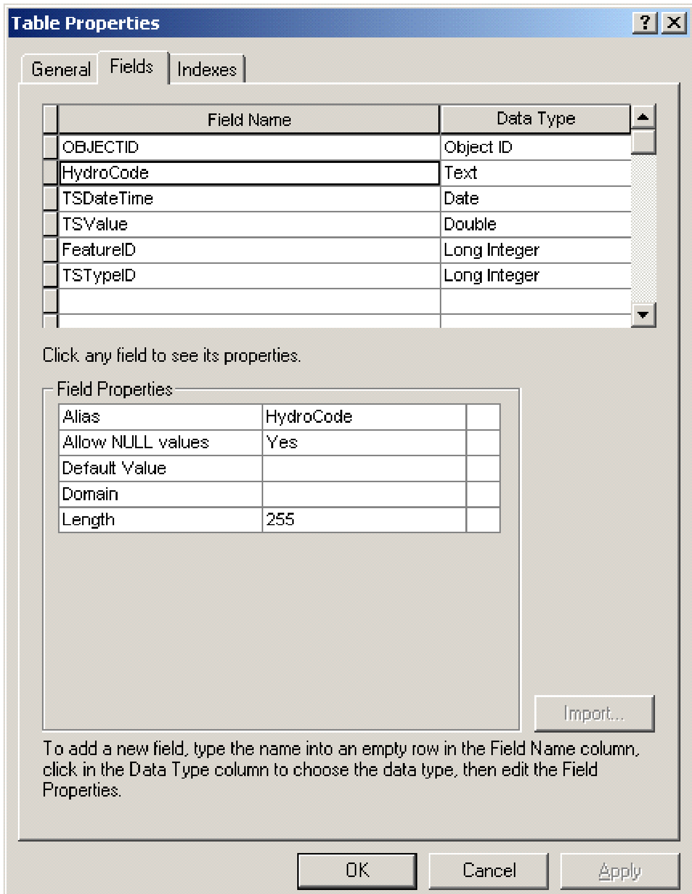
```
THE OPTIMAL SUBSET OF STATIONS IS:
-> LKISS7, LKISS9, S65_H

PARAMETERS OF ENTIRE PERIOD:
- AverageRMSE                 = 0.055 ft
- Standard Deviation of RMSE = 0.106 ft
- Minimum RMSE                = 0.000 ft
- Maximum RMSE                = 0.820 ft
- Admissible RMSE             = 0.100 ft
- P(RMSE <= 0.10 ft)          = 0.8648

PARAMETERS OF OPTIMAL DATES (WHEN THE SUBSET IS OPTIMAL):
- AverageRMSE                 = 0.029 ft
- Standard Deviation of RMSE = 0.027 ft

RMSE of Global Frequency      = 0.0355

Detailed Results for the identified daily optimal subsets:

-------- +- Param of Optimal Dates -+
+----------------------------- Parameters of Entire Period
------------------------------+  ------------------
Rank  R  Frequency  AvgRMSE  StDvRMSE  Frequency  MinRMSE
MaxRMSE  AvgRMSE  StDvRMSE  #RMSE<=Adm    Fr     P(RMSE<=adm)
(Fr-P)^2  Subset of Stations
  1   3     192      0.029     0.027      730      0.000
0.820    0.055     0.106      649     0.8890     0.8648
0.001    LKISS7, LKISS9, S65_H
  2   3      91      0.029     0.021      730      0.000
1.037    0.085     0.138      582     0.7973     0.7609
0.001    LKISS5B, LKISS7, LKISS9
```

<u>*Note 6. Loading Data into the Geodatabase GANetOpt*</u>

The loading of new data into the geodatabase GANetOpt can be done following these steps:

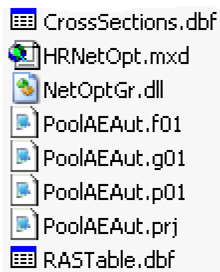
- Identify the station(s) for which you want to load data.
- Query the StageStations feature class to see if the stations are in it.
- If a station does exist, copy its HydroID for further reference.
- If a station does not exist, assign to it a new HydroID, and create a new station in the feature class StageStations using data downloaded from DBHYDRO http://www.sfwmd.gov/org/ema/dbhydro/index.html. In DBHYDRO, look for active stations in surface water data, Data Type = STG (stage), Freq = DA (daily) and Stat = MEAN. Keep the fields that are present in the StageStations feature class.
- Once the stations have been identified, retrieve their corresponding stage time series. Format the data according to the structure of the TimeSeries table. A convenient table format is an Access table. Finally, to load into

the time series table the data from an Access table you can use ArcCatalog's Simple Data Loader. The structure of the TimeSeries table is:



Network Optimization in Streams Toolset

The purpose of this section is to show how to use the Network Optimization in Streams Toolset. The optimization of the stations along the main stream of Pool AE is used as an example. The files necessary to use the Network Optimization in Streams Toolbar are in file HRNetOpt.zip located at ftp://ftp.crwr.utexas.edu/pub/outgoing/smartinez/, and are shown below:

