APPLICATION OF GENETIC ALGORITHMS TO CALIBRATION AND VERIFICATION OF QUAL2E MODEL

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

RECEP KAYA GÖKTAŞ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF
MASTER OF SCIENCE
IN
ENVIRONMENTAL ENGINEERING

NOVEMBER 2004

Approval of the Graduate School of Natural and Applied Sciences

_____

Prof. Dr. Canan Özgen
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

_____

Prof. Dr. Filiz B. Dilek
Head of Department

This is to certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

_____

Assist. Prof. Dr. Ayşegül Aksoy
Supervisor

Examining Committee Members

Prof Dr. Celal F. Gökçay	(METU, ENVE)	_____
Assist. Prof Dr. Ayşegül Aksoy	(METU, ENVE)	_____
Prof Dr. Filiz B. Dilek	(METU, ENVE)	_____
Assist. Prof Dr. İpek İmamoğlu	(METU, ENVE)	_____
Aylin K. Onur, M.S.	(DSİ)	_____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Recep Kaya GÖKTAŞ

Signature :

ABSTRACT

APPLICATION OF GENETIC ALGORITHMS TO CALIBRATION AND VERIFICATION OF QUAL2E MODEL

Göktaş, Recep Kaya
M. Sc., Department of Environmental Engineering
Supervisor: Assist. Prof. Dr. Ayşegül Aksoy

November 2004, 153 pages

The objective of this study is to develop a calibration and verification tool for the QUAL2E Model by using Genetic Algorithms. In the developed optimization model, an objective function that is formulated on the basis of the sum-of-least squares approach aiming at minimizing the difference between the observed and simulated quantities was used. In order to perform simultaneous calibration and verification, verification of the calibrated results was treated as a constraint and inserted into the objective function as a penalty function.

The performance of the optimization model was tested for different observation data qualities represented by the synthetic perfect and biased data sets. Although it was not possible to obtain the exact values of the kinetic coefficients for any of the tests performed, the coefficient estimates were successful in reflecting the water quality variable profiles in the river.

The results of the tests showed that the performance of the optimization model is generally sensitive to the error in the observed data sets, to the number and location of sampling points, and to the objective function formulation. For the problems that

involve multiple water quality variables, a weighting approach used in the objective function formulation resulted in better performances.

The optimization model was also applied for a case study. For the same input data, calibration obtained with the genetic algorithm optimization – simulation was better compared to the trial-and-error approach.

Keywords: Calibration, Verification, QUAL2E, Genetic Algorithms, Optimization.

ÖZ

GENETİK ALGORİTMALARIN QUAL2E MODELİNİN KALİBRASYON VE VERİFİKASYONUNA UYGULANMASI

Göktaş, Recep Kaya
Yüksek Lisans, Çevre Mühendisliği Bölümü
Tez Yöneticisi: Yrd. Doç. Dr. Ayşegül Aksoy

Kasım 2004, 153 sayfa

Bu çalışmanın amacı QUAL2E modelinin kalibrasyon ve verifikasyonu için Genetik Algoritmaları kullanan bir optimizasyon aracı geliştirmektir. Geliştirilen optimizasyon modeli, en küçük kareler toplamı yaklaşımıyla formüle edilen ve gözlem verileri ile simülasyon sonuçları arasındaki farkı azaltmayı amaçlayan bir hedef fonksiyon içermektedir. Kalibrasyon ve verifikasyon işlemlerini eşzamanlı olarak yapabilmek için, verifikasyondan gelen hata bir sistem kısıtı olarak değerlendirilimiş ve hedef fonksiyon içerisine bir ceza fonksiyonu olarak yerleştirilmiştir.

Optimizasyon modelinin performansı öncelikle yapay data kullanılarak test edilmiştir. Testlerde, hatasız ve hatalı olmak üzere iki ayrı veri seti kalitesi kulanılmıştır. Bütün testler için tüm nehir kesitlerinde gerçek kinetik parametre katsayı değerlerine tam olarak ulaşılamamasına rağmen, nehirdeki su kalitesini gerçeğine çok yakın bir biçimde simüle edebilecek katsayı değerleri bulunabilmiştir.

Testlerin sonuçları, optimizasyon modelinin performansının, gözlem verilerindeki hataya, gözlem noktalarının sayısına ve yerine, ve hedef fonksiyon formülasyonuna bağlı olarak değişebildiğini göstermektedir. Birden fazla su kalitesi değişkeni içeren

problemler için hedef fonksiyonunda ağırlıklandırma yöntemi kullanılmış ve daha iyi performans elde edilebileceği görülmüştür.

Optimizasyon modeli ayrıca gerçek bir probleme de uygulanmıştır. Aynı veri setlerinin kullanılması durumunda, genetik algoritma optimizasyon – simülasyon modelinin, deneme-yanılma yaklaşımı ile elde edilenlere göre daha iyi kalibrasyon sonuçları verdiği görülmüştür.

Anahtar Kelimeler: Kalibrasyon, Verifikasyon, QUAL2E, Genetik Algoritmalar, Optimizasyon.

To my deceased grandparents

ACKNOWLEDGMENTS

The author would like to express his deepest gratitude to his supervisor Assistant Professor Dr. Ayşegül Aksoy for her guidance, advice, encouragement, and understanding all throughout the study.

The author wishes to send his sincere appreciation to Miss Aylin Kübra Onur for providing the necessary data for the case study, devoting her time to answering critical questions and her support.

The author also thanks to the examining committee members Prof. Dr. Celal F. Gökçay, Prof. Dr. Filiz B. Dilek and Assistant Professor Dr. İpek İmamoğlu for their criticisms and advices.

The author gratefully acknowledges the financial support from the METU Scientific Research Projects Fund (Project Code: BAP-2003-03-11-01).

The author sends his special thanks to Mr. Hakan Moral for his technical assistance and spiritual support throughout the study. The author is regardful to Miss Gamze Güngör for providing her personal computer for the test runs in a very critical moment of the research.

The author appreciates the continuous morale support from his friends Mr. Erkan Şahinkaya, Miss Nimet Uzal, Mr. Burak Uzal, Miss Bilgen Yüncü, Mr. Bahadır Duygulu, Mr. Baran Görmez and Mr. Mehmet Gölge.

The author wishes to mention his longing for the days with his office-mate Miss Esra Atlı.

The author extends his special thanks to his home-mates Mr. Ferhan Fıçıcı and Mr. Murat Soysal for their understanding and continuous morale support.

The author feels himself responsible to mention his deep gratitude to his parents and his sister for their devotion and endless support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Water-quality simulation models are widely used in water-quality control and management studies. These studies include analyzing the effects of a waste discharge on a water body and determining the allowable amount of waste to be discharged by a polluter, such as the waste load allocations and environmental impact assessments. Beyond being a decision-making tool in water quality management, models also enhance our understanding of the complex nature of these environmental systems.

Early water quality models aimed at predicting dissolved oxygen levels in rivers (Chapra, 1997). Scientific work in the field resulted in the incorporation of both conventional and toxic pollutants and also the ecological phenomena into the simulation capabilities of the models. Improvements in the computer technology had also contributed to the continuous development of these complex models.

As models get more complex, the number of parameters they process increases and the input data become very hard to set. The quantitative description of the processes is possible when the data set that defines the unique characteristics of the system is complete. The data may contain many parameters most of which may be very hard to determine by observations and/or measurements. Quantification of these data can be achieved using the model itself assuring the model is adequate to describe the system being studied.

Calibration and verification are the crucial steps in any water quality modeling study. Calibration can be defined as the search process to tune or quantify the kinetic parameters that will be used to simulate the system being modeled. An

optimal agreement is sought between the simulated and observed (measured) state variables of the system. Verification or validation step is where the calibrated parameter values are tested by running the model for new conditions. The kinetic parameter values are kept fixed, but the quantity of physical parameters (e.g. water flow rate, water depth, temperature, etc.) and forcing functions (e.g. headwater inputs, point sources, withdrawals, etc.) are varied. The model is verified when output results and the observed values are close enough to satisfy the verification criteria.

The conventional methodology for calibration has been the trial-and-error method. The trial-and-error method makes the process considerably time-consuming and the subjectivity of the methodology leads to questionable results. The accepted parameter estimates is very dependent on the modeler and different modelers may come up with different parameter estimates with the same model formulation and the data set. However, the advances in computer technology and operations research field can be incorporated in model calibration and verification. By this way, objective and automated methods giving much more reliable results with less time consumption can be possible. The results would be more reliable since objectivity avoids errors from personal judgment of the modeler.

The objective of this study is to develop an effective calibration and verification tool for 'The Enhanced Stream Water Quality Model' QUAL2E by using Genetic Algorithms (GAs). QUAL2E is a widely used stream water quality model. It is developed and distributed by United States Environmental Protection Agency (US EPA) in a public domain. It can simulate up to fifteen water quality constituents including dissolved oxygen, biochemical oxygen demand, temperature, nitrogen species, phosphorus species, coliforms and several user-specified conservative and nonconservative constituents. In addition to modeling the impacts of point loads on the stream water quality, QUAL2E can be used to define non-point sources affecting the stream as well. Moreover, QUAL2E can simulate algal growth, and the impacts of excess nutrients on water quality.

Genetic algorithms are probabilistic, global search methods that are inspired from natural evolution. Their robustness has been proven through many theoretical and empirical studies (Goldberg, 1989). Today, genetic algorithms are widely used in many different fields to perform optimization. In the field of environmental engineering, they have been applied successfully to optimal groundwater remediation design and management (McKinney and Lin, 1994; Cieniawski et al., 1995; Aksoy and Culver, 2000; Chan Hilton and Culver, 2000). Their applicability for the calibration of surface water quality models is also shown (Mulligan and Brown, 1998).

The first phase of the study was linking the QUAL2E model with a genetic algorithm program. This was accomplished by running the QUAL2E model as a sub-program of the genetic algorithm program. An optimization algorithm that carries out simultaneous calibration and verification of the QUAL2E model was developed. The optimization algorithm uses an objective function that was formulated according to the sum-of-least-squares criterion, in general, to minimize the error between the observed and simulated conditions. The error from verification conditions was inserted into the objective function as a penalty function. The aim was to develop a program package that is applicable to any river system modeled by QUAL2E, allowing automated calibration and verification of any combination of kinetic parameters desired by the user.

The developed optimization model was applied on a hypothetical river network by using perfect and biased measurements of state variables. The effects of number of observation points and different objective function formulations on the performance of GAs were also tested. The complexity of the calibration problem was gradually increased by adding new kinetic parameters to be calibrated and new water quality responses to be utilized. The impact of using more than one verification data set was also investigated. Finally, the optimization model was applied for a case study. For this purpose, a previous modeling study conducted on the Lower Seyhan River by Onur (1996) using QUAL2E was utilized. In this work, a trial-and-error

approach had been employed for calibration and verification. The performance of GA optimization was assessed by comparing GA calibration results with the trial-and-error calibration results.

CHAPTER 2

BACKGROUND

2.1 Calibration and Verification in Surface Water Quality Modeling

It is essential to consider the possible reactions of the water body before taking any action on it. As surface waters are complex environmental systems, it is hard to understand and predict their behavior. The attempts to predict the response of water bodies to pollution loads originating from human activities resulted in the development of mathematical models. These models quantitatively describe the physical, chemical and biological behavior of the water bodies through a collection of mathematical relationships that contain many parameters (e.g., reaction rate coefficients, biological and chemical constants) that are specific to the system modeled. However, in order to apply them, they must be able to define the system in the most accurate way.

A typical water quality modeling process consists of data collection, model formulation, calibration and verification (Little and Williams, 1992). The data collection and model formulation phases affect the results of the subsequent calibration phase. However, for a successful modeling practice, the calibration phase should give the optimum model parameters possible. During calibration or verification steps some problems in the data set or model formulation may be revealed. According to the information gathered about the system during these steps, additional sampling or model reformulation may be considered (Little and Williams, 1992).

2.2 Calibration Methodology

When environmental systems are of concern, it is generally not expected to find a data bank containing the required parameters for the system being modeled. Environmental data collection is a considerably time and money consuming process. Beyond the economic constraints, it may be practically impossible to measure some system-specific parameters due to technical reasons (e.g. access limitations for some parts of the basin, need for excessive number of measurements, etc.).

The commonly applied methodology is initially collecting the physical data (e.g., flow rate, depth, temperature) about the system and determining the other informative parameters like dissolved oxygen (DO), nitrogen and phosphorus concentrations. Then, the initial estimates for the biological, chemical, and kinetic coefficients are made. The initial estimates are determined using the literature values or, if present, field and laboratory measurements. Literature is a very valuable source in determining the model parameters and should be referred to especially in the initial phases of the modeling applications. In literature, the ranges of values for specific model parameters can be found most of the time. These values can be used as a starting point in the calibration phase. Following calibration, the parameter values most accurately defining the system are determined. However, it should be noted that the value of a parameter can be dependent on the model formulation and assumptions used in the development of the model. In such cases, even the actual readily available parameter values determined by field or laboratory studies may need tuning to obtain the optimum agreement between the observed and simulated state variables.

Calibration methods can be divided into two major categories:

- subjective methods (trial-and-error methods)
- objective methods (optimization methods)

2.2.1 Trial-and-Error Method

In the trial-and-error method, the model parameters are adjusted by the modeler him/herself by running the model for different parameter values many times until a match between the observations and the predictions is obtained. This is a time-intensive work. The method is subjective since the goodness-of-fit is judged without statistically based criteria. Little and Williams (1992) state that this method makes the model calibration ambiguous. The modeler is left unsure whether the calibration result is the best that can be achieved given the available data and the specific structural form of the model.

Although trial-and-error methods are constrained with the number of trials to find the best set of model parameter values, they are still employed for the water quality modeling studies (Lung and Larson, 1995; Drolc and Zagorc Koncan, 1996; Chaudhury et al., 1998; Onur et al., 1999). Today, State Hydraulic Works (DSİ, Devlet Su İşleri) of Turkey is using this approach for calibration of the water quality models as well.

A large, structured, trial-and-error method was tested by Jaffe et al. (1988). The technique took into account the random fluctuations of field measurements and yielded a probability distribution of the model parameters. A large number of model simulations were conducted using a wide range of model input parameters each of which were selected from the specified probability distributions. Then, from the subset of outputs that were satisfactorily close to the field observations, the mean and variance of input parameters were computed. The procedure was a useful calibration tool; however, as the complexity of the model increased, that is when the number of unknown model parameters increased, the number of simulations required would also increase.

2.2.2. Calibration Using Optimization Methods

It is possible to increase the efficiency of the modeling process by using objective, statistically valid methods. Utilization of computer-aided optimization techniques in calibration increases the reliability of the calibration outcome. By this way, the bias introduced due to the personal judgment of the modeler is minimized. Moreover, the time allocated for the process can be decreased considerably.

There are several studies on calibration methods using objective criteria as presented below. However, most of the methods used are only applicable to simple models and can have problems with complex water-quality models with large number of parameters.

Bowles and Grenney (1978) applied sequential extended Kalman filters as a technique for steady state river water quality modeling. They used the method for a real river system. In their study, they also had demonstrated the calibration capability of the filter procedure. With this procedure, coefficients in the model were estimated at the same time as the state variables. Some limitations of calibration procedure were shown, and some suggestions for improvement were given. However, applying these suggestions would increase the complexity of the procedure.

Most of the model calibration studies using optimization employ specification of an objective function, typically formulated employing a sum-of-least-squares approach (Yih and Davidson, 1975; Wood et al., 1990; Little and Williams, 1992; Mulligan and Brown, 1998; van Griensven and Bauwens, 2001). In these studies, the general optimization goal is common; aiming to minimize the error between the observed and simulated state variables. However, proposed methods in search of the best solution to the objective function vary.

In the study of Yih and Davidson (1975), the longitudinal dispersion coefficient was identified from a salinity intrusion model for estuaries. The optimum parameter was

selected by minimizing the error between the simulated and observed values by means of the least-squares criterion. Three algorithms were used for the parameter identification study: conjugate gradient, Marquardt's algorithm, and the steepest descent method. Among these methods, the Marquardt's algorithm was shown to be the most efficient one. Also, studies of the effects of parameter sensitivity to noise levels and the number of data measurement locations were performed. Marquardt's method was shown to be stable in the presence of a moderate amount of data. However, the accuracy of the algorithm was seriously affected by the number of locations of data measurements. Although the study of Yih and Davidson (1975) gave good results, the model used was quite simple relative to the complex water-quality models of today.

Wood et al. (1990) focused on the development of an expert system to aid in the calibration process and the subsequent use of a stream-quality simulation model. A biochemical oxygen demand (BOD) – dissolved oxygen model, graphics software, and a code to perform the model calibration were all linked externally to an expert system shell. In the calibration program, the expected error was evaluated by determining an average absolute-value deviation of the calculated values from the measured ones. The calibration was accomplished by using a pattern-search technique to determine the search direction that would yield a lower expected error. As the water-quality model considered four parameters, the search was conducted in a four-dimensional space. The river length was segmented into reaches and calibration was applied on these reaches separately and sequentially. This type of segmentation simplified the procedure. However, this method did not take the whole system into consideration. Therefore, the model's ability to visualize "the big picture" diminished.