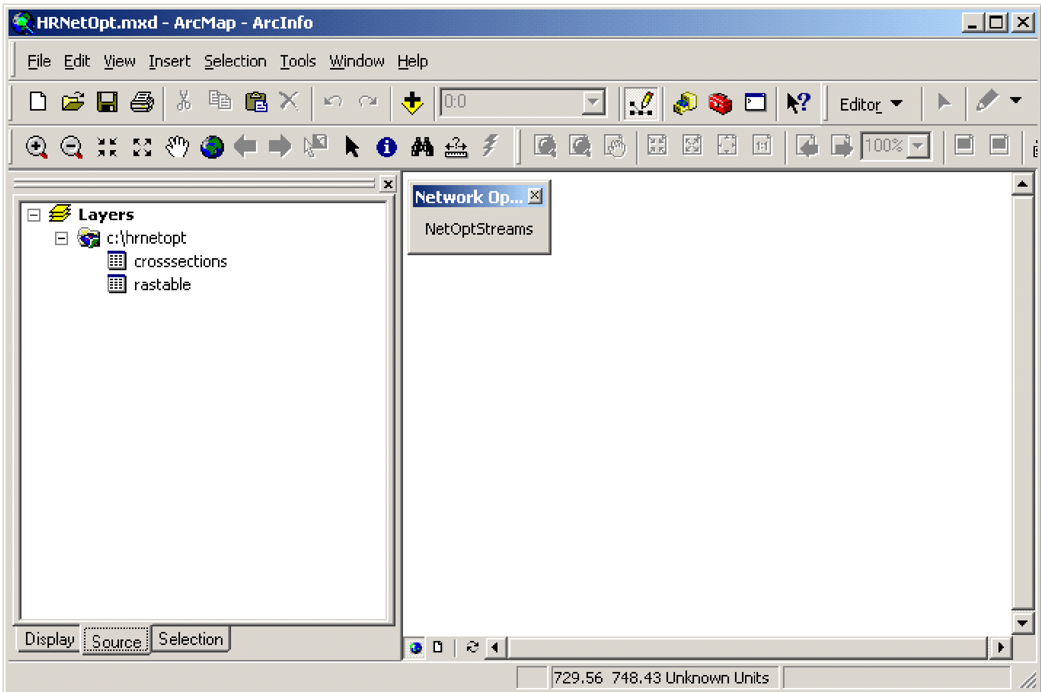


CrossSections.dbf is a table that contains the observed daily water elevations in the PoolAE stations. The period of interest is from 10/01/2001 to 12/30/2003. RASTable.dbf is a table that has the flows, downstream boundary condition, and internal water elevations that are needed to compute the water surface profile of the stream in study for each of the 428 dates selected inside the period of interest. Both tables share the same period of interest. PoolAEAut.* are several files that are used by HEC-RAS to model the stream in study. HRNetOpt.mxd is an ArcMap document with the NetOptStreams toolbar, its associated VBA code and the data from both dbf tables. NetOptGr.dll is the library that implements the NetOptStreamsGr tool. Additionally, several files with the results of the exercise are included.

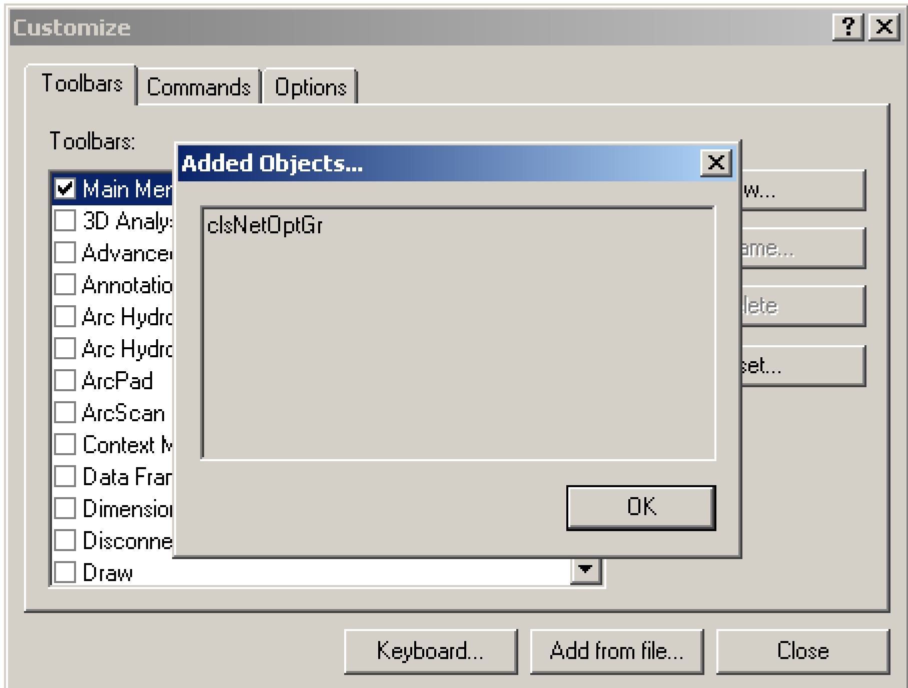
As an additional requirement, HEC-RAS must be installed in your computer before using this tool. The NetOptStreams tool has been tested with HEC-RAS versions 3.1.1 and 3.1.2.

Getting Started

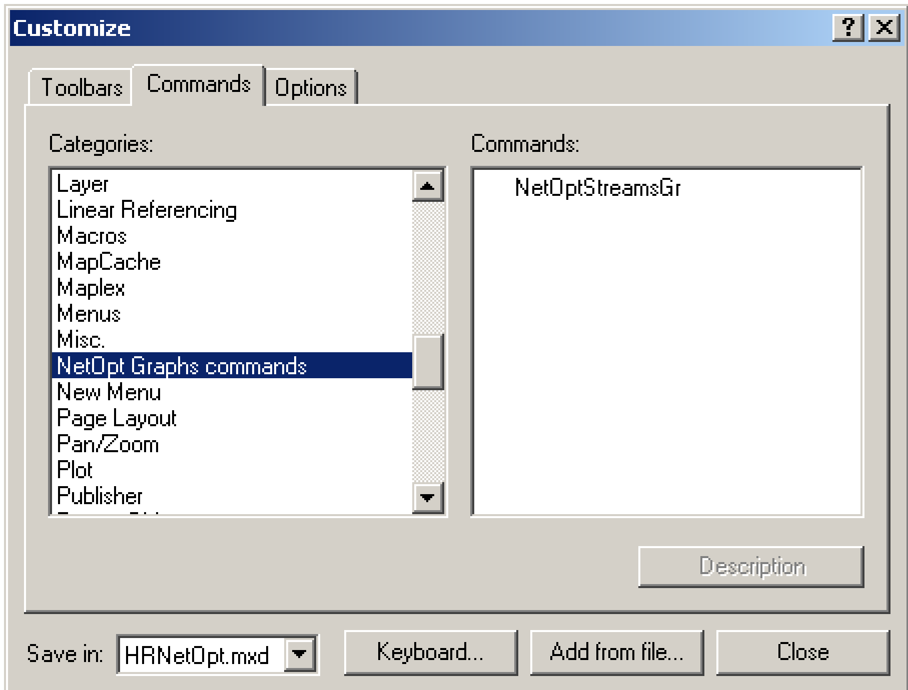
To use the NetOptStreams toolbar, unzip HRNetOpt.zip and copy all the files shown in the previous figure in one folder on your local hard drive (For example, c:\HRNetOpt). Open ArcMap, and open HRNetOpt.mxd.



You will need to add the NetOptStreamsGr tool to the toolbar. Click **Tools>Customize**. Press **Add from File…** and browse to your folder. Choose NetOptGr.dll. Click on **Open**. A window will appear telling you that the new object clsNetOptGR was added:



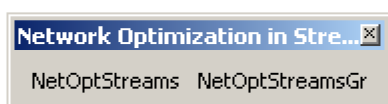
Click on **OK**. Click on the Commands tab. Browse in the categories list. Highlight the NetOpt Graphs commands category. Look to the Commands list, the NetOptStreamsGr command has appeared.



Drag this command to the right of the **NetOptStreams** button in the Network Optimization in Streams Toolbar. Once you have the new tool in the toolbar, click on **Close**. Now, the Network Optimization in Streams toolbar has its two tools. You are ready to work.

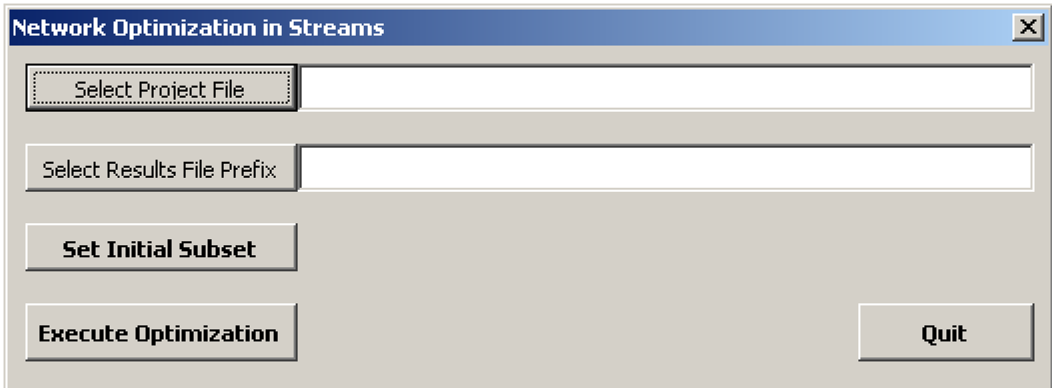
Network Optimization in Streams Toolbar

The Network Optimization in Streams toolbar has two tools: the NetOptStreams and NetOptStreamsGr tools.

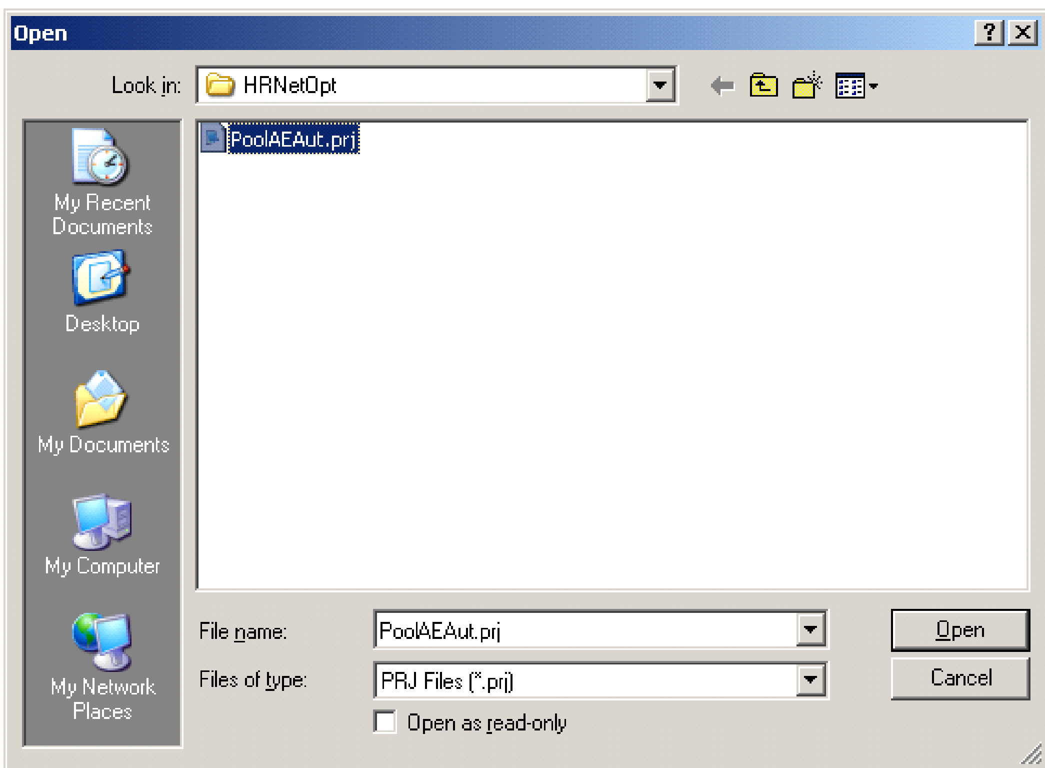


NetOptStreams Tool

The NetOptStreams tool carries out all the optimization process. Click on the **NetOptStreams** button to activate the user interface as shown below:

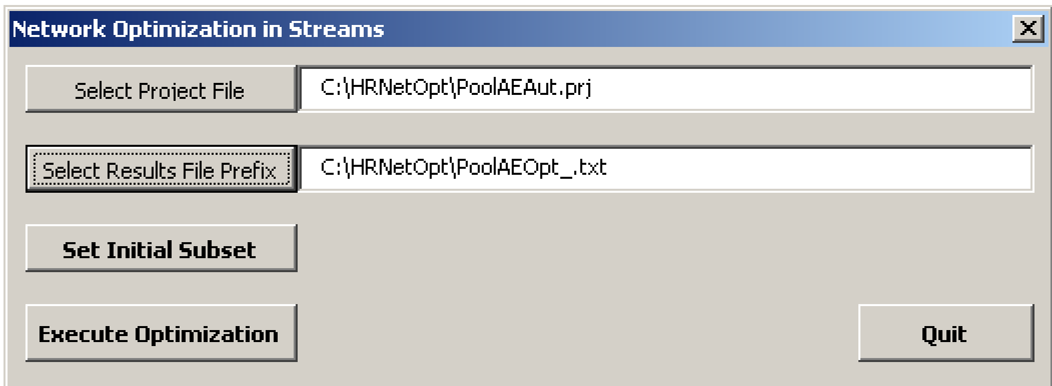


Click on the **Select Project File** and browse to the folder where HRNetOpt.mxd and all the files are stored (c:\HRNetOpt). Choose PoolAE.prj.

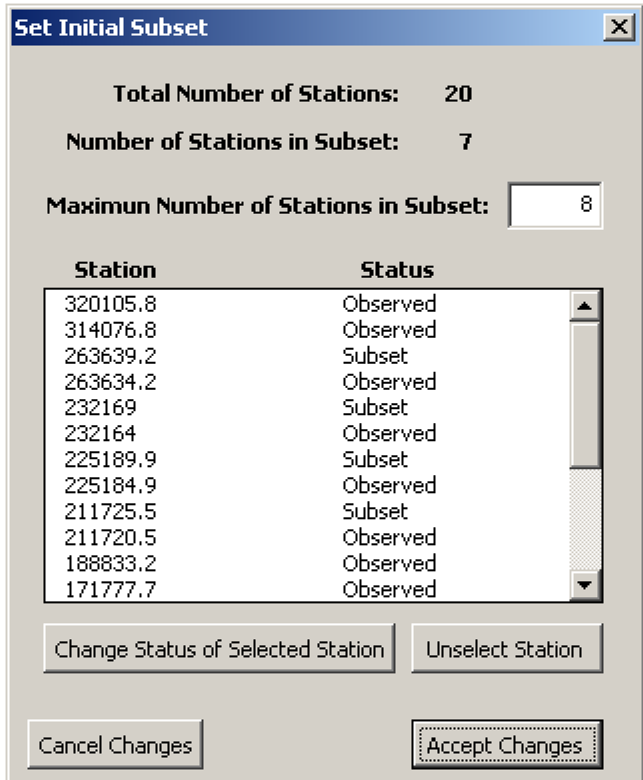


Click the **Open** button. Now click on the Select Results File Prefix button. Give any name, for example PoolAEOpt_. Click the Save button. In the NetOptStreams user interface, the name will appear as PoolAEOpt_.txt. All the files with results will have the prefix "PoolAEOpt_" and then a specific string such as "RMSE" and ".txt". In this example, the name of the file with RMSE by date will be

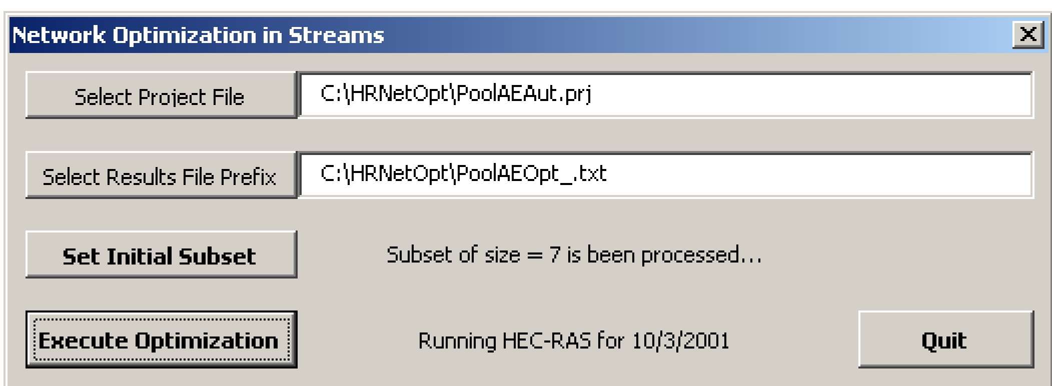
"PoolAEOpt_RMSE.txt". You must take into account that the optimization results are stored sequentially in your files, that is, they are not erased when a new optimization is run. It is, therefore, advisable that you change the prefix of your results files if you are running a new optimization session. Otherwise your results will be difficult to interpret. After introducing the results file prefix, the NetOptStreams user interface has the following data:



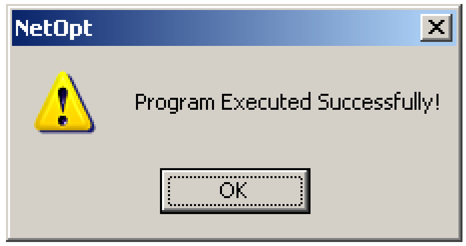
The next step is to establish the initial base subset of stations. The CrossSections table that you copied has defined a subset of 7 stations. The 47964.26 station is the downstream boundary station. Assuming a subcritical flow, this station must always be in the base subset because HEC-RAS needs it to compute the profiles. A station in the base subset is indicated with "Subset", a station that is outside the subset with "Observed". You can experiment changing the status of one or more stations. Just select the desired station in the list box and click on the Change Status of Selected Station button. If you click the button twice, you will end up with the original status in the selected station. Before continuing, you must be sure that the stations in the subset are 263639.2, 232169, 225189.9, 211725.5, 134404.5, 87366.08, and 47964.26. Once you set the initial subset, you can set the maximum size of the base subset. In this exercise, you can type 8. Before hitting Accept Changes, you should have:



The Set Initial Subset Interface disappears. You are ready to run the optimization process. Just click on the **Execute Optimization** button. You will see the progress of the process in the NetOptStreams interface.