The water quality variables are often highly correlated (van Griensven et al., 2002). Many model parameters affect more than a single state variable and when this is the case, it is more appropriate to estimate the parameter using the data from all affected variables simultaneously (Little and Williams, 1992). For example, BOD

decay rate affects both BOD and DO values within the system. If parameter estimation is conducted using both DO and BOD data, the estimates will be more accurate than the estimates based only on the DO data (Mulligan and Brown, 1998). Moreover, when all the output variables are used simultaneously during the calibration process, all the available information will be used. In addition, the risk of error accumulation at the end step will be reduced (van Griensven et al., 2002). However, incorporating all the output variables simultaneously to the calibration process will increase the computational complexity substantially. For such intricate systems, exercising typical nonlinear techniques for the solution may be problematic. Moreover, in these multi-dimensional complex systems, the setback of converging to local optima is prominent.

Today, the water quality models used are highly developed and complex. They can simulate a high variety of water quality constituents, and demand a high number of input parameters. For the calibration of these complex water quality models, using new, global optimization techniques may give better results compared to the traditional methods. These new techniques are more robust to messy problems such as discontinuities and difficult-to-evaluate or nonexistent derivatives (Little and Williams, 1992). Most of them use statistical, probabilistic algorithms and they are much less likely to be trapped at the local optima (Cooper et al., 1997; Goldberg, 1989).

Cooper et al. (1997) evaluated the use of global optimization methods for a conceptual rainfall runoff model. They stated that the calibration of this model is a difficult task because the response surface possesses the following: several major regions of attraction into which a search can converge, non-convexity in the vicinity of the optimal parameter set, discontinuous first and second order derivatives, presence of minor optima located near the optimal set. Because of these characteristics, analytically-based global optimization methods were inappropriate for the problem. Analytically-based global optimization methods may require restrictive conditions such as continuity and convexity. On the other hand,

probabilistically-based global optimization methods do not impose such restrictive conditions on the response surface. Cooper et al. (1997) compared three global optimization methods: shuffled complex evolution, genetic algorithms and simulated annealing. The results of the Cooper et al. (1997)'s study showed that the performances of these methods were dependent on the objective function formulation, the set of parameter values to be estimated and also on the starting position of the optimization search with respect to the global optimum. It was stated that for most situations, the shuffled complex evolution method provided better estimates of the optimum solution than genetic algorithms and simulated annealing methods. When the number of iterations for convergence was considered, the shuffled complex evaluation and genetic algorithms methods were generally more efficient than the simulated annealing method.

The study of van Griensven et al. (2002) is another example of calibration of a complex model using a global optimization method. ESWAT (Extended Soil and Water Assessment Tool) was developed by incorporating a detailed river water quality module to SWAT (the Soil and Water Assessment Tool) which had been initially developed by USDA (United States Department of Agriculture) for integrated water quantity and quality modeling of catchments. However, this integration has resulted in an increased complexity and number of uncertain parameters. The autocalibration option of the model utilized shuffled complex evolution method. This method combined several strategies in order to search over the whole parameter space and thus minimized the risk of converging to a local optimum. Shuffled complex evolution method was proved to be useful in calibration of the model. Since the model was a highly complex integrated model composed of several different models, it was recommended to conduct the calibration in combination with a preceding sensitivity analysis to screen the large set of input parameters.

Little and Williams (1992) used the Box's Complex Method, a numerical global search technique, for the calibration of QUAL2E Model. Mulligan and Brown

(1998) used genetic algorithms to calibrate the steady-state Streeter-Phelps model. In that study, genetic algorithms were compared with a more traditional optimization technique, the Marquardt algorithm, and found to be superior.

Although there exist quite a number of studies for implementing objective methods in water quality model calibration, application of such methods in practice has gained importance recently. Recent developments in the global search techniques and advancements in the computer technology will promote their employment.

2.3 QUAL2E, The Enhanced Stream Water Quality Model

QUAL2E is a comprehensive and versatile stream water quality model developed and distributed freely by United States Environmental Protection Agency (US EPA). It is intended for use as a water quality planning tool. QUAL2E is widely used and accepted as a standard model for waste load allocations, discharge permit determinations, and other conventional pollutant evaluations in the United States. There are also many modeling applications using QUAL2E throughout the world. In Turkey, it is accepted and used as a stream water quality modeling tool by DSİ.

QUAL2E can be operated either as a steady-state or dynamic model (Brown and Barnwell, 1987). When it is operated as a steady-state model, it can simulate the impact of point and non-point waste loads on the in-stream water quality. When it is operated as a dynamic model, it can be used to study the effects of diurnal variations in meteorological data and water quality. Diurnal dissolved oxygen variations due to algal growth and respiration can also be predicted. However, the effects of dynamic forcing functions, such as headwater flows or point loads cannot be modeled. Waste load allocation studies are performed for constant low flow conditions and pollutant loads.

The history of the development of QUAL2E goes back to late 1960s. Since the early 1980s, the model has been supported by EPA Center for Water Quality Modeling (CWQM), and has been enhanced to reflect a state-of-the-art water

quality model (Brown and Barnwell, 1987). Today, QUAL2E model has been incorporated as a module to the BASINS, which is a multipurpose environmental analysis system developed by EPA to be used as an integrated watershed management tool. BASINS integrates a geographic information system (GIS), watershed and meteorological data, and state-of-the-art environmental assessment and modeling tools into one package.

QUAL2E can simulate up to 15 water quality constituents in any combination desired by the user: (Brown and Barnwell, 1987)

- Dissolved Oxygen
- Biochemical Oxygen Demand
- Temperature
- Algae as Chlorophyll-a
- Nitrogen Species (Organic Nitrogen, Ammonia, Nitrite, Nitrate)
- Phosphorus Species (Organic Phosphorus, Dissolved Phosphorus)
- Coliforms
- Arbitrary Nonconservative Constituent
- Three Conservative Constituents

QUAL2E is applicable to dendritic streams that are well mixed laterally and vertically. The model can simulate any branching, one dimensional stream system. The first step in using QUAL2E is to develop a conceptual representation of the river system to be modeled. The stream system is divided into reaches that have uniform hydraulic characteristics and then these reaches are further divided into computational elements of equal length which are treated as completely mixed reactors that are linked to each other via mechanisms of advection and dispersion. The computational elements within a reach have the same hydrogeometric properties (stream slope, channel cross section, roughness, etc.) and biological rate constants (BOD decay rate, benthic source rates, algae settling rates, etc.) (Brown and Barnwell, 1987; Chapra, 1997).

For each computational element, a hydrologic balance, and a materials balance for each constituent are written. The one-dimensional transport is governed by advection, dispersion, constituent changes, and external sources/sinks. The mathematical expression for steady-flow is:

$$\frac{\partial C}{\partial t} = \frac{\partial\left(A_x D_L \frac{\partial C}{\partial x}\right)}{A_x \partial x} - \frac{\partial\left(A_x u C\right)}{A_x \partial x} + \frac{dC}{dt} + \frac{s}{V} \qquad (2.1)$$

where;

C = concentration (ML^{-3})

x = distance (L)

t = time (T)

A_x = cross-sectional area (L^2)

D_L = dispersion coefficient (L^2T^{-1})

u = mean velocity (LT^{-1})

s = external source or sinks (MT^{-1})

The advective-dispersive mass transport equation for each constituent is solved by finite difference method that employs the classical implicit backward difference scheme (Brown and Barnwell, 1987).

2.3.1 Constituents Simulated

The QUAL2E model simulates the major mechanisms and constituents that are related with the dissolved oxygen concentrations. They include the nutrient cycles, algae production, benthic oxygen demand, carbonaceous oxygen uptake, and atmospheric aeration (Brown and Barnwell, 1987). Figure 2.2 shows these mechanisms and constituents together with the way their interactions are conceptualized by QUAL2E.

The differential equation used in QUAL2E to describe the rate of change of dissolved oxygen includes atmospheric reaeration and the oxygen produced by algal photosynthesis as the major sources. The major sinks included in the equation are

biochemical oxygen demand, sediment oxygen demand, algal respiration, and the oxidation of ammonia and nitrate. Additionally, QUAL2E has the capability of modeling the reaeration that occurs when water is flowing over dams. The change in the oxygen concentration in the river is expressed as;

$$\frac{dO}{dt} = K_2(O_s - O) + (\alpha_3\mu - \alpha_4\rho)A - K_1L - K_4 / d - \alpha_5\beta_1 N_1 - \alpha_6\beta_2 N_2 \qquad (2.2)$$

where,

- O = the concentration of dissolved oxygen, mg/L
- O_s = the saturation concentration of dissolved oxygen, mg/L
- α_3 = the rate of oxygen production per unit of algal photosynthesis, mg-O/mg-A
- α_4 = the rate of oxygen uptake per unit of algae respired, mg-O/mg-A
- α_5 = the rate of oxygen uptake per unit of ammonia nitrogen oxidation, mg-O/mg-N
- α_6 = the rate of oxygen uptake per unit of nitrite nitrogen oxidation, mg-O/mg-N
- μ = algal growth rate, 1/day
- ρ = algal respiration rate, 1/day
- A = algal biomass concentration, mg-A/L
- L = concentration of ultimate carbonaceous BOD, mg/L
- d = mean stream depth, m
- K_1 = carbonaceous BOD deoxygenation rate, 1/day
- K_2 = the reaeration coefficient, 1/day
- K_4 = sediment oxygen demand rate, g/m^2-day
- β_1 = ammonia oxidation rate coefficient, 1/day
- β_2 = nitrite oxidation rate coefficient, 1/day
- N_1 = ammonia nitrogen concentration, mg-N/L
- N_2 = nitrite nitrogen concentration, mg-N/L

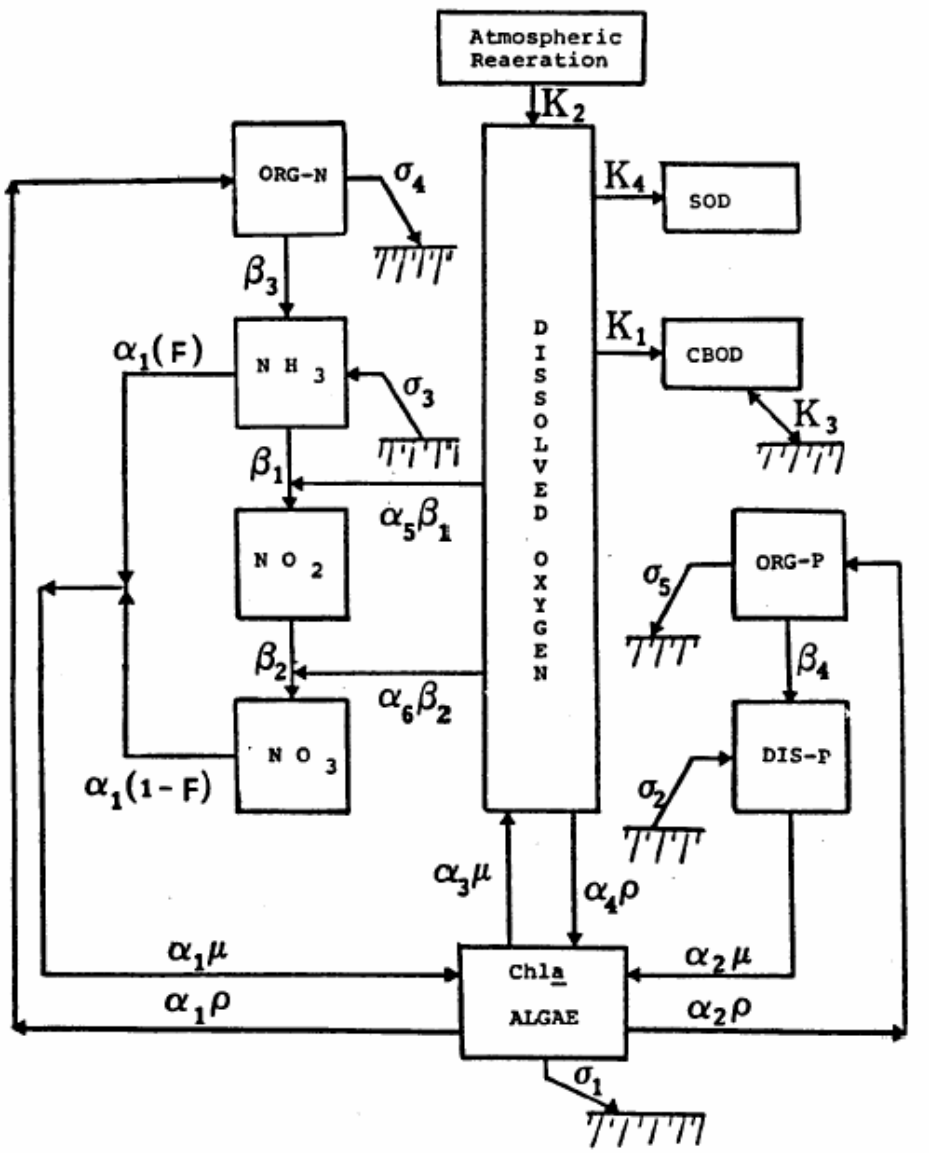

Figure 2.1: Major constituent interactions in QUAL2E. (From Brown and Barnwell, 1987)

The BOD function in the QUAL2E model uses a first order reaction to describe deoxygenation of ultimate carbonaceous BOD and also takes into account additional BOD removal by settling:

$$\frac{dL}{dt} = -K_1 L - K_3 L \tag{2.3}$$

where, K_3 = the rate of loss of carbonaceous BOD due to settling, 1/day.

The relationship that is used to model the algal biomass includes growth, respiration and settling of algae:

$$\frac{dA}{dt} = \mu A - \rho A - \frac{\sigma_1}{d} A \tag{2.4}$$

where, σ_1 = settling rate of algae, m/day

QUAL2E models the stepwise transformation of organic nitrogen to ammonia, to nitrite, and finally to nitrate. The differential equations used in modeling of these four components of nitrogen cycle consider organic nitrogen hydrolysis and settling, nitrification (oxidation of ammonia to nitrite and then nitrite to nitrate), uptake by algae, benthos source and regeneration from algal respiration. The equations are shown below:

$$\frac{dN_4}{dt} = \alpha_1 \rho A - \beta_3 N_4 - \sigma_4 N_4 \tag{2.5}$$

$$\frac{dN_1}{dt} = \beta_3 N_4 - \beta_1 N_1 + \sigma_3 / d - F \alpha_1 \mu A \tag{2.6}$$

$$\frac{dN_2}{dt} = \beta_1 N_1 - \beta_2 N_2 \tag{2.7}$$

$$\frac{dN_3}{dt} = \beta_2 N_2 - (1 - F)\alpha_1 \mu A \tag{2.8}$$

where,

N_4 = organic nitrogen concentration, mg-N/L

α_1 = fraction of algal biomass that is nitrogen, mg-N/mg-A

β_3 = organic nitrogen hydrolysis rate constant, 1/day

σ_4 = rate coefficient for organic nitrogen settling, 1/day

σ_3 = the benthos source rate for ammonia nitrogen, mg-N/mg-A

F = fraction of algal nitrogen uptake from ammonia pool

N_3 = nitrate nitrogen concentration, mg-N/L

The phosphorus cycle considers two forms of phosphorus: organic and dissolved. The equations used to model the interactions between the two components of the phosphorus cycle include organic phosphorus regeneration from algal respiration,

organic phosphorus settling, organic phosphorus conversion to dissolved state, benthal source for dissolved phosphorus, and uptake by algae:

$$\frac{dP_1}{dt} = \alpha_2 \rho A - \beta_4 P_1 - \sigma_5 P_1 \tag{2.9}$$

$$\frac{dP_2}{dt} = \beta_4 P_1 + \sigma_2 / d - \alpha_2 \mu A \tag{2.10}$$

where,

P_1 = organic phosphorus concentration, mg-P/L

α_2 = phosphorus content of algae, mg-P/mg-A

β_4 = organic phosphorus decay rate, 1/day

σ_5 = organic phosphorus settling rate, 1/day

P_2 = inorganic or dissolved phosphorus concentration, mg-P/L

σ_2 = benthos source rate for dissolved phosphorus, mg-P/L

Coliforms, the arbitrary nonconservative constituent and the conservative constituents do not interact with other constituents. For the case of coliform modeling, only a first order decay function is considered. For the case of the arbitrary nonconservative constituent modeling, in addition to a first order decay function, settling to and regeneration from the sediment are also included. The governing equations are;

$$\frac{dE}{dt} = -K_5 E \tag{2.11}$$

where,

E = coliforms concentration, colonies/100 ml

K_5 = coliform die-off rate, 1/day

$$\frac{dR}{dt} = -K_6 R - \sigma_6 R + \sigma_7 / d \tag{2.12}$$

where,

R = the nonconservative constituent concentration, mg-ANC/L

K_6 = decay rate for the constituent, 1/day

σ_6 = rate coefficient for constituent settling, 1/day

σ_7 = benthal source for constituent, mg-ANC/m^2-day

2.3.2 Calibration of QUAL2E

In the literature, many stream water-quality modeling and management studies using QUAL2E model can be found (Lung and Larson, 1995; Drolc and Zagorc Koncan, 1996; Dussaillant and Munoz, 1997; Ghosh and McBean, 1998; Chaudhury et al., 1998; Onur et al., 1999; Ning et al., 2000). However, most of these studies do not focus on the calibration problem of the model.