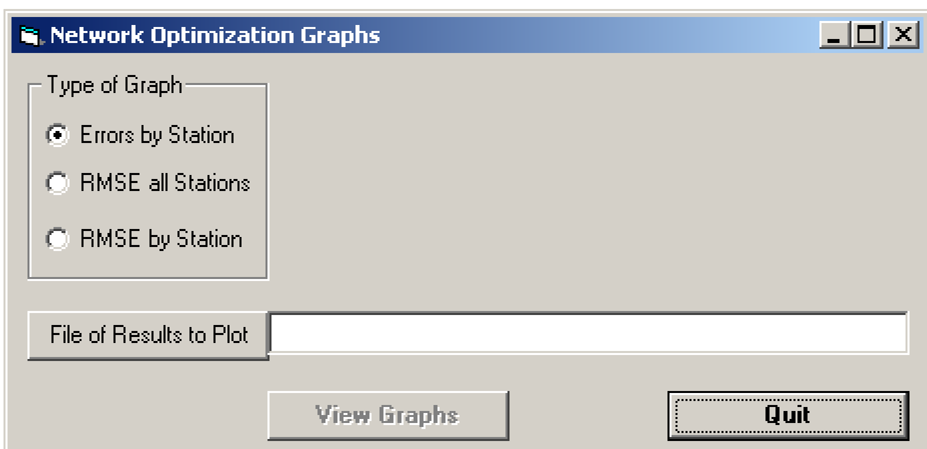
When the optimization process ends, the following message box appears:



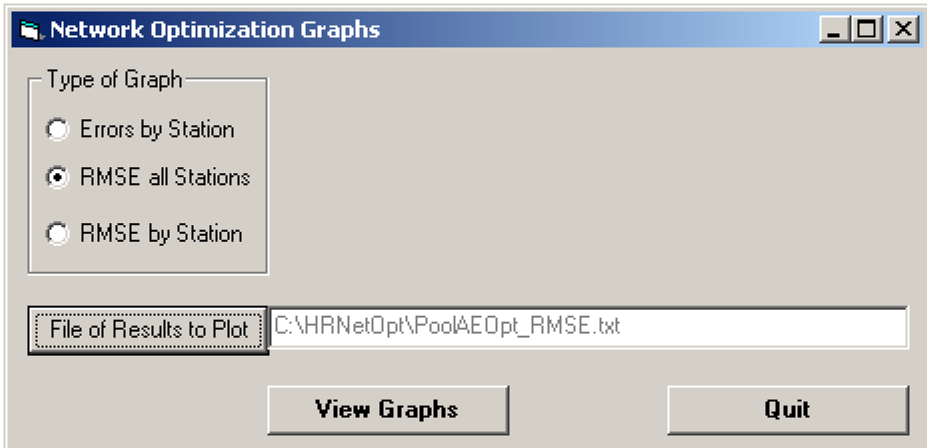
Click **OK**. Finally click on **Quit** to finish the execution of the NetOptStreams tool. For this example, the results are stored in several text files: PoolAEOpt_RMSE.txt used for RMSE daily values, PoolAEOpt_RMSE.txt for station errors and *RMSESt* values, and one file for each station outside the base subset. The names of these files are "PoolAEOpt_" + Name_of_Station + "Err.txt". Here you will find the files for 13 stations. All these files are intended for further analysis.

NetOptStreamsGr Tool

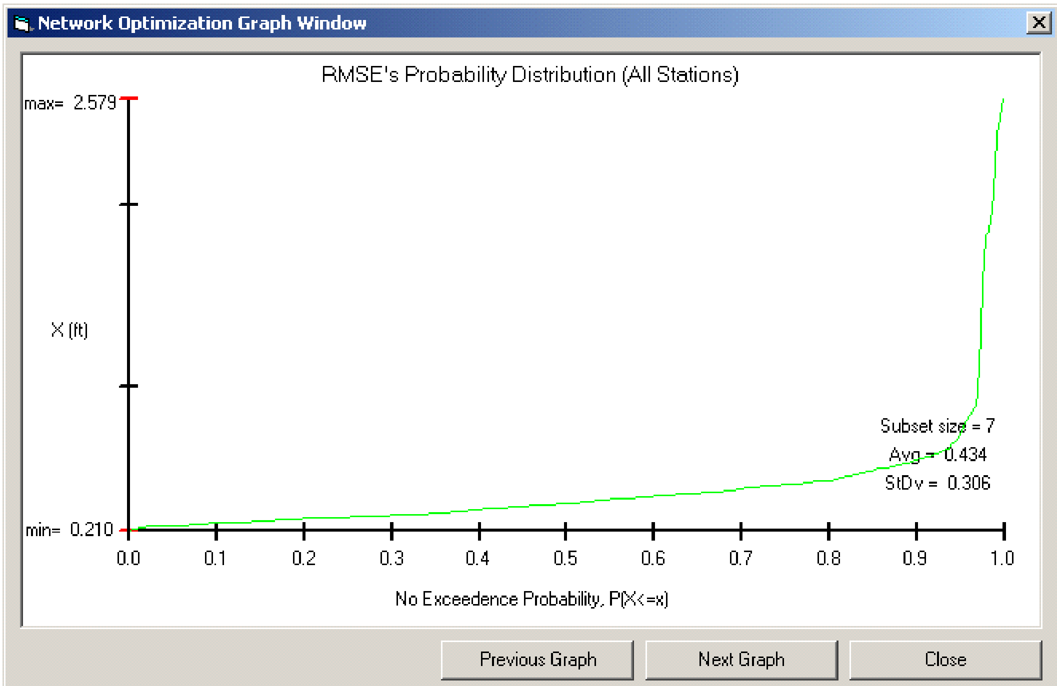
The NetOptStreams tool makes three types of graphs of the results of the optimization process. Click on the **NetOptStreamsGr** button to activate the user interface as shown below:



The first two graphs are probability graphs of the errors in a given station and the RMSE in all the stations. The third graph is a plot of the value of *RMSESt* in each station, given a size of the base subset, versus the station's distance along the stream. Here, you can choose any of the three types of graphs. For example, if you choose RMSE all Stations, you will need to click on the option **RMSE all Stations.** Click on the **File of Results to Plot** button. Navigate to your folder and choose the file PoolAEOpt_RMSE.txt. At the end, you will have the next user interface:



Now click on **View Graphs**. The graph that must be in the Graph Window is the RMSE's Probability Distribution for all stations for the base subset with seven stations. You can change to other graph by clicking the **Previous** or **Next Graph** buttons. In this example, there are only two different graphs. Close the graph window.



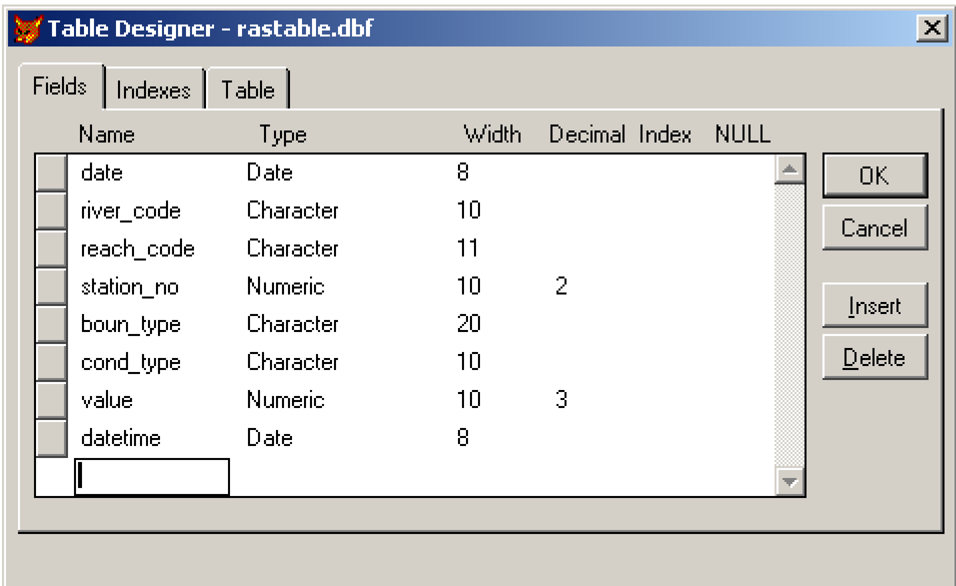
Summarizing, to plot the other two types of graphs you will follow the same steps: choose the type of graph, select the appropriate file and view the graph. Play with the other two types of graphs. To finish, click on **Quit** to close the Network Optimization Graphs user interface.

OK, you are done!

Notes about the Network Optimization in Streams Toolset

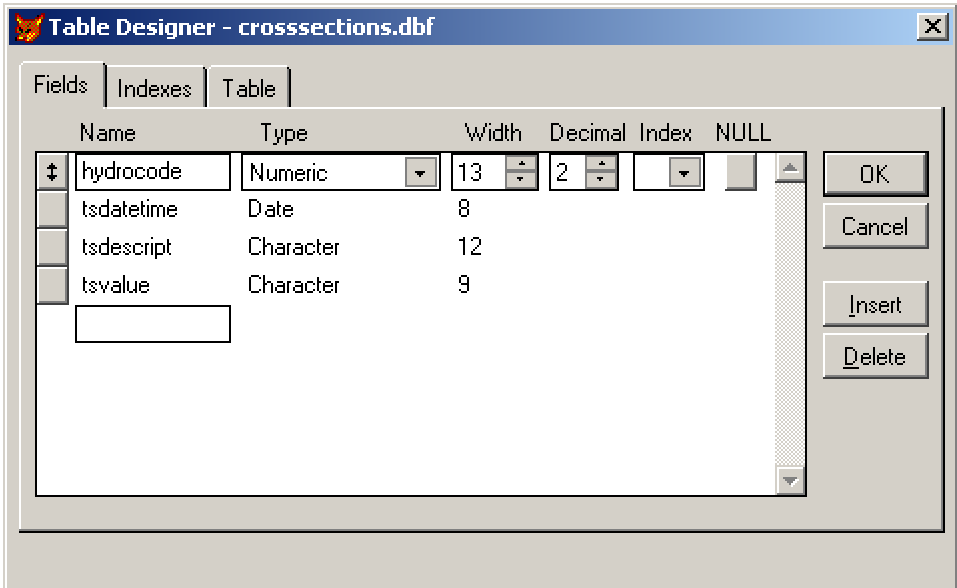
Note 1. The Structure of the Tables RASTable and CrossSections

The structure of RASTable.dbf is shown in the figure:



It has eight fields. The Date field could be used to specify the date a particular record was loaded. River_code and Reach_code are respectively the River name and Reach name assigned in HEC-RAS. Station_no is the River station assigned in HEC-RAS. Boun_type and Cond_type also assigned in HEC-RAS. Value is the numerical value of the flow or stage loaded in the record. Possible values for Boun_type are “Flow Change”, “Downstream Boundary” and “Internal Boundary”. A valid value for Cond_type is “Know WS”. Finally, the Datetime field is used to identify a particular date in the period of analysis.

The structure of CrossSections.dbf is shown in the figure:



This table has records of four fields. The hydrocode field is used to store the River station, corresponding to Station_no in RASTable.dbf. TSDateTime corresponds to DateTime in RASTable.dbf. TSDescript is a field that can be "Subset" for a station in the base subset, "Observed" for a measured water elevation or "Model" for a computed water elevation. The stations along the stream must be codified with an empty TSDateTime and TSDescript, either "Subset" or "Observed".