In the study of Onur (1996), QUAL2E was used as a tool for developing a water quality management plan for the Lower Seyhan River in Turkey. The effects of different pollution control strategies on the in-stream water-quality were analyzed. The study included a comprehensive modeling application. State variables used in the study were temperature, dissolved oxygen, 5-day biological oxygen demand, nitrogen species (ammonia, nitrite, nitrate, and organic nitrogen), dissolved phosphorus and algae. All the constants except the hydraulic coefficients needed by the model were determined by model calibration. A trial-and-error approach was adopted. Calibrations were performed by visual comparison of measured data and concentration profiles. No quantitative or statistical methods were applied.

The QUAL2E model was modified and applied to the Sava River in Slovenia by Drolc and Zagorc Koncan (1996). The focus was on the determination of the impact of wastewater discharges on the dissolved oxygen levels in the stream. The model was calibrated by field and laboratory measurements (Drolc and Zagorc Koncan, 1999). The sediment oxygen demand rate constant was determined experimentally in situ, the BOD degradation rate was determined in a river laboratory model. Hydrological characteristics were evaluated on the basis of empirical coefficients obtained from field measurements of river flow, velocity, and river depth. The reaeration rate constant was calculated with the help of an energy dissipation model. The validation results were successful although some of the determined values for the constants were quite different from the literature values. It was stated that reliance on literature values would cause substantial errors in model predictions for the Sava River case (Drolc and Zagorc Koncan, 1999).

Little and Williams's study (1992) is an example application of an objective, computer-aided method for the calibration of QUAL2E model. In their study, the focus was on the dissolved oxygen concentrations. Six parameters affecting dissolved oxygen concentrations were calibrated. A least-squares objective function that minimized the sum of squares of differences in model predictions and observations was used and solved using the Box's Complex Method. However, it was pointed out that the algorithm might not perform well if a large number of decision variables were to be estimated. Nevertheless, the study demonstrated the benefits of using an objective, statistically valid, computer-assisted method for calibration of QUAL2E model.

2.4 Genetic Algorithms

Genetic algorithms are probabilistic, global search methods that are inspired from natural evolution (Holland, 1975). Genetic algorithms perform simulated evolution on a population which is composed of potential solutions to the problem of concern. As in nature, the principle of natural selection is at work. The better members of the population, or better solutions to the problem, reproduce and pass their genetic material to the next generations while relatively worse solutions are removed from the population. In the reproduction phase, information exchange between the selected members takes place and better solutions are produced with every new generation.

Several researchers have contributed to add new and improved features to the earlier simple algorithms and broadened the area of practical application. Through many theoretical and empirical studies, it is proven that genetic algorithms are robust search methods in complex spaces (Goldberg, 1989). Today, genetic algorithms are among the most popular probabilistic global optimization methods that are preferred to the traditional optimization methods, especially for problems that contain non-linear, non-convex and discontinuous functions.

2.4.1 Overview of a Genetic Algorithm Implementation

Genetic algorithms search from a population of points, rather than a single point as in the traditional search techniques. This characteristic decreases the probability of being trapped in local optima (Goldberg, 1989). In a genetic algorithm application, the transition rules employed to guide the search process are probabilistic, not deterministic. Another important difference is that genetic algorithms are blind to the problem they are working on (Goldberg, 1989). They do not need any auxiliary information (e.g. derivative information) about the problem. Only the objective function values associated with the parameter sets are required. Also, genetic algorithms do not use the actual parameter set, instead they work on a coding of the parameter set.

There are two important mechanisms that relate the genetic algorithms to the problem they are working on; encoding and evaluation (Davis, 1991). Each potential solution to a problem is encoded into a chromosome-like structure called a 'string'. Although there are several types of encoding, such as the integer, real, and character encoding (Levine, 1996), the binary encoding is the most common method to represent the decision variables as strings. Genetic algorithm operators work on these strings to converge to the best solution. In this process, the measure of goodness may be called as the `fitness` of the string (Davis, 1991). This fitness information, returned by the evaluation function (objective function), is the only information required by a genetic algorithm to continue its search.

The general steps followed by a genetic algorithm are as follows:

1. Randomly generate an initial population of a number of strings,
2. Using the objective function, evaluate each string,
3. Create new strings applying the genetic algorithm operators;
    a. Selection
    b. Crossover
    c. Mutation

4. Form a new generation by replacing all or a portion of the old strings with the new ones,
5. Evaluate each string in the new generation,
6. If the stopping criterion is met, terminate and return the best string, if not, go back to step 3.

The cycle of applying genetic algorithm operators, forming a new generation and evaluation (steps 3, 4, 5, 6) is repeated until the stopping criterion is satisfied. The stopping criterion may be in the form of completing a previously determined number of generations or lack of improvement in best fitness value for a number of consecutive generations.

The genetic algorithm operators are responsible for the transformation of the strings and aiding in the improvement of best fitness value throughout successive generations. Three basic operators; selection, crossover and mutation, would be enough to obtain good results in many practical problems (Goldberg, 1989). Various implementation procedures for these three operators are present. Research is underway on adding new variants and totally new operators to the algorithm in order to increase the performance of genetic algorithm on specific problems. An overview on the three basic genetic algorithm operators and their implementation is given below.

Selection operator selects a number of strings based on their fitness values. Selection can be implemented in many ways. The roulette wheel parent selection method uses a biased roulette wheel where each current string in the population has a slot size proportional to its fitness (Goldberg, 1989). In the tournament selection method, two randomly selected strings are compared and the one with the better fitness value is selected.

The crossover operator is applied on selected strings generally with a high probability value (0.5-1.0) determined a priori. With this operator, the selected pair

of strings exchanges some bits. Crossover provides a search technique of high performance and gives a genetic algorithm its distinctive power. By exchanging information between the selected strings of high fitness, it is made possible to obtain even better solutions (Goldberg, 1989; Davis, 1991; Holland, 1975; Holland, 1992).

There are various types of crossover operators. In a single-point crossover, parts of two strings are exchanged at a randomly selected point. In a two-point crossover, two points are randomly selected and the string segments between these two points are exchanged. In a uniform crossover, it is randomly decided which segments of two strings will be exchanged. By this way even the smallest segments can be transferred between the selected strings (Davis, 1991). An example uniform crossover operation is represented in Figure 2.2.

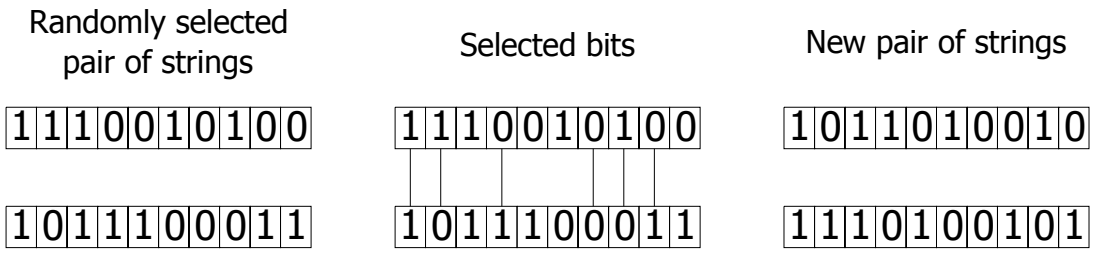

Figure 2.2: Uniform crossover on binary coded strings.

Crossover is followed by mutation. The mutation operator modifies some bits of the strings with a certain probability. Mutation probability is generally very low (0.001 – 0.04). It is used to prevent premature convergence to a local optimum and is a means of recovery for the important lost material in the reproduction and crossover processes (Goldberg, 1989).

In forming the new generation, an elitist strategy may be adopted by copying the best member of the population into the next generation. Elitism improves the genetic algorithm performance (Davis, 1991).

It is important to note that genetic algorithms are highly problem specific. The efficiency of a genetic algorithm is dependent on the proper selection of operator parameters (Ng and Perera, 2003). For most problems, using a traditional genetic algorithm with parameter values in a robust range can give satisfactory results. On the other hand, if increasing the performance of the genetic algorithm is essential, a preliminary study on the most suitable operator parameters and combinations, as well as the coding and evaluation techniques to be used, should be conducted for a specific problem (Davis, 1991).

2.4.2 Application of Genetic Algorithms to Model Calibration

Genetic algorithms have not been widely used for the optimization of calibration process in the modeling studies. However, considering the complexity of current water quality models, genetic algorithms can be deemed as strong candidates for calibration with their robustness for such problems. A few studies in the literature also strengthen this idea.

Cooper et al. (1997) applied genetic algorithms to the calibration of a conceptual rainfall-runoff model and obtained good results. They stated that calibration of conceptual rainfall-runoff models is a very complex problem and global optimization methods should be used to solve this problem. In their study, three global optimization methods were compared: shuffled complex evolution, genetic algorithms and simulated annealing. The performance of genetic algorithms were better than the simulated annealing method when the parameter estimates and the number of iterations for convergence are considered, but they did not perform as well as the shuffled complex evolution method.

Mulligan and Brown (1998) studied the applicability of genetic algorithms to calibrating water quality models. They used the steady-state Streeter-Phelps dissolved oxygen model and included the processes of reaeration, carbonaceous biological oxygen demand, nitrogenous biochemical oxygen demand, and sediment oxygen demand. Genetic algorithms were compared with a more traditional,

calculus-based nonlinear optimization technique, the Marquardt algorithm; and found to be superior. Several different formulations for incorporating field observations as constraints to the optimization problem were investigated on the base of genetic algorithm performance. The study also showed that multiple-response estimation that utilizes more than one water quality variable to estimate a model parameter is more effective. The calibration on reaeration coefficient and BOD decay rate by using both DO and BOD response gave better results compared to utilizing the DO response only. The enhancement was observed in the form of a reduction in the variability of the parameter estimates and an increase in the GA convergence rate. Another important result of the study was that the data obtained by a genetic algorithm search in calibration was shown to be informative about the response surface, parameter correlation and model sensitivity.

CHAPTER 3

METHODOLOGY

3.1 Linking QUAL2E with Genetic Algorithms

3.1.1 The QUAL2E Code Modification

QUAL2E is written in ANSI FORTRAN 77 and is compatible with mainframe and personal computer systems that support this language (Brown and Barnwell, 1987). Source code of QUAL2E is distributed freely by US EPA as a part of the program package that contains the executable program files, informative text files, sample input and output files and other utilities. A self-extracting, interactive, installation executable program file installs the whole program package to the computer. The latest DOS version of the model is the version 3.22 released by US EPA on May 1996.

The QUAL2E program is originally developed to run under the MS-DOS operating system. It is not possible to run this version in MS Windows 2000 / XP operating systems. A recent version of the program provides a graphical user interface for input and output data handling in MS Windows 98. After editing the user specified data, an interface runs the original DOS version. A special version of QUAL2E that enables the program to function under Windows 2000 is also present. However, several modifications in the operating system specifications are needed for running the Windows 2000 interface. There is no available version, pre-processor, or post-processor of the QUAL2E program for Windows XP.

QUAL2E with the version number 3.22 is structured as one main program supported by 56 different subroutines. All the subroutines are coded in separate files. There are also several include files, called as the dependencies, used for

declaring data statements, block declarations or program constants. The total number of FORTRAN source code files (*.FOR) is 57 and the total number of dependencies (*.INC, *.VAR) is 11.

Due to the modular structure of the code, the modeler is able to alter the model and adapt it to a specific stream system by modifying the related sub-programs. Moreover, new state variables can be added or modifications to existing relationships can be made through addition of appropriate subroutines. The program has to be recompiled when changes have been made in the FORTRAN source code. The original compiler used in the development of QUAL2E is Lahey FORTRAN, F77L-EM/32 version 5.01. The "make" file that aids in compilation is also distributed with the code, in order to allow modifications in the code by the user.

In this study, Microsoft Fortran PowerStation 4.0 was used in order to compile the modified QUAL2E code to run as a console application under the Windows XP operating system. First, the file specifications in the code were changed. Each compiler has its own way of reorganizing the program code and storing the floating-point variables used in the numerical processes when changing it to the form that can be understood and executed by the computer. Therefore, it would be essential to check whether the model output is correct. In order to make this analysis, the outputs of the original QUAL2E program and the modified QUAL2E program compiled with MS Fortran PowerStation 4.0 were compared. Initially, a few number of output values of some water quality variables were not the same. The reason for this variation was due to the rounding of floating-point numbers. Although, the differences were negligible, further improvement was established by recompiling the program using the '/Op' option of the compiler. This option enables improved floating-point consistency. The result of each calculation is stored in the target variable rather than being kept in the floating-point processor for usage in a subsequent calculation.

3.1.2 QUAL2E in a Genetic Algorithm

Following the modification and compilation of the QUAL2E model to run under Windows XP, study on the linkage of the QUAL2E code with a genetic algorithm (GA) started. In this study Carroll`s (2001) genetic algorithm driver in FORTRAN (version 1.7a updated in February 2001), available in a public domain, was employed.

Carroll's genetic algorithm driver uses binary encoding to represent the decision variables. The selection scheme used in the program is tournament selection with a shuffling technique for choosing the random pairs for mating. The driver includes jump mutation, creep mutation and the option for single-point or uniform crossover.

Two mutation operators, jump mutation and creep mutation, can be applied separately or at the same time. The mutation operator, referred to as the jump mutation, performs conventional mutation as described in the previous chapter. The creep mutation operates on the decoded individual (or on the real value of the individual). When subject to creep mutation, the decoded value of the individual is randomly increased or decreased by one increment. The increment value is related with the search range for that individual and the required accuracy, which is related with the string length.

Together with single-point or uniform crossover options, the number of children per pair of parents can be specified. An option for performing elitism is also present. The genetic algorithm driver can perform niching (sharing) on the population. Niching is performed for maintaining diversity in the population, and preventing the premature convergence. The fitness values of individuals are modified so that a sub-population does not become dominant over other sub-populations. By this way, many sub-populations from different regions of the search space go on searching for sub-optima in their vicinity. Niching (sharing) is especially recommended when working on multimodal search spaces containing sub-optimal but local optimum

points. ‘Sharing’ term comes from the fact that regions of search space are shared among the sub-populations which will collect on (sub-)optima.

In the recent versions of the driver, an option for the use of micro-GA operator has been added. The micro-GA technique omits the mutation operator. With this method, the genetic algorithm works on a small population of strings, called as the micro-population, and checks for the convergence of the micro-population at every generation. When convergence is achieved, a new micro-population is restarted while the best string is kept intact. However, the micro-GA operator was not used in this study.

The genetic algorithm driver contains a subroutine that evaluates the strings in a population, assigns fitness values to them, establishes the best string, and outputs the information produced. In order to assign the fitness values to each string, this subroutine calls the function evaluator (‘func’). This evaluator assigns the values of the related variables that are used to calculate the objective function value. Therefore, QUAL2E is embedded in the function evaluator subroutine.

The optimization algorithm is schematically described in the Figure 3.1. The potential solutions of the calibration problem produced by the genetic algorithm are fed into the ‘evaluation of the fitness’ phase. In this context, the potential solutions are the values of a set of kinetic parameters to be calibrated that are generated by the GA driver. They can also be called as ‘the decision variables’. The decision variables are transferred to the QUAL2E program. QUAL2E performs simulations using the transferred values. The simulations are performed both for calibration and verification conditions keeping the kinetic parameters fixed. Following the water quality simulation, results for the related water quality variables are returned for evaluation. Using the water quality values, the objective function is calculated. Then, the objective function value is returned to the GA driver to judge its fitness and application of the GA operators.