Note 2. Structure of the Results Files

File: ResultsFilePrefix + "RMSE.txt"

Line(s)	Contents
First	Errors computed by Date (RMSE)
Second	Empty
Third	Subset size = initial size
Four (heading)	Date, nStations, RMSE
Subsequent lines, as many as valid dates	actual date, actual number of active stations, actual RMSE
Last line for current size	avgRMSE = actual avgRMSE, DvStRMSE = actual DvStRMSE
Next line	Empty
Next line	Subset size = initial size + 1
…	

Etcetera, etcetera, Until maximum size	

File: ResultsFilePrefix + “RMSESt.txt”

Line(s)	Contents
First	Errors computed by Station
Second	Empty
Third	Subset size = initial size
Four (heading)	Station, ndays, avgErr, DvStErr, RMSESt
Subsequent lines, as many as the number of stations outside base subset	actual station, actual number of dates with data, actual avgErr, actual DvStErr, actual RMSESt
Last line for current size	Station to base subset: actual station, RMSEStmax = actual RMSEStmax
Next line	Empty
Next line	Subset size = initial size + 1
…	
Etcetera, etcetera, until maximum size	

Files: ResultsFilePrefix + Station_no + “Err.txt”

Line(s)	Contents
First	Errors computed for Station Station_no
Second	Empty
Third	Subset size = initial size
Four (heading)	Date, Error
Subsequent lines, as valid dates	actual date, actual error
Previous to Last line for current size (heading)	Station, ndays, avgERR, DvStErr, RMSESt
Last line for current size	Station_no, actual number of valid days, actual avgErr, actual DvStErr, actual RMSESt
Next line	Empty
Next line	Subset size = initial size + 1
…	
Etcetera, etcetera, until maximum size	

Bibliography

Ang, A. H.-S., Tang, W. H., 1975. *Probability Concepts in Engineering Planning and Design. Volume I. Basic Principles*. John Wiley and Sons.

Barnes, J. W., and Chambers, J. B., 1995. Solving the job shop scheduling problem with tabu search, *IIE Transactions*, 27, 257-263.

Barnes J. W., Dokov, S., Dimova, B., Solomon, A., 2003. A theory of elementary landscapes, *Applied Mathematics Letters*, 16, 337-343.

Battiti, R., and Tecchiolli, G., 1994. The Reactive Tabu Search, OGRSA Journal on Computing, 6 (2): 126-140.

Burke, R. 2003. *Getting to Know ArcObjects. Programming ArcGIS with VBA.* ESRI Press, Redlands, CA.

Carlton, W. B., and Barnes, J. W., 1996. A note on hashing functions and tabu search algorithms, *European Journal of Operational Research*, 95 (1996) 237-239.

Caruso, C., and Quarta, F., 1998. Interpolation Methods Comparison. *Computers Mathematical Applications*, Volume 35, No. 12, pp. 109-126.

Chow, V. T., 1959. *Open-Channel Hydraulics*. McGraw-Hill, Inc.

Chow, V. T., Maidment, D. R., Mays, L. W., 1988. *Applied Hydrology*. McGraw-Hill, Inc.

Cieniawski, S. E., Eheart, J. W., and Ranjithan, S., 1995. Using genetic algorithms to solve a multiobjective groundwater monitoring problem, *Water Resources Research*, 31 (2): 399-409.

Collins, F.C., and Bolstad, P. V., 1996. A comparison of spatial interpolation techniques in temperature estimation, in *Proceedings of the Third International Conference/Workshop on Integrating GIS and Environmental Modeling*, Santa Fe, New Mexico, January 21-25, 1996. Santa Barbara, California: National Center for Geographic Information Analysis (NCGIA). CD-ROM.

Cunha, M. C., and Ribeiro, L., 2004. Tabu search algorithms for water network optimization, European Journal of Operational Research, 157 (2004) 746-758.

Dessalegne, T., and Nicklow, J. W., 2003. Evolutionary Algorithms for Optimal Control of the Illinois Waterway, *Proceedings of the World Water & Environmental Resources Congress 2003*, June 23-26, 2003, Philadelphia, PA, ASCE.

Dessalegne, T., and Nicklow, J. W., 2004. Optimal Operation of Multi-Reservoir River Systems Using and Artificial Life Algorithm, *Proceedings of the World Water & Environmental Resources Congress 2004*, June 27-July 1, 2004, Salt Lake City, Utah, ASCE.

Dixon, W., Smyth, G. K., and Chiswell, B., 1999. Optimized Selection of River Sampling Sites, *Water Research*, 33 (4): 971-978.

Erxleben, J., Elder, K., and Davis, R., 2002. Comparison of spatial interpolation methods for estimating snow distribution in the Colorado Rocky Mountains. *Hydrological Processes*, Volume 16, pp. 3627-3649.

ESRI, 2004. *ArcGIS Desktop Help*. ArcGIS Desktop, Version 9.0, ESRI, Redlands, CA.

Fread, D.L. and J.M. Lewis (1988). "FLDWAV: A Generalized Flood Routing Model," Proceedings of National Conference on Hydraulic Engineering, Colorado Springs, Colorado.

Fread, D.L. (1993). "NWS FLDWAV Model: The Replacement of DAMBRK for Dam-Break Flood Prediction," Proceedings: 10th Annual Conference of the Association of State Dam Safety Officials, Inc., Kansas City, Missouri, pp. 177-184.

Fread, D.L. and J.M. Lewis (1998). "NWS FLDWAV MODEL. Theoretical Description. User Documentation," National Weather Service, NOAA, Silver Spring, Maryland.

Gen, M., and Cheng, R., 2000. *Genetic Algorithms & Engineering Optimization*, John Wiley & Sons, pp. 53-96.

Glover, F. 1989. Tabu Search. Part I, *ORSA Journal on Computing*, 1 (2): 190-206.

Glover, F., and Laguna, M. 1993. Tabu Search, in Colin R. Reeves (editor), *Modern Heuristic Techniques for Combinatorial Problems*, Blackwell Scientific Publications, Oxford, pp. 70-150.

Goldberg, D. E., 1989. Chapter 1: A Gentle Introduction to Genetic Algorithms, in *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, Mass., pp. 1-6.