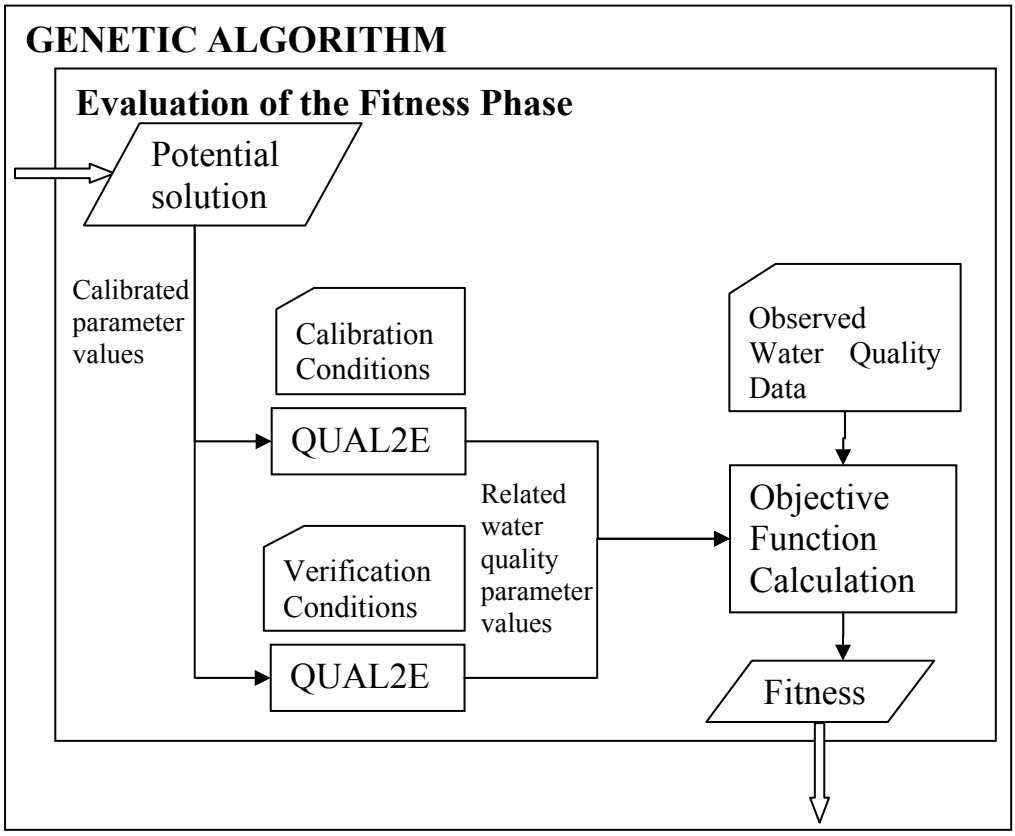

Figure 3.1: The basic optimization algorithm applied in this study.

In embedding QUAL2E in the GA driver, the following modifications were made:

1. COMMON blocks were included in both codes for the variables that were used by several subroutines,
2. Relevant commands and coding were applied for reading from and writing to the related files.
3. Relevant subroutines of QUAL2E were modified to read the values of the decision variables transferred by the genetic algorithm driver.
4. Subroutines of QUAL2E related with outputting were modified so as to return the required water quality variable values for quantification of the objective function.
5. Objective function was coded.

Following the modifications, the optimization model composed of linked QUAL2E and a GA were compiled using Microsoft Fortran PowerStation. An example code is given in Appendix A. Information about the required input files and output files are given in Appendix B.

The program needs a considerable paging file size. The used paging file size increases with every generation of the genetic algorithm run. So, the execution of the program was done through a batch-file that performs a series of genetic algorithm runs via the 'restart continuation' feature of the GA driver. The batch file also copies the necessary output files to the specified folders for further analysis. With this strategy, the memory is used more efficiently. It is possible to stop and then restart a genetic algorithm run whenever needed. Moreover, the genetic algorithm input parameters can be altered as the run is in progress.

In order to determine the most suitable GA operator parameters and the combination of operators, a preliminary study was conducted. Taking several genetic algorithm studies in the literature and the recommendations of Carroll (2001) into consideration, a set of calibration runs were made with the developed optimization model. Following these tests, the GA input parameters given in Table 3.1 were used throughout the study. The population size, however, was different for different runs, depending on the number of the decision variables. The adopted strategy in determining the population size was taking it larger than at least three times of the length (number of bits) of a string in the population.

Table 3.1: The genetic algorithm input parameters used in the study.

Parent selection scheme	Tournament selection
Crossover type	Uniform crossover, one child per pair of parents
Crossover probability	0.5
Jump mutation	YES
Jump mutation probability	0.01
Creep mutation	YES
Creep mutation probability	0.02
Elitism	YES
Niching	YES

3.1.3 The Objective Function

The objective function used in the calculation of the fitness of a string aims to minimize the difference between the observed and simulated conditions. A sum-of-least squares approach was adopted for the objective function formulation. In order to perform simultaneous calibration and verification, verification of the calibrated parameters were treated as a system constraint. This constraint was introduced as a multiplicative penalty factor. Although different forms of the objective function formulation are used throughout the study, the general mathematical expression for the objective function can be stated as:

$$Min\ Z = \sum_{j=1}^{K} w_j \left\{ \left[\sum_{i=1}^{N} \left(C_{COi\,j} - C_{CPi\,j} \right)^2 \right] \left[1 + \sum_{i=1}^{N} \left(C_{VOi\,j} - C_{VPi\,j} \right)^2 \right] \right\} \qquad (3.1)$$

where,

C_{COij} = Observed concentration of the water quality parameter *j* at observation point i for the field conditions of calibration (M/L^3)

C_{CPij} = Simulated (predicted) concentration of the water quality parameter *j* at observation point i for the field conditions of calibration (M/L^3)

C_{VOij} = Observed concentration of the water quality parameter *j* at observation point i for the field conditions of verification (M/L^3)

C_{VPij} = Simulated (predicted) concentration of the water quality parameter *j* at observation point i for the field conditions of verification (M/L^3)

N = Total number of observation points (-)

K = Total number of water quality parameters (-)

w_j = Weight of the sum-of-squared errors for the water quality parameter *j* (-)

Equation 3.1 uses the simulated and observed values of multiple water quality parameters to calculate the objective function value. The contribution of error for each water quality parameter considered is reflected in the overall error by a weighting factor (w_j). Since there may be difference of magnitude in the quantities of different water quality parameters, the contribution of error occurring for a specific water quality parameter can appear to be significant even if the error is

insignificant in terms of percentage. Therefore, weight factors are used to prevent the domination of one or more water quality parameter in the overall objective function value. By this way all the parameters are treated as equally as possible depending on the proper selection of the weight factors. However, the definition of different weights for different water quality parameters is a difficult task (van Griensven et al., 2002). Initial trials have to be conducted in order to set these weights.

For complex optimization problems with multi-modal search space, genetic algorithms can converge to local optima. Although this situation is a less likely problem for genetic algorithms compared to other optimization methods, precautions must be taken. In this study, each run was repeated three times with different randomly generated initial population of strings. In the case of obtaining different final results at the end of the optimization process, the best one was assumed as the optimum solution. All runs were lasted for 10,000 generations.

3.2 Synthetic Data Production

The assessment of the performance of the developed optimization model was initially tested based on synthetic data. Using synthetic data enables a better assessment of the method since the optimum solution is known a priori. Two different sets of data were used for the tests:

1. Perfect data
2. Biased data

The first set of tests used observed water quality constituent values that contained no error. In this set of experiments, no error was expected between the observed and simulated data for the optimum set of parameter values.

The second set of experiments was designed to test the performance of the optimization model for the observed data that contained a known amount of error. In this situation, at the end of the optimization run, the squared residuals between

the observed and the simulated data may not be zero depending on the amount of error in the observation data. Additionally, the quantities for the optimal parameter set is expected to be different from the 'real' parameter set since the optimization model tries to fit to the biased observation data.

The synthetic data to be used as the observation values were generated for a hypothetical river system of 6 reaches and 45 computational elements. This system and its relevant input files are based on 'Sample Data Set 2' (wrkshop2.dat) distributed with the QUAL2E Model System. 'Sample Data Set 2' provides all the necessary data for simulating several constituents; 5-day biological oxygen demand, dissolved oxygen, algae, phosphorus, nitrogen, fecal coliform, total dissolved solids and temperature (Figure 3.2).

The observation data at imaginary observation points (Figure 3.2) were generated for the calibration and verification conditions. In the calibration data set, the dry bulb temperature and wet bulb temperature values were 20°C and 15°C, respectively. In the verification data set, temperature values were decreased and taken as 15°C and 10°C for dry and wet bulb temperature values, respectively. Therefore, the ambient temperature differences made the distinction between the calibration and verification conditions. All other assumptions were the same. Temperature difference can have significant impacts on the reaction rates and, therefore, on the water quality. The temperature correction for the reaction rate constants is handled internally in the QUAL2E model.

The sampling points were distributed throughout the river system so as to have at least one sampling point on each reach. Placement was performed arbitrarily with care to assign them to critical locations such as the downstream of point sources, dams, junctions and upstream of withdrawals (Figure 3.2). After running the QUAL2E for the calibration and verification conditions, values of the simulated constituents at the pre-determined observation points were labeled as the observed value of the constituent at that point. For the perfect data set, the simulated water

quality constituent concentrations for the specified kinetic constants were directly used. However, to prepare the biased observation data set, the observed concentrations for the perfect data case were altered. This is achieved by randomly introducing an error within a range of ±20% for all of the water quality parameters in the observation data set.

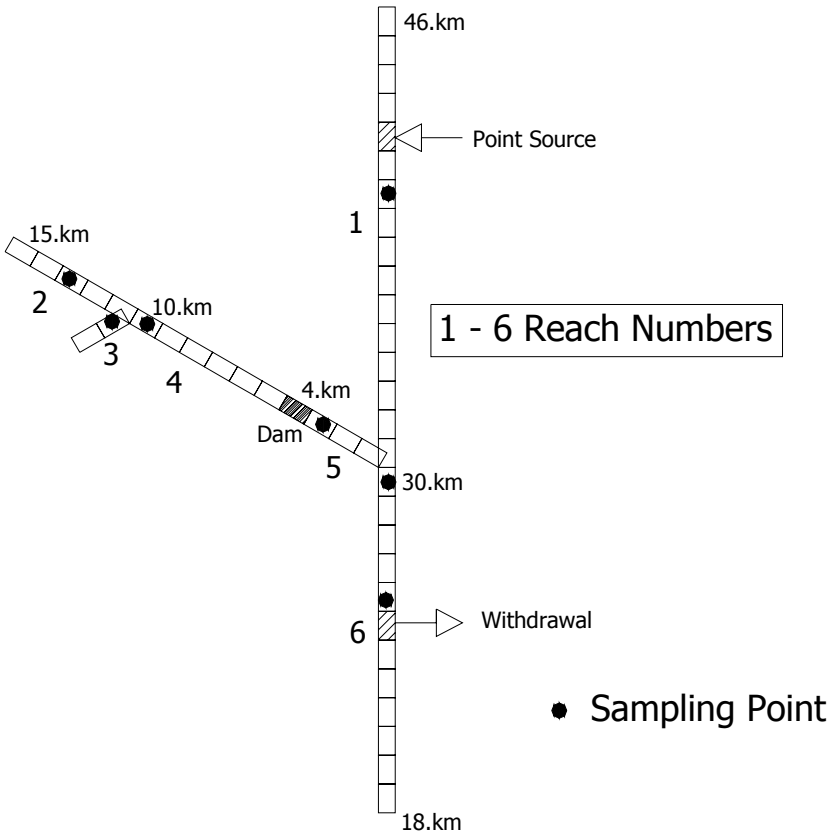

Figure 3.2: Schematic representation of the river system and sampling points

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Results for the Calibration and Verification for Two Model Parameters

The performance of the optimization model in calibration and verification of the QUAL2E based on different model parameters are analyzed based on the comparison of the optimized and real (goal) values for different reaches and comparison of water quality constituent profiles along the main stream (reaches 1 and 6) and the main tributary (reaches 2, 4, and 5). The profiles will be given for the calibration conditions only, since similar trends are observed for the verification conditions. In comparing the water quality constituent profiles the `real profile` refers to the goal profile intended to be reached by the optimized model parameter values. The `simulated profile` refers to the profile obtained using the best model parameter values produced at the end of the optimization runs.

4.1.1. Perfect Data Case

The optimization model was initially tested for the perfect observation data. The tests were designed for simultaneous calibration and verification of two kinetic constants; the reaeration coefficient (K_2) and the sediment oxygen demand (K_4). It was assumed that these constants can have variable values for the different reaches of the river system. Accordingly, for the hypothetical river system of 6 reaches, 12 decision variables were used. The relevant information about the test is given in Table 4.1. The search ranges for the coefficients were determined by referring to the observed values stated in literature and the QUAL2E model documentation (Brown and Barnwell, 1987). The range of values encoded in the genetic algorithm (GA) and other information are given in Table 4.1. The range of values used for the

sediment oxygen demand in this study corresponds to sandy bottoms and mineral soils (Chapra, 1997). However, enriched sediments may have higher values.

K_2 and K_4 values employed to prepare the observation data are given in Table 4.2. Using these coefficients, dissolved oxygen (DO) concentrations were simulated and recorded as the observation data at the sampling points given in Figure 3.2, both for the calibration and verification conditions. These concentrations are stated in Table 4.3. Basically, the optimization model uses the observation values given in Table 4.3 to reach to the K_2 and K_4 values in Table 4.2 at the end of the optimization process.

In Figure 4.1, the performance of the GA optimization run through 10,000 generations is shown for one of the tests. As depicted, as generation number increases, better solutions are obtained. In other words, when the best strings in the generations were compared, the overall error between the predicted and observed values decreases very rapidly with the evolution of the strings. However, as the error gets smaller and better strings are obtained with the transformation supplied by the GA operators, the rate of improvement decreases. The best coefficient estimates after 10,000th generation, together with the optimum coefficient values, are given in Table 4.4.

Table 4.1: Information on decision variable encoding in GA strings.

Calibrated Parameter	**Reaeration coefficient (K_2)**	**Sediment oxygen demand (K_4)**
Range	0.0 – 100 day^{-1}	0.0 – 1.0 g/m^2-day
Precision of search space	0.1day^{-1}	0.1 g/m^2-day
Total number of decision variables	12	
Total string length	84 bits	
Population size	275 strings	

Table 4.2: Reaeration coefficient (K_2) and Sediment Oxygen Demand (K_4) values used for generating the perfect observation data.

Reach Number	K_2 (day^{-1})	K_4 (g/m^2-day)
1	3.1	0.5
2	13.1	0.0
3	26.1	0.0
4	0.8	1.0
5	11.0	0.0
6	2.8	0.5

Table 4.3: The perfect observed DO data that is used in the simultaneous calibration and verification of K_2 and K_4.

Reach Number	Element Number	DO (mg/l) (calibration)	DO (mg/l) (verification)
1	7	3.38	3.78
2	3	6.71	6.86
3	2	7.23	7.69
4	1	7.52	8.11
5	2	7.21	8.08
6	1	3.47	4.33
6	5	2.98	4.04

Table 4.4: Optimization results for the simultaneous calibration and verification of K_2 and K_4. (Perfect data case).

		Observed / Optimum Value	Best Results after 10,000 generations
Objective Function Value		0	0.0005
K_2 (day^{-1})	Reach-1	3.1	2.8
	Reach-2	13.1	13.4
	Reach-3	26.1	26.1
	Reach-4	0.8	0.2
	Reach-5	11.0	15.2
	Reach-6	2.8	2.9
K_4 (g/m^2-day)	Reach-1	0.5	0.2
	Reach-2	0.0	0.3
	Reach-3	0.0	0.0
	Reach-4	1.0	0.7
	Reach-5	0.0	0.0
	Reach-6	0.5	0.7

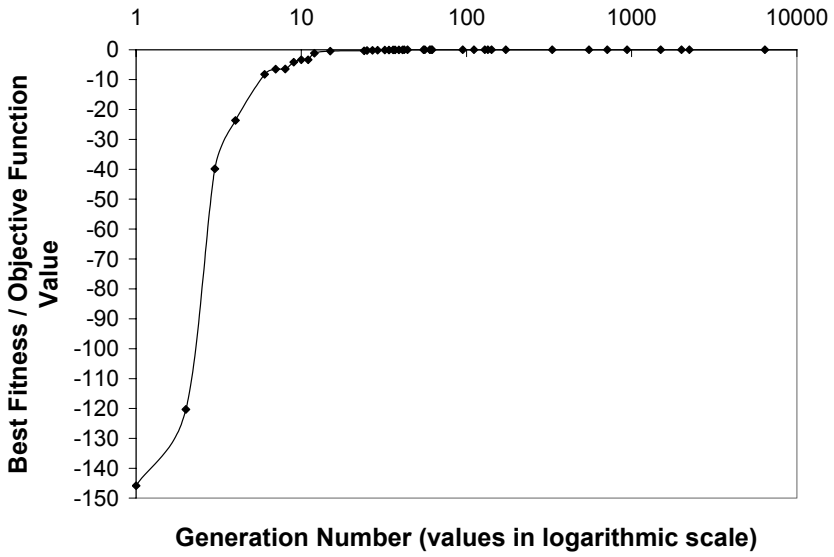

Figure 4.1: Genetic algorithm performance on simultaneous calibration and verification of K_2 and K_4.

The resultant DO profile for the main stream (reaches 1 and 6), for the coefficient estimates after 10,000 generations, is depicted in Figure 4.2 for the calibration conditions, with the observed DO values and the 'real' DO profile. It can be seen that the simulated and the real DO profiles are similar. This was the case although it was not possible to exactly reach to the goal K_2 and K_4 values (Table 4.4). Although the optimized K_2 values followed the trend in the real values along the main stream, this was not the case for K_4. Nevertheless, still a good match was obtained for the DO profiles. The maximum error in the simulated DO profile was less than 0.2 mg/L in Reach 1. The DO profile match was similar for the verification conditions. For the tributary (reaches 2, 4, and 5), variation was observed in the real and simulated DO profiles (Figure 4.3). In the 4^{th} reach, up to 1 mg/l error was observed at the downstream locations. The ability to reproduce the real DO profile with the estimated coefficients is as important as the ability to reproduce the real coefficient values. This is especially important when the mentioned coefficients are hard to determine by field or laboratory measurements. It is very common to use the coefficient values determined after calibration and verification instead of conducting field or laboratory measurements for such model coefficients. Reaeration coefficient is an example for a hard-to-measure model coefficient.

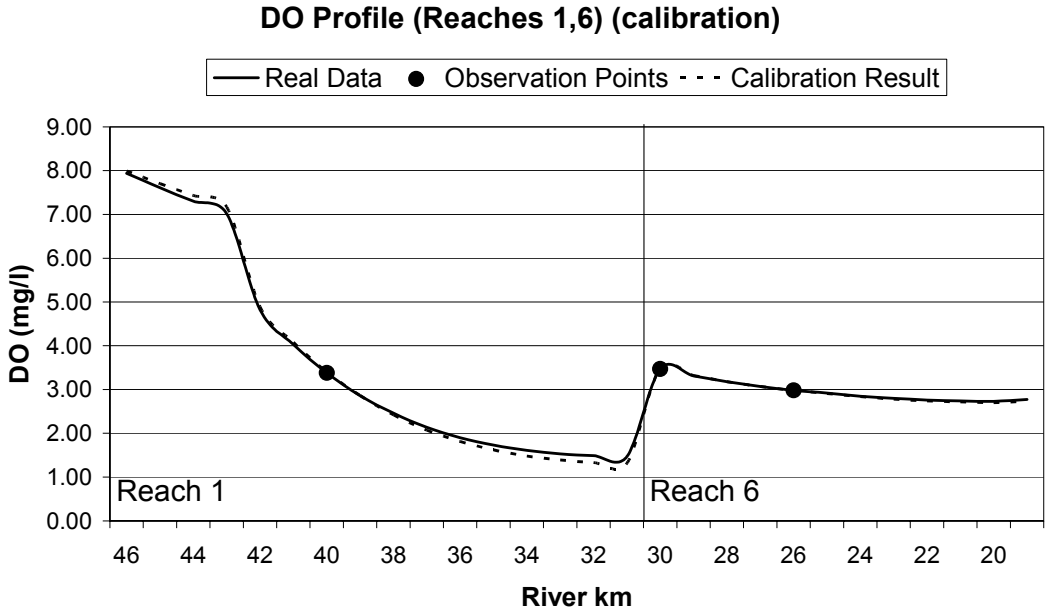

Figure 4.2: DO profile along the main river for the calibration conditions (estimated coefficients: K_2 and K_4 for perfect observation data)

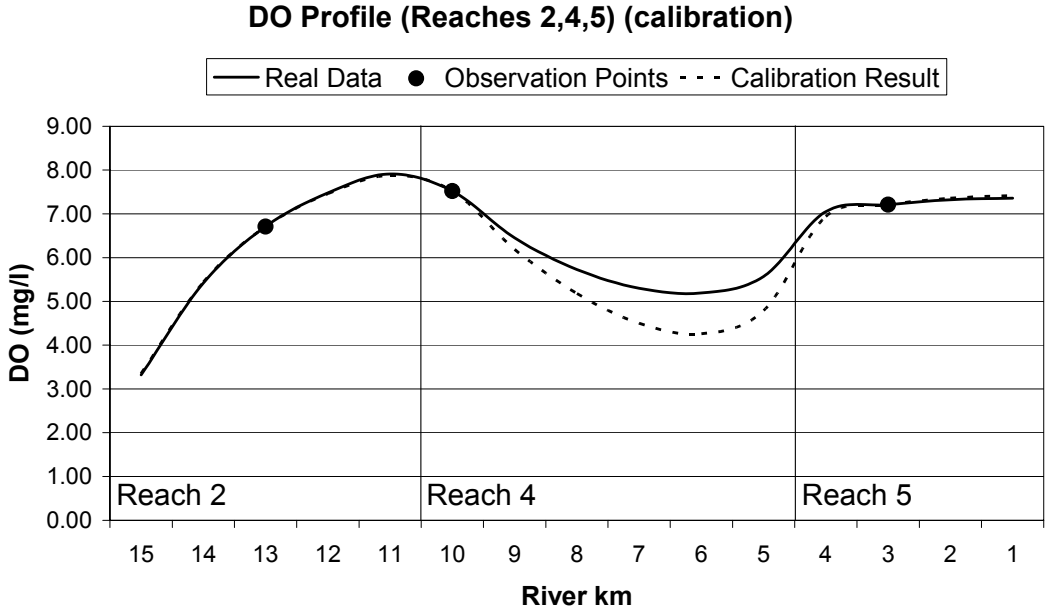

Figure 4.3: DO profile along the tributary for the calibration conditions (estimated coefficients: K_2 and K_4 for perfect observation data)

The only information that is used by the genetic algorithm in search for the optimum solution is the water quality observation data at the sampling points. When the real DO profiles and the simulated DO profiles obtained for the genetic algorithm estimates are compared in Figures 4.2 and 4.3, it can be seen that the profiles match at the sampling point locations. However, the error in the simulated DO values with respect to the real DO values may increase as the distance to the sampling point location increases.

The river reach number 4 is distinguished from the other reaches since a dam is located just downstream of it. Reach 4 actually represents a pond trapped by the dam. Consequently, the hydraulic properties of Reach 4 are different from the other reaches. The water depth is greater and the average velocity throughout the reach is slower compared to the other reaches. The decreased velocity and the increased depth cause a considerable downgrade in the reaeration rate. The optimization results reflect this lower reaeration rate. However, the estimated K_2 value (0.2 day^{-1}) is even lower than the real value (0.8 day^{-1}). When K_4 values are considered, again there is a difference with the real value (1 g/m^2-day) and the estimated value (0.7 g/m^2-day). On the other hand, although there are such differences between the real and the estimated values of the coefficients, there is no error in the simulated DO value at the sampling point (Figure 4.3). The observed DO value measured at the most upstream location of the reach reflects the impact of the oxygen rich waters of reaches 2 and 3, and it is unable to indicate the potential drop in the DO level in the downstream segments of Reach 4 due to lower reaeration in the pond (Figure 4.3). As a result, the sampling point on Reach 4 does not provide sufficient information to perform a successful calibration and verification.

4.1.2 Biased Data Case

In order to investigate the performance of optimization model for erroneous (biased) observation data, the observation DO concentrations used in the perfect data case were altered to accommodate an error of up to 20%. By this way, it is aimed to have an optimization problem that is closer to the conditions of a real

water quality modeling study. The results obtained for this set of runs were compared to the real DO profiles in the river system.

The tests with biased observation data were also comprised of simultaneous calibration and verification of K_2 and K_4, at the same time. Except the biased observation data, all other assumptions were the same as the perfect data case. Therefore, information given in Tables 4.1 and 4.2 also applies to the design of this set of runs. The biased data set used is given in Table 4.5.

Table 4.5: The biased DO observation data used in the optimization runs. (Simultaneous calibration and verification of K_2 and K_4).

Reach Number	Element Number	DO (mg/l) (calibration)	DO (mg/l) (verification)
1	7	3.27	4.30
2	3	5.44	6.61
3	2	6.51	8.87
4	1	8.85	6.81
5	2	7.34	6.58
6	1	3.94	4.82
6	5	3.19	3.85

The best decision variable estimates obtained after the 10,000th generation are given in Table 4.6. Estimated K_2 and K_4 values by the optimization model are quite different from the real K_2 and K_4 values. This difference becomes very dramatic, again, for the 4th reach. Moreover, K_4 value was estimated as 0.0 g/m^2-day in the reservoir located upstream of the dam (Reach 4), whereas high values were assigned to the reaches that actually had negligible sediment oxygen demand. This is an expected result due to the grounds mentioned in the previous section and the erroneous observation data that guides the genetic algorithm search.

Table 4.6: Optimization results for the simultaneous calibration and verification of K_2 and K_4. (Biased data case)

		Real / Optimum Value	**Best Results after 10,000 generations**
K_2 (day^{-1})	Reach-1	3.1	2.6
	Reach-2	13.1	8.8
	Reach-3	26.1	19.4
	Reach-4	0.8	97.6
	Reach-5	11.0	35.7
	Reach-6	2.8	2.4
K_4 (g/m^2-day)	Reach-1	0.5	0.2
	Reach-2	0.0	0.0
	Reach-3	0.0	0.7
	Reach-4	1.0	0.0
	Reach-5	0.0	1.0
	Reach-6	0.5	0.0

The DO profiles in the main stream and in the tributary obtained for the best estimates of K_2 and K_4 values are depicted in Figures 4.4 and 4.5, respectively, for the calibration conditions. The figures show that the simulated DO profiles differ significantly from the real DO profiles, especially for the tributary. The resulting DO profile met the biased observed DO concentrations at the sampling points on the main river. However, this was not the case for the tributary. On the other hand, the DO values simulated according to the optimization results generally had a narrower error range at the sampling points when compared with the biased observation data. In other words, the DO values obtained by simulating the DO profile, with the best coefficient estimates returned by the optimization model, are slightly closer to the real DO values than the biased observation values. The above discussion is also valid for the DO profiles for the verification conditions. The DO profile through the main stream for the verification conditions is given in Appendix C (Figure C.1).

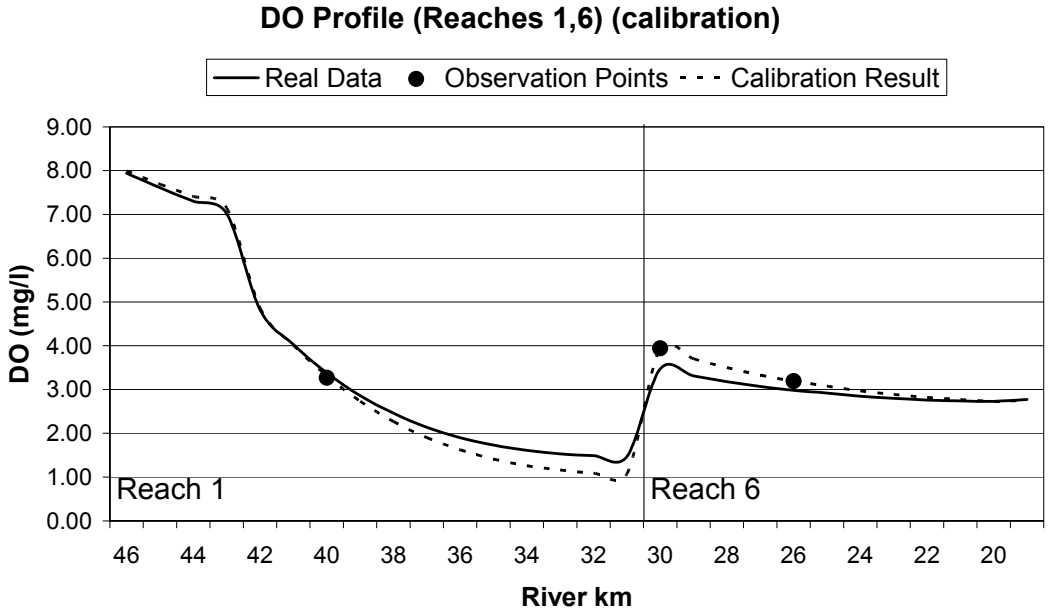

Figure 4.4: DO profile along the main river for calibration conditions. (Estimated coefficients: K_2 and K_4 for biased observation data)

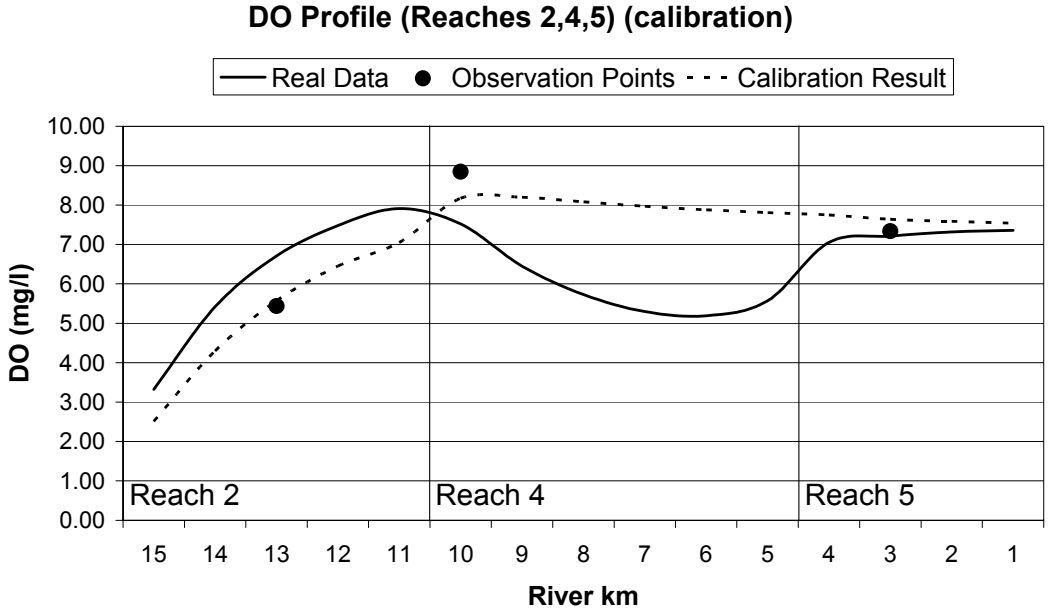

Figure 4.5: DO profile along the tributary for the calibration conditions. (Estimated coefficients: K_2 and K_4 for biased observation data)

4.2 The Impact of the Number and Location of Sampling Points

The results presented in section 4.1 suggested that the number and location of sampling points were not sufficient in defining the water quality characteristics of the reaches. In order to investigate the impact of the number and location of sampling points, two new sampling points were added. Since the maximum errors were observed at Reaches 1 and 4 for the main river and the tributary, respectively, an additional sampling point was added for each. The first additional sampling point was located just upstream of the junction point where reaches 1 and 5 met. This point is at the 31st km of Reach 1. The second sampling point was placed at the 6th km on Reach 4. This point is closer to the dam structure with respect to the existing sampling point that was located at the 10th km. The river system with 9 sampling points after the addition of 2 new sampling points is shown in Figure 4.6. The impact of the additional sampling points was investigated for both perfect and biased observation data cases.

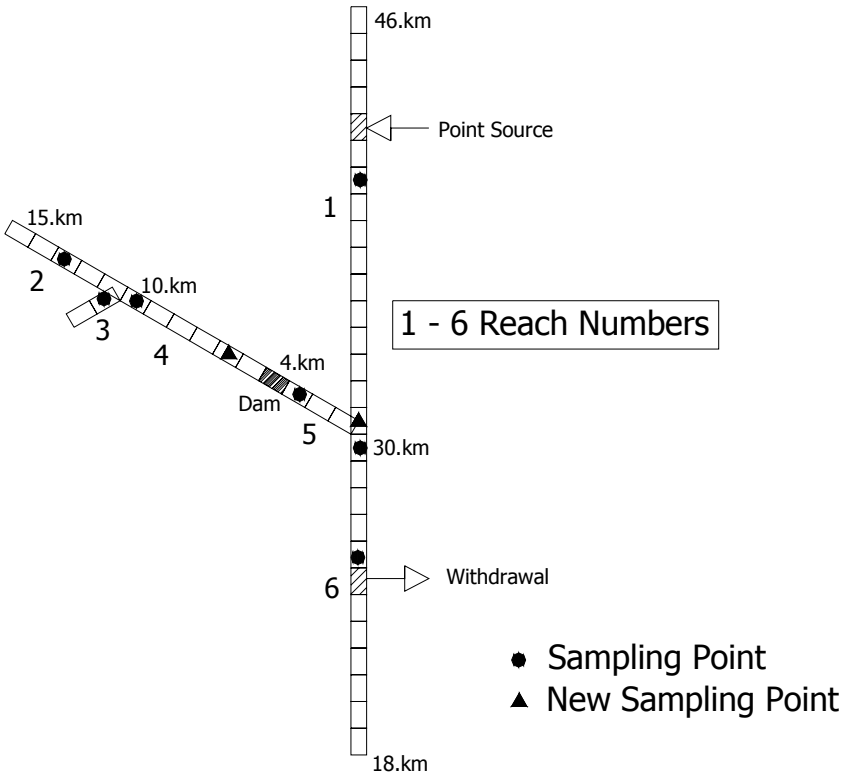

Figure 4.6: The river system with 9 sampling points.