Jin, M., and Fread, D. L. 1997. Dynamic Flood Routing with Explicit and Implicit Numerical Solution Schemes, *Journal of Hydraulic Engineering*, ASCE, 123 (3): 166-173.

Kirkpatrick, S., Gelatt, C. D., Vechchi, M. P., 1983. Optimization by Simulated Annealing, *Science*, 220 (4598): 671-680.

Kutija, V., 1996. *Flow Adaptive Schemes*. Dissertation, TDU/IHE Delft, The Netherlands.

Kutija, V., 2003. Paper 25: Hydraulic Modelling of Floods, in Thorndycraft, V. R., Benito, G., Barriendos, M., and Llasat, M. C. (editors), *Palaeofloods, Historical Floods and Climatic Variability: Applications in Flood Risk Assessment* (Proceedings of the PHEFRA Workshop, Barcelona, 16-19th October 2002).

Laguna, M., and Barnes, J. W., 1991. Tabu search methods for a single machine scheduling problem, *Journal of Intelligent Manufacturing*, 2, 63-74.

Lai, C., Balztzer, R. A., Schaffanek, R. W., 2002. Conservation-form equations of unsteady open-channel flow, *Journal of Hydraulic Research*, 40 (5): 567-578.

Lee, Y. M., Ellis, J. H., 1996. Comparison of Algorithms for Nonlinear Integer Optimization: Application to Monitoring Network Design, *Journal of Environmental Engineering*, ASCE, 122 (6): 524-531.

Loucks, D. P., Stedinger, J. R., and Haith, D. A. 1981. Chapter 2: Identification and Evaluation of Water Management Plans, in *Water Resource Systems Planning and Analysis*, Prentice-Hall, Englewood Cliffs, pp. 15- 93.

Maidment, D. R. (ed.), 2002. *Arc Hydro: GIS for Water Resources.* ESRI Press, Redlands, CA.

Mays, L. W., and Tung, Y.-K., 2005. Chapter 1: Systems Analysis, in *Water Resources Systems Management Tools*, edited by L. W. Mays, McGraw Hill, pp. 1-1 - 1-52.

Merwade, V. M., and Maidment, D. R., 2004. Geospatial Description of River Channels in Three Dimensions. The University of Texas at Austin. Center for Research in Water Resources online report 04-8. <http://www.crwr.utexas.edu/reports/2004/rpt04-8.shtml>

Minder, E., and Nicklow, J. W., 2004. System-Wide Optimization of Dam Operations to Contro Water Level Fluctuations, *Proceedings of the World Water & Environmental Resources Congress 2001*, May 20-24, 2001, Orlando, FL, ASCE.

National Oceanic and Atmospheric Administration (NOAA), 2005. NOAA's National Weather Service. *HL River Mechanics*. <http://www.nws.noaa.gov/oh/hrl/rvrmech/>. Visited on Nov 23, 2005

National Technical Information Service (NTIS), 2005 NTIS, Web page. <http://www.ntis.gov/>. Visited on Dec 12, 2005.

Nelder, J. A., and Mead, R., 1995. A Simplex Method for function minimization, *Computer Journal*, 7 (4), 308-313.

Nunes, L. M., Cunha, M. C., and Ribeiro, L. 2004. Groundwater Monitoring Network Optimization with Redundancy Reduction. *Journal of Water Resources Planning and Management*, Vol. 130, No. 1, January, pp. .

Nunes, L. M., Paralta, E., Cunha, M. C., and Ribeiro, L., 2004. Groundwater Nitrate Monitoring Network Optimization with Missing Data, *Water Resources Research*, 40 (2): 24.

Pardo-Igúzquiza, E., 1998. Optimal Selection of number and location of rainfall gauges for aerial rainfall estimation using geostatistics and simulated annealing, *Journal of Hydrology*, Volume 210, pp. 206-220.

Powell, M. J. D., 1964. An efficient method for finding the minimum of a function of several variables without calculating derivatives, *Computer Journal*, 7, 155-162.

Reed, P. M., Minsker, B. S. and Valocchi, A. J., 2000. Cost-effective long-term groundwater monitoring design using a genetic algorithm and global mass interpolation, *JWater Resources Research*, 36 (12): 3731-3741.

Reed, P. M., Minsker, B. S., 2004. Striking the Balance: Long-Term Groundwater Monitoring Design for Conflicting Objectives, *Journal of Water Resources Planning and Management*, ASCE, 123 (2): 140-149.

South Florida Water Management District (SFWMD), 2004. <http://www.sfwmd.gov/>

Sárközy, F., 1999. GIS Functions – Interpolation, *Periodica Polytechnica Ser. Civ. Eng.*, Vol. 43, No. 1, pp. 63–86.

Singh, S. K., 2004. *Analysis of Uncertainties in Digital elevation Models in Flood (Hydraulic) Modelling*, Master Thesis, ITC, The Netherlands / IIRS, India.

Sobey, R. J., 2001. Evaluation of Numerical Models of Flood and Tide Propagation in Channels, Journal of Hydraulic Engineering, ASCE, 127 (10): 805-824.

Sorenson, J. K., and Maidment, D. R., 2004. Temporal Geoprocessing for Hydroperiod Analysis of the Kissimmee River. The University of Texas at Austin. Center for Research in Water Resources online report 04-5. <http://www.crwr.utexas.edu/reports/2004/rpt04-5.shtml>

U. S. Army Corps of Engineers (USACE), 2004. HEC-RAS River Analysis System. User's Manual. Version 3.1.2. April 2004, Hydrologic Engineering Center, Davis, California.

Zheng, C. and Wang, P., 1996. Parameter structure identification using tabu search and simulated annealing, *Advances in Water Resources*, Vol. 19, No. 4, pp. 215-224.

Zimmerman, D., Pavlik, C., Ruggles, A., and Armstrong, M. P., 1999. An Experimental Comparison of Ordinary and Universal Kriging and Inverse Distance Weighting, Mathematical Geology, Vol 31, No. 4, pp. 375-390.