4.2.1 Perfect Data Case

Table 4.7 presents the best K_2 and K_4 values obtained at the end of the 10,000th generation for 9 sampling points. As expected, the results were improved after the addition of new sampling points. The DO profiles obtained by using the coefficient estimates for the 9 sampling point case are plotted in Figures 4.7 and 4.8 for the main river and the tributary, respectively. The figures show that the deviations between the simulated and the real DO profiles were decreased considerably by the addition of new sampling points.

In order to compare the calibration and verification results of the runs with 9-sampling points with that of the runs with 7-sampling points, the difference between the simulated and real DO concentrations through all the computational elements were analyzed and the maximum amount of error through the river reaches were determined. The maximum error was nearly 1 mg/L for the 7-sampling-point-case, whereas it was only 0.05 mg/L for the 9-sampling-point-case. These results indicate that using a high number of sampling points can provide a better representation of the river system and water quality, leading to a more successful calibration and verification. However, there is a trade-off between the costs of additional sampling points and the success of a calibration and verification of a water quality model.

Table 4.7: Optimization results for the simultaneous calibration and verification of K_2 and K_4. (Perfect data case with 9 sampling points)

		Real/ Optimum Value	**Best Results**
Objective Function Value		0	0.0022
K_2 (day^{-1})	Reach-1	3.1	3.1
	Reach-2	13.1	13.2
	Reach-3	26.1	25.9
	Reach-4	0.8	0.7
	Reach-5	11.0	12.9
	Reach-6	2.8	2.9
K_4 (g/m^2-day)	Reach-1	0.5	0.5
	Reach-2	0.0	0.2
	Reach-3	0.0	0.0
	Reach-4	1.0	0.8
	Reach-5	0.0	0.4
	Reach-6	0.5	0.7

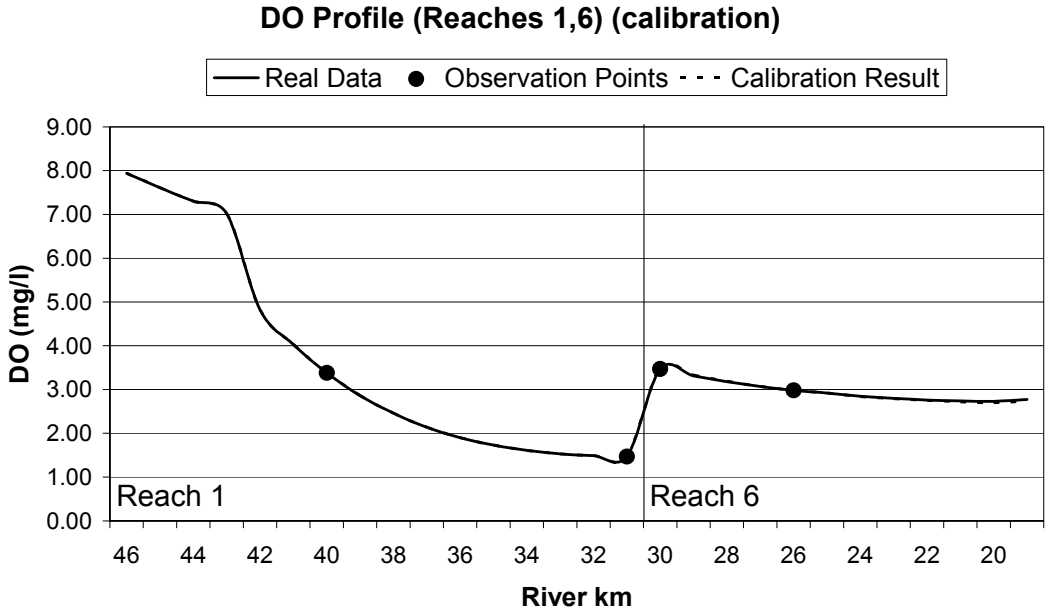

Figure 4.7: DO profile along the main river for the 9-sampling points case for the calibration conditions. (Estimated coefficients: K_2 and K_4 for perfect observation data)

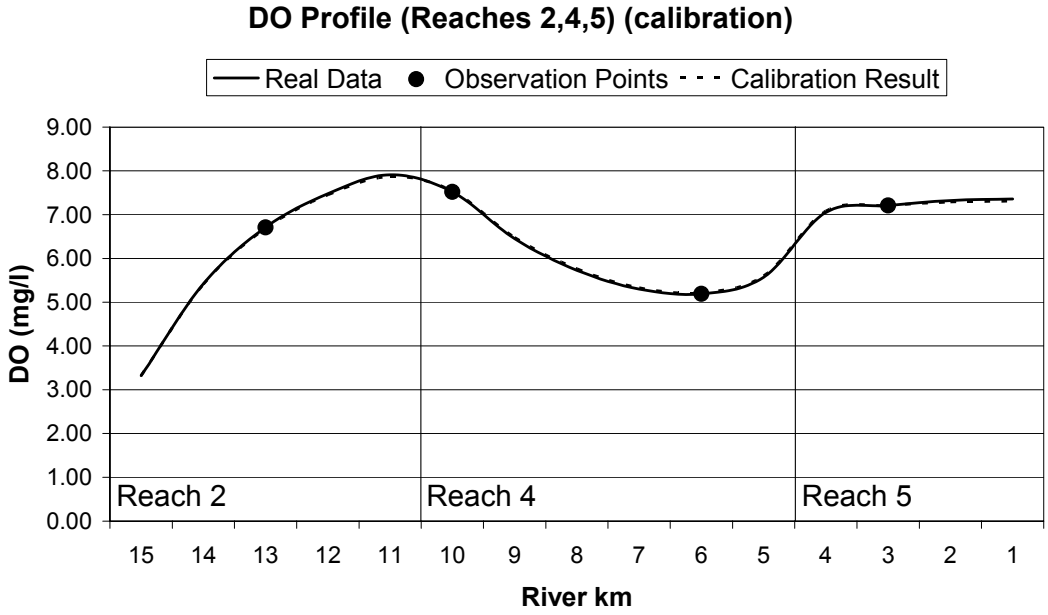

Figure 4.8: DO profile along the tributary for the 9-sampling points case for the calibration conditions. (Estimated coefficients: K_2 and K_4 for perfect observation data)

4.2.2 Biased Data Case

As discussed in section 4.1.2, results were unsuccessful for the 7-sampling points. In this section it is tested whether improvement can be achieved in predicting the parameter values and DO profiles for increased sampling points. The best K_2 and K_4 values obtained by the optimization process are stated in Table 4.8. Since the observation data used in the optimization runs were erroneous, the estimated coefficient values were quite different from the real coefficient values. However, compared to the runs with 7-sampling points, significant improvement was achieved. This is also observable in the DO profiles of the main river and the tributary shown in Figures 4.9 and 4.10, respectively. The DO profile for the verification conditions through the main stream is given in Appendix C (Figure C.2).

Table 4.8: Optimization results for the simultaneous calibration and verification of K_2 and K_4 (Biased data case with 9 sampling points)

		Real/Optimum Value	**Best Results after 10,000 generations**
K_2 (day^{-1})	Reach-1	3.1	3.1
	Reach-2	13.1	9.5
	Reach-3	26.1	52.1
	Reach-4	0.8	0.0
	Reach-5	11.0	13.7
	Reach-6	2.8	2.6
K_4 (g/m^2-day)	Reach-1	0.5	0.3
	Reach-2	0.0	0.0
	Reach-3	0.0	1.0
	Reach-4	1.0	0.0
	Reach-5	0.0	0.3
	Reach-6	0.5	0.0

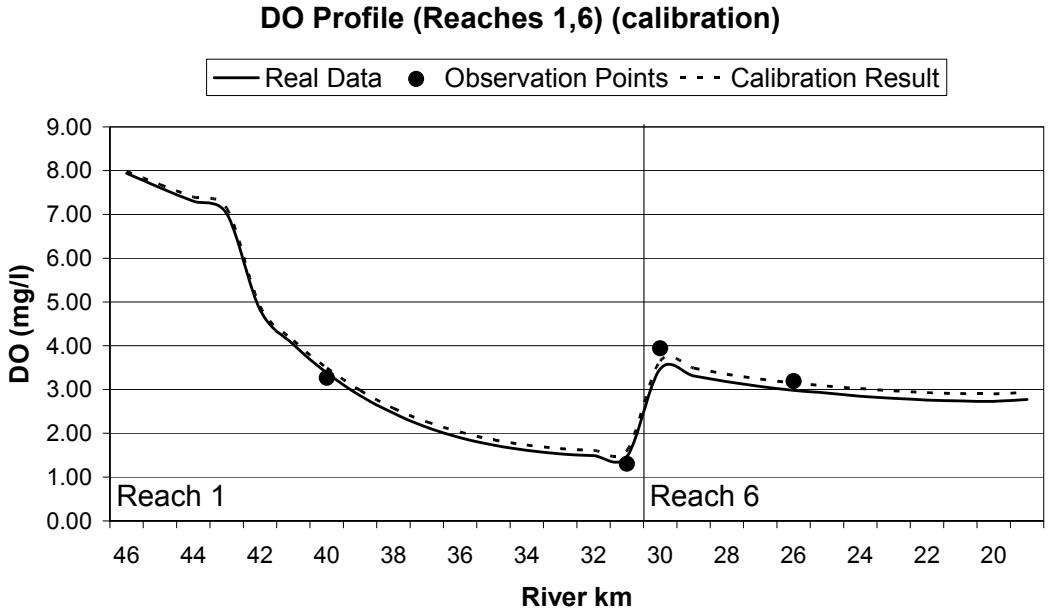

Figure 4.9: DO profile along the main river for 9 sampling points case for the calibration conditions. (Estimated coefficients: K_2 and K_4 for biased observation data)

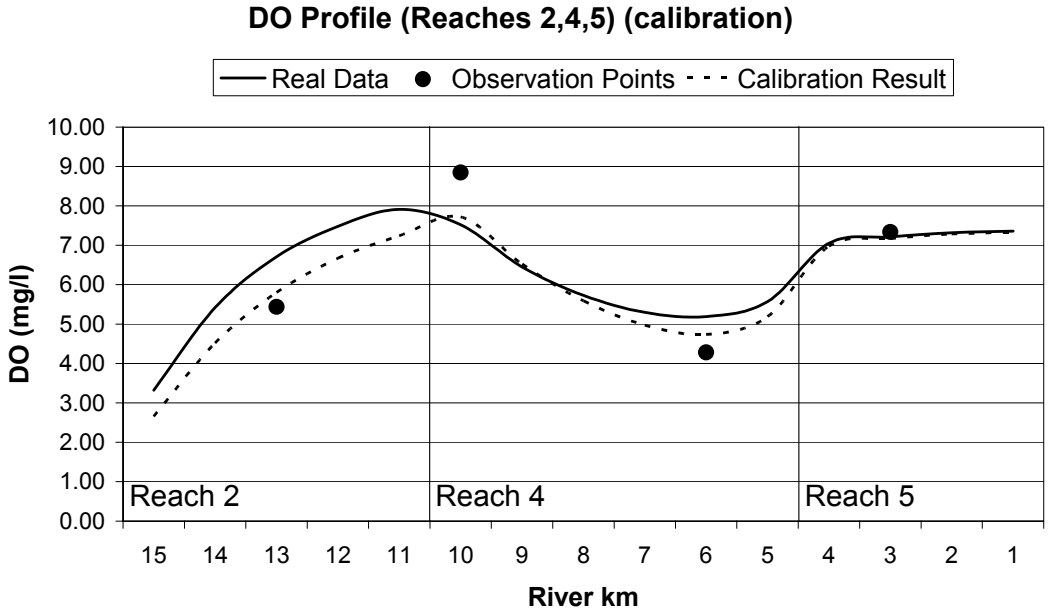

Figure 4.10: DO profile along the tributary for the 9 sampling points case for the calibration conditions. (Estimated coefficients: K_2 and K_4 for biased observation data)

The simulated DO profile along the main river has a very small deviation from the real DO profile although the utilized observation data was biased. Results point out that additional sampling points have a positive impact on the simulated DO profiles, especially in the tributary. Although, the simulated and real DO concentrations in Reach 2 may differ by as much as 1 mg/l at some locations, the simulated DO profiles follow the same trend with the real DO profile in the tributary. The error amounts in the simulated DO profiles are very closely related with the magnitude of bias in the observed DO values.

4.3 The Impact of Different Objective Function Formulations

The test results presented in Sections 4.1 and 4.2 were obtained using the objective function formulation given in Chapter 3, Methodology. In this section, the impact of different objective function formulations on the optimization performance is discussed. The analysis of this impact was performed for the 7-sampling-point case for which the performance was not satisfactory. Two new objective functions were formulated to check whether the performance of the optimization model can be improved by the objection function formulation itself. The results obtained for the same problem with three different objective functions were compared. The formulations of these three objective functions are given below:

Objective Function – 1:

$$Min\,Z = \left[\sum_{i=1}^{N}\left(C_{COi} - C_{CPi}\right)^2\right]\left[1 + \sum_{i=1}^{N}\left(C_{VOi} - C_{VPi}\right)^2\right] \tag{4.1}$$

Objective Function – 2:

$$\begin{aligned} Min\,Z = &\left[\sum_{i=1}^{N}\left(C_{COi} - C_{CPi}\right)^2\right]\left[1 + \sum_{i=1}^{N}\left(C_{VOi} - C_{VPi}\right)^2\right] \\ &+\left[\sum_{i=1}^{N}\left(C_{VOi} - C_{VPi}\right)^2\right]\left[1 + \sum_{i=1}^{N}\left(C_{COi} - C_{CPi}\right)^2\right] \end{aligned} \tag{4.2}$$

Objective Function – 3:

$$Min\ Z = \left[\sum_{i=1}^{N} \left(C_{COi} - C_{CPi} \right)^2 \right] \left[1 + w\left(D\max_C \right) \right] \left[1 + w\left(D\max_V \right) \right] \qquad (4.3)$$

where,

$Dmax_C$: The maximum error at the sampling point locations for the calibration conditions (mg/l),

$Dmax_V$: The maximum error at the sampling point locations for the verification conditions (mg/l),

w : The weight of the penalty function = 100

As previously stated, the developed optimization algorithm performs simultaneous calibration and verification. Verification results were treated as a constraint and incorporated into the objective function formulation as a penalty function. The general objective function (Equation 3.1) used in the previous tests has a drawback that may cause poor performance. As the difference between the simulated and observed values gets smaller, the verification error may become less significant. Even if the error from the verification phase is large, the objective function may still get small values in the case of small errors in magnitude for the calibration conditions. In order to avoid such a situation, the Objective Function – 2 (Equation 4.2) was developed. This function is equal to the sum of two sub-functions. The first sub-function is the Objective Function – 1. In the second sub-function, the sum-of-the-squared errors for the calibration conditions are treated as a penalty function.

A different approach was adopted when formulating the Objective Function – 3 (Equation 4.3). Instead of taking the sum-of-the-squared errors at all of the sampling points into consideration in calculating the penalty function value, the maximum amount of errors for the calibration and verification conditions are selected and used in the penalty function. In this case, the objective function value is only affected by the maximum error values, not by the error values at all of the sampling points.

4.3.1 Perfect Data Case

The runs executed to obtain the results given in Section 4.1.1 were repeated using the Objective Functions – 2 and – 3. The estimated K_2 and K_4 values for three different objective functions are given in Table 4.9. It is seen that the set of coefficient estimates for Objective Function – 3 is relatively better compared to the set obtained with other objective function formulations.

Table 4.9: Optimization results for the simultaneous calibration and verification of K_2 and K_4 by using different objective function formulations. (Perfect data case)

		Observed / Optimum Value	**Objective Function - 1**	**Objective Function - 2**	**Objective Function - 3**
K_2 (day^{-1})	Reach-1	3.1	2.8	2.9	3.1
	Reach-2	13.1	13.4	13.2	13.3
	Reach-3	26.1	26.1	25.7	26.0
	Reach-4	0.8	0.2	0.3	0.8
	Reach-5	11.0	15.2	14.1	10.6
	Reach-6	2.8	2.9	2.9	2.6
K_4 (g/m^2-day)	Reach-1	0.5	0.2	0.3	0.5
	Reach-2	0.0	0.3	0.1	0.2
	Reach-3	0.0	0.0	0.0	0.0
	Reach-4	1.0	0.7	0.7	1.0
	Reach-5	0.0	0.0	0.0	0.0
	Reach-6	0.5	0.7	0.7	0.2

In Figures 4.11 and 4.12, DO profiles plotted using the results obtained for three different objective function formulations are given. When these DO profiles are compared, it is seen that the simulated DO values are very close to the real DO values at the sampling point locations for all of the three objective functions. However, for the same final GA generation number, the results obtained by using Objective Function – 3 (Equation 4 .3) are nearly equal to the real DO values following the similarity of the estimated K_2 and K_4 values with their real values throughout all the river system (at all of the computational elements). For the other objective functions' results, there are some deviations from the real DO values. These deviations are smaller for the results obtained by the tests with Objective Function – 2 when compared with that of Objective Function – 1.

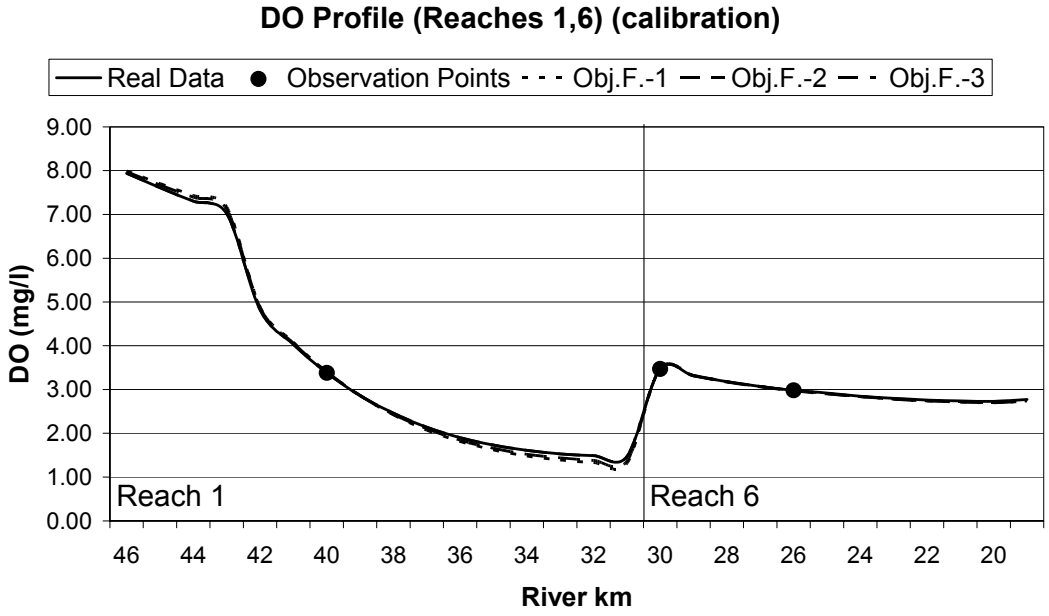

Figure 4.11: DO profiles along the main river for three different objective function formulations for the calibration conditions. (Estimated coefficients: K_2 and K_4 for perfect observation data)

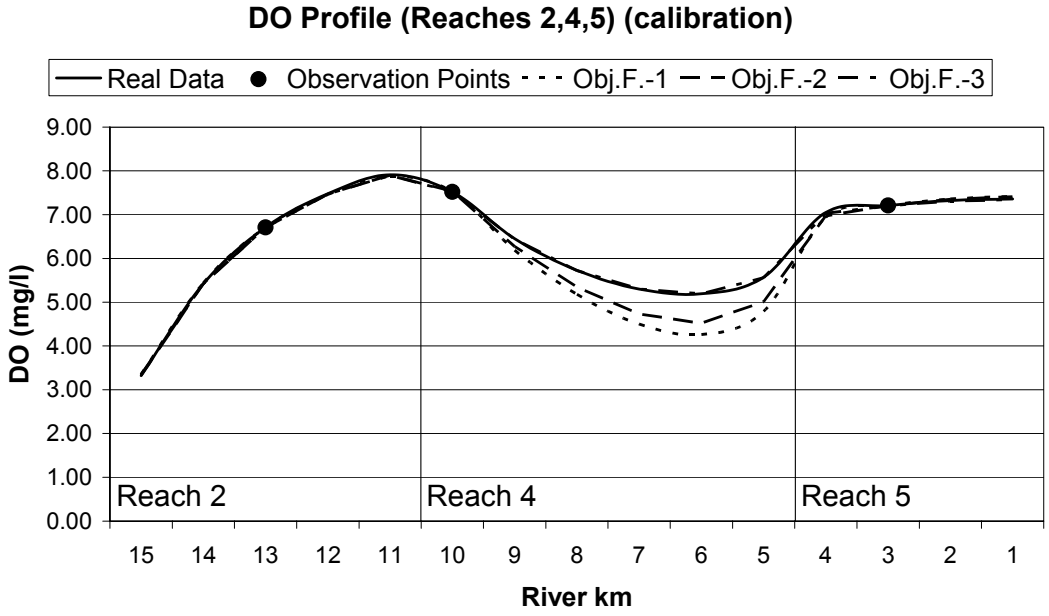

Figure 4.12: DO profiles along the tributary for three different objective function formulations for the calibration conditions. (Estimated coefficients: K_2 and K_4 for perfect observation data)

Figures 4.11 and 4.12 enable a visual comparison of the DO profiles for three different objective function formulations for the main river and the tributary, respectively. As can be noticed, the profiles obtained using Objective Function-2 are better compared to the ones for the Objective Function-1. However, the performance of optimization with Objective Function-3 was far better than the other two. First of all, it was possible to obtain coefficient estimates that were very close to the real ones (Table 4.9). As a result, very small error was observed in the simulated DO concentrations throughout the river system in general as given in Table 4.10. These results suggest that the objective function formulation can be significant in terms of convergence for the perfect data case. It was decided to use Objective Function – 3 as the base objective function in the more complex problems presented in the following section given that it was the best performing objective function among the tested ones.

Table 4.10: The maximum error values in the DO values determined after optimization runs with three different objective functions (mg/l). (Perfect data case)

Objective Function	**All the computational elements**				**Sampling Points**			
	Reaches 1, 6		**Reaches 2, 4, 5**		**Reaches 1, 6**		**Reaches 2, 4, 5**	
	Calib.	**Verif.**	**Calib.**	**Verif.**	**Calib.**	**Verif.**	**Calib.**	**Verif.**
1	0.16	0.20	0.93	0.89	0.02	0.01	0.00	0.01
2	0.11	0.14	0.67	0.66	0.02	0.01	0.01	0.01
3	0.02	0.05	0.02	0.02	0.01	0.03	0.01	0.01

4.3.2 Biased Data Case

The previous tests given in Section 4.1.2 were repeated using Objective Functions -2 and -3. The resultant K_2 and K_4 values are given in Table 4.11. As can be seen, the coefficients are in error for most of the reaches. However, the resulting DO profiles were acceptable for all of the objective functions as depicted in Figures 4.13 and 4.14 for the main river and the tributary, respectively. The DO profile for the verification conditions through the main stream is given in Appendix C (Figure C.3).

The performance difference for different objective functions, which was clearly seen in the optimization results for the perfect observation data case, was not observable for the biased observation data case. In these tests, the error in the observation data dominated the outcome and the impact of objective function formulation became insignificant. This was the case especially for the tributary.

Table 4.11: Optimization results for the simultaneous calibration and verification of K_2 and K_4 by using different objective function formulations. (Biased data case)

		Observed / Optimum Value	**Objective Function - 1**	**Objective Function - 2**	**Objective Function - 3**
K_2	Reach-1	3.1	2.6	3.1	3.1
	Reach-2	13.1	8.8	9.4	8.8
	Reach-3	26.1	19.4	19.4	15.8
	Reach-4	0.8	97.6	76.8	98.8
	Reach-5	11.0	35.7	23.3	27.5
	Reach-6	2.8	2.4	2.6	2.5
K_4	Reach-1	0.5	0.2	0.5	0.6
	Reach-2	0.0	0.0	0.5	0.2
	Reach-3	0.0	0.7	0.6	0.3
	Reach-4	1.0	0.0	0.2	0.0
	Reach-5	0.0	1.0	0.6	1.0
	Reach-6	0.5	0.0	0.2	0.2

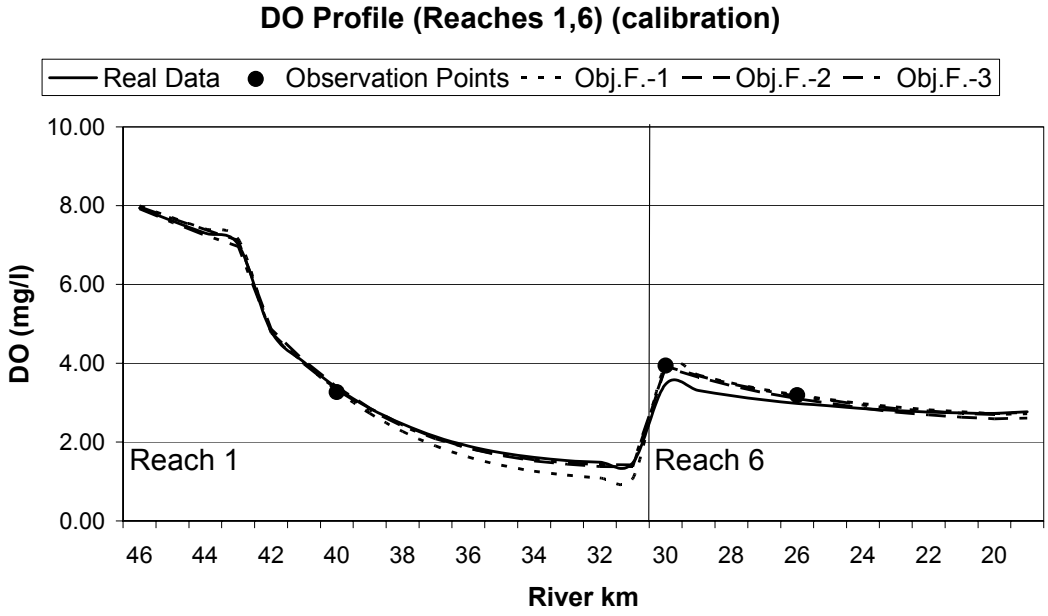

Figure 4.13: DO profiles along the main river for three different objective function formulations for the calibration conditions. (Estimated coefficients: K_2 and K_4 for biased observation data)

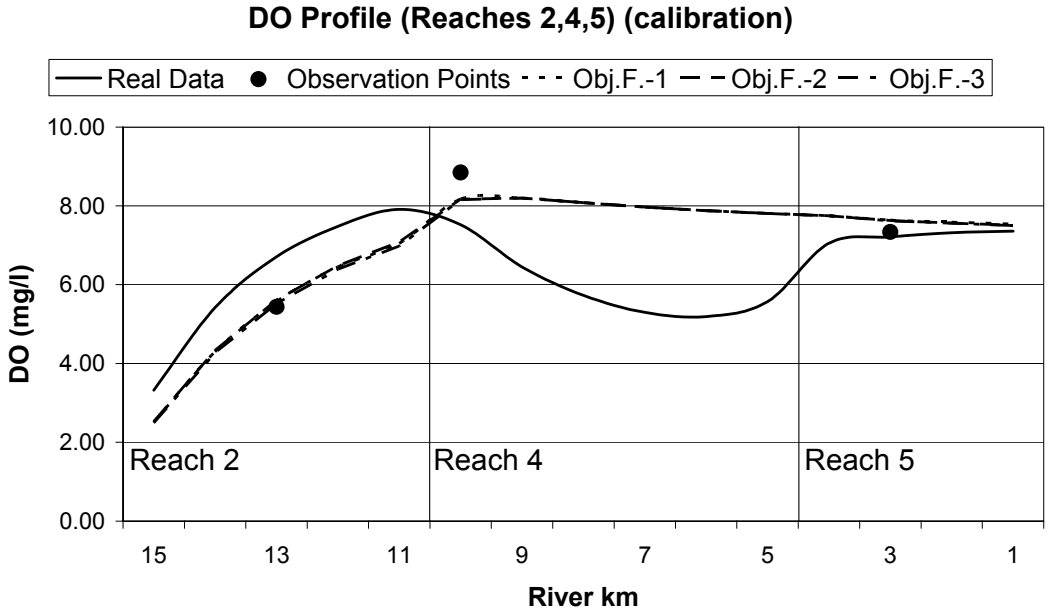

Figure 4.14: DO profiles along the tributary for three different objective function formulations for the calibration conditions. (Estimated coefficients: K_2 and K_4 for biased observation data)

4.4 Performance of the Optimization Model in Complex Problems

In this section, the performance of the optimization problem was tested for complex calibration and verification problems. Here, the term 'complex' stands for the problem of performing calibration and verification for the estimation of a large number of kinetic coefficients, use of multiple water quality variables in the observation data set to guide the optimization, or verification of the model for multiple verification conditions. Since the previous tests had been performed on relatively simple calibration and verification problems, some modifications in the optimization model was required in order to acquire the results in this section. These modifications caused an evolution of the optimization model increasing its general applicability to real modeling studies.

The river system that was based on the QUAL2E sample input file, 'wrkshop2.dat', was continued to be used. In the light of the results of the previous tests, it was decided to use Objective Function – 3 as a basis for the more complex problems since it was the best performing objective function among the tested ones. Also, the observation data set represented the measurements at the sampling points of the 9-sampling point case.

The level of complexity of the problem was gradually increased resulting in three different "Complex Problem" cases. In Complex Problem I, the observed data of both DO and BOD were used to determine four reach-variable kinetic coefficients related with DO and BOD. In Complex Problem II, the nitrogen (N) species were added to the observation data set and the estimated kinetic coefficients were a total of nine reach variable kinetic coefficients related to DO, BOD and N species. Finally, in Complex Problem III, Complex Problem II was repeated by verifying the calibrated coefficients by using two observation data sets instead of one. Every test was performed for both perfect and biased observation data.

4.4.1 Adapting the Optimization Model to Complex Problems

The increased number of coefficients to be estimated requires more computational effort, since higher number of decision variables results in a long string length for the genetic algorithm. In order to make use of multiple water quality variables, the objective function formulation used in the previous tests needs to be modified to enable the minimization of the differences between the simulated and observed values of all the related variables. The objective function that reflects this requirement is given in Equation 4.4.

$$\text{Min } Z = \sum_{j=1}^{K} w_j E_j \tag{4.4}$$

$$E_J = \left[\sum_{i=1}^{N} \left(C_{COi\,j} - C_{CPi\,j} \right)^2 \right] \left[1 + w\left(D\max_{C\,j} \right) \right] \left[1 + w\left(D\max_{V\,j} \right) \right] \tag{4.5}$$

where,

E_j= error function value for the water quality variable j (mg/L)

K = total number of water quality variables in the observation data set (-)

The error function, E_j, is similar to the Objective Function-3. Therefore, Equation 4.4 is analogous to finding the value of Objective Function-3 separately for each water quality variable and then summing up these values after multiplication with a proper weight (w_j) to calculate the magnitude of the objective function. In this formulation, the assignment of the proper weights is performed according to the typical range of error magnitudes for a particular water quality variable. This range is dependent on the river system, the observed value of the variable, and the search range of the related kinetic coefficient to be estimated. The determined set of weight values should ensure an equal contribution from each water quality variable to the overall objective function value as much as possible, at least in terms of the orders of magnitude.

In this study, a linear normalization approach was adopted to determine the weight values. The error function values that a particular water quality variable can have

throughout the optimization runs were mapped in a range of 0 to 1. The following formula was used for the normalization:

$$X_{j-norm} = \frac{X_j - X_{j-\min}}{X_{j-\max} - X_{j-\min}} \tag{4.6}$$

where,

$X_{j\text{-norm}}$: Normalized value of the error function for water quality variable j

X_j : Original value of the error function for water quality variable j

$X_{j\text{-min}}$: Minimum value of the error function for water quality variable j

$X_{j\text{-max}}$: Maximum value of the error function for water quality variable j

It is known that the minimum value that the error function can take is 0 (no error). When this value is inserted into Equation 4.6, the normalization formula reduces to:

$$X_{j-norm} = \frac{X_j}{X_{j-\max}} \tag{4.7}$$

In order to determine $X_{j\text{-}max}$, the initial population of a genetic algorithm population is used. When generating the initial population, genetic algorithm randomly samples the search space. If we assume that the sample size which is equal to the population size is large enough, then the maximum error function value within the GA population can be accepted as equal to $X_{j\text{-max}}$. There is a possibility that the error function value of that particular variable can take greater values in the proceeding generations resulting in a normalized error function value of greater than 1. This situation is not expected to cause any problems in the computations. In fact, when this situation occured in runs, such values were only slightly greater than 1 in most cases. With this approach, the objective function is stated as follows:

$$Min\, Z = \sum_{j=1}^{K} \frac{1}{X_{j-\max}} \left\{ \left[\sum_{i=1}^{N} \left(C_{COi\,j} - C_{CPi\,j} \right)^2 \right] \left[1 + w\left(D\max_{C\,j} \right) \right] \left[1 + w\left(D\max_{V\,j} \right) \right] \right\} \tag{4.8}$$

The tests of Complex Problem I and II are performed by using both weighted and unweighted (all weights equal to 1) forms of the objective functions. Both results are presented for comparison.

In the case of utilizing more than one data sets for the verification (Complex Problem III), the maximum error value at the sampling point locations is determined for all of the verification conditions and the largest one is taken as the $Dmax_{vj}$. With this definition for $Dmax_{vj}$, Equation 4.8 is applicable to any number of verification data sets.

4.4.2 Complex Problem I

In this set of runs, observed DO and BOD data were used at the same time to estimate the related QUAL2E kinetic coefficients, namely, BOD decay rate coefficient (K_1), BOD settling rate (K_3), K_4, and K_2. All coefficients are assumed to be reach-variable resulting in a GA optimization problem with 24 decision variables. The information about the problem set is given in Table 4.12. The search ranges of the coefficients were determined by referring to the QUAL2E documentation (Brown and Barnwell, 1987), except that of the sediment oxygen demand.

Table 4.12: Information on decision variable encoding in GA strings for Complex Problem I.

Calibrated Parameter	**BOD decay rate coefficient (K_1)**	**BOD settling rate (K_3)**	**Sediment oxygen demand (K_4)**	**Reaeration coefficient (K_2)**
Range	0.02 – 3.4 day^{-1}	-0.36 – 0.36 day^{-1}	0.0 – 1.0 g/m^2-day	0.0 – 100 day^{-1}
Required accuracy	0.01 day^{-1}	0.01 day^{-1}	0.1 g/m^2-day	0.1day^{-1}
Total number of decision variables	24			
Total string length	180 bits			
Population size	550			

4.4.2.1 Perfect Data Case

The observation data set used for the perfect observation data assumption is given in Table 4.13. The weights assigned to the error functions of DO and BOD in the weighted runs are given in Appendix D (Table D.1). The original values of the coefficients to be estimated and the optimization outcomes for the weighted and unweighted objective functions are given in Table 4.14. As was the case for other runs, it was not possible to obtain the exact coefficient values. However, the consequential DO and BOD profiles were satisfactory. The DO profiles are depicted in Figures 4.15 and 4.16 for the main river and the tributary, respectively. Respective BOD profiles are shown in Figures 4.17 and 4.18.

The DO and BOD profiles obtained for the optimized QUAL2E coefficients using the weighted and unweighted runs match with each other with very small deviations from the real ones. By visual comparison, it can be said that the results for the weighted runs are slightly better than that of the unweighted runs.

Table 4.13: The perfect observation data used in the simultaneous calibration and verification of BOD decay rate coefficient (K_1), BOD settling rate (K_3), sediment oxygen demand (K_4) and reaeration coefficient (K_2).

Reach Number	Element Number	DO (mg/l)		BOD (mg/l)	
		calibration	verification	calibration	verification
1	7	3.38	3.78	20.37	20.55
1	16	1.47	2.30	15.16	15.94
2	3	6.71	6.86	1.90	1.91
3	2	7.23	7.69	17.51	17.74
4	1	7.52	8.11	5.26	5.43
4	5	5.19	6.20	2.24	2.67
5	2	7.21	8.08	1.67	2.12
6	1	3.47	4.33	10.85	11.56
6	5	2.98	4.04	10.12	11.03

Table 4.14: Optimization results for the simultaneous calibration and verification of BOD decay rate coefficient (K_1), BOD settling rate (K_3), sediment oxygen demand (K_4) and reaeration coefficient (K_2). (Perfect data case)

		Observed / Optimum Value	Unweighted Runs	Weighted Runs
K_1 (day^{-1})	Reach-1	0.6	0.78	0.72
	Reach-2	0.6	0.89	1.38
	Reach-3	0.6	0.65	0.73
	Reach-4	0.6	0.97	0.70
	Reach-5	0.6	1.02	0.74
	Reach-6	0.6	0.56	0.87
K_3 (day^{-1})	Reach-1	0.0	-0.21	-0.14
	Reach-2	0.0	-0.10	-0.36
	Reach-3	0.0	-0.02	-0.17
	Reach-4	0.1	-0.34	-0.03
	Reach-5	0.0	-0.28	0.17
	Reach-6	0.0	0.10	-0.29
K_4 (g/m^2-day)	Reach-1	0.5	0.1	0.1
	Reach-2	0.0	0.6	0.3
	Reach-3	0.0	0.3	0.7
	Reach-4	1.0	0.3	0.7
	Reach-5	0.0	0.6	0.3
	Reach-6	0.5	0.1	0.3
K_2 (day^{-1})	Reach-1	3.1	3.8	3.5
	Reach-2	13.1	13.8	13.8
	Reach-3	26.1	34.0	32.5
	Reach-4	0.8	1.1	0.8
	Reach-5	11.0	12.7	11.5
	Reach-6	2.8	2.3	3.6

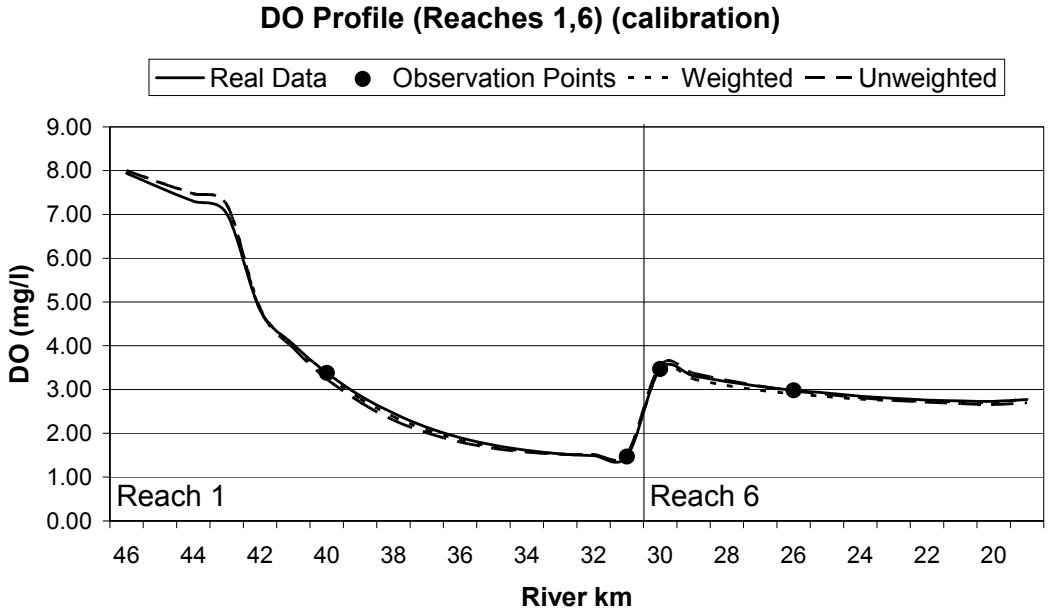

Figure 4.15: DO profile along the main river for calibration conditions. (Estimated coefficients: K_1, K_3, K_4 and K_2 for perfect observation data)

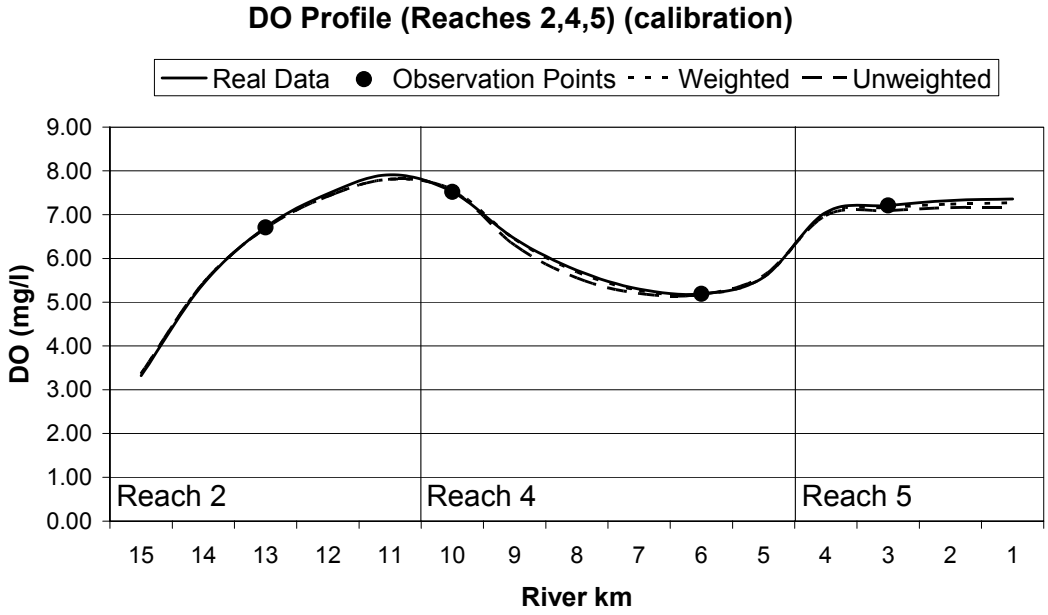

Figure 4.16: DO profile along the tributary for calibration conditions. (Estimated coefficients: K_1, K_3, K_4 and K_2 for perfect observation data)

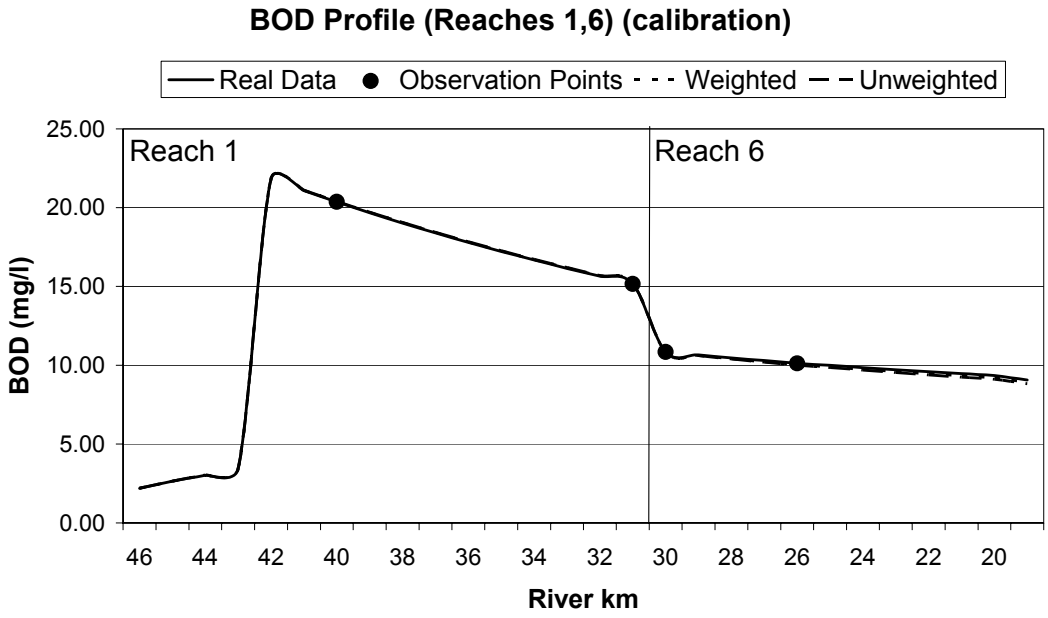

Figure 4.17: BOD profile along the main river for calibration conditions. (Estimated coefficients: K_1, K_3, K_4 and K_2 for perfect observation data)

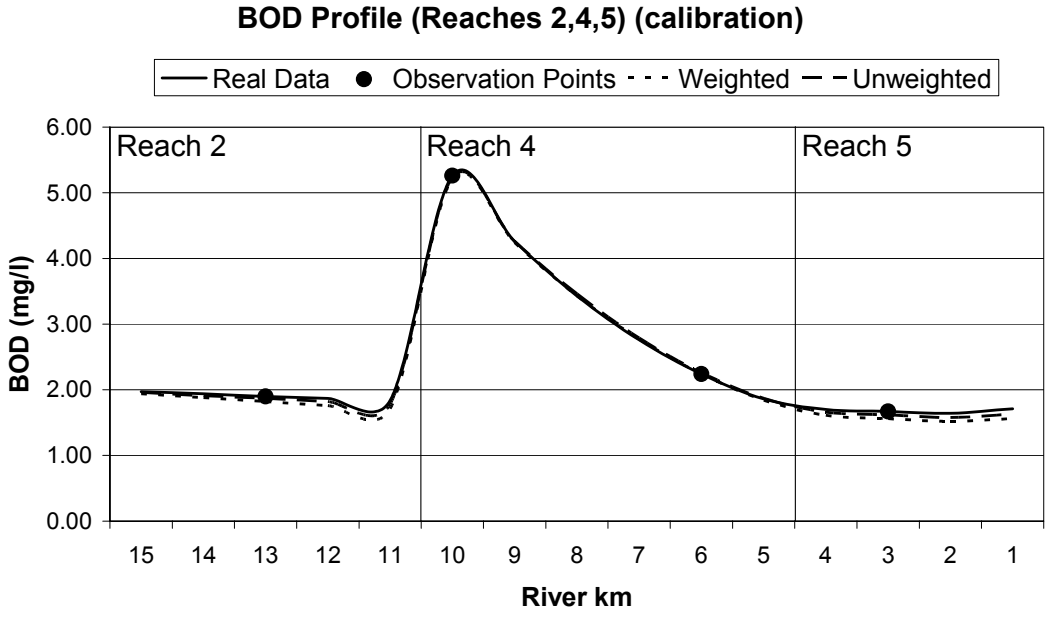

Figure 4.18: BOD profile along the tributary for calibration conditions. (Estimated coefficients: K_1, K_3, K_4 and K_2 for perfect observation data)

The deviations of the DO and BOD values from the real ones are represented in Table 4.15 quantitatively. In general, the maximum errors throughout the river and the sampling points are smaller when weights are used in the objective function calculations. When the abiliy to reproduce the 'real' DO and BOD profiles is considered, both weighted and unweighed objective functions can be accepted to be successful.

Table 4.15: The maximum error values in the DO and BOD values determined after optimization runs with weighted and unweighted objective functions (mg/l) (Complex Run I) (Perfect data case)

Objective Function	All the computational elements				Sampling Points			
	Main river		Tributary		Main river		Tributary	
	Calib.	Verif.	Calib.	Verif.	Calib.	Verif.	Calib.	Verif.
DO (mg/L)								
Weighted	0.21	0.19	0.10	0.08	0.08	0.10	0.05	0.08
Unweighted	0.20	0.20	0.19	0.25	0.14	0.11	0.12	0.19
BOD (mg/L)								
Weighted	0.15	0.10	0.15	0.13	0.08	0.10	0.11	0.09
Unweighted	0.24	0.19	0.08	0.15	0.11	0.15	0.05	0.15

4.4.2.2 Biased Data Case

The biased observation data used in this test is given in Table 4.16. The biased observed DO and BOD values are generated following the same approach that varies the quantities in the range of ±20%. The weights assigned to the error functions of DO and BOD in the weighted runs are given in Appendix D (Table D.1). The results obtained using the weighted and unweighted runs are given in Table 4.17. The DO and BOD profiles are shown in Figures 4.19 to 4.22 for the calibration conditions. The profiles for the verification conditions through the main stream are given in Appendix C (Figures C.4 and C.5).

Table 4.16: The biased observation data used in the simultaneous calibration and verification of BOD decay rate coefficient (K_1), BOD settling rate (K_3), sediment oxygen demand (K_4) and reaeration coefficient (K_2).

Reach Number	Element Number	DO (mg/l)		BOD (mg/l)	
		calibration	verification	calibration	verification
1	7	3.27	4.30	17.75	23.97
1	16	1.30	2.65	18.00	15.54
2	3	5.44	6.61	1.76	2.04
3	2	6.51	8.87	19.49	14.27
4	1	8.85	6.81	6.19	4.40
4	5	4.28	7.37	2.44	2.18
5	2	7.34	6.58	1.70	2.37
6	1	3.94	4.82	8.88	11.54
6	5	3.19	3.85	11.33	11.54

Table 4.17: Optimization results for the simultaneous calibration and verification of BOD decay rate coefficient (K_1), BOD settling rate (K_3), sediment oxygen demand (K_4) and reaeration coefficient (K_2). (Biased data case)

		Observed / Optimum Value	Unweighted Runs	Weighted Runs
K_1 (day^{-1})	Reach-1	0.6	0.64	0.64
	Reach-2	0.6	0.06	0.10
	Reach-3	0.6	0.52	0.97
	Reach-4	0.6	0.78	0.70
	Reach-5	0.6	3.33	2.51
	Reach-6	0.6	0.53	0.53
K_3 (day^{-1})	Reach-1	0.0	-0.07	-0.07
	Reach-2	0.0	-0.13	0.13
	Reach-3	0.0	0.13	-0.34
	Reach-4	0.1	-0.27	-0.17
	Reach-5	0.0	-0.36	-0.03
	Reach-6	0.0	-0.36	-0.22
K_4 (g/m^2-day)	Reach-1	0.5	0.2	0.3
	Reach-2	0.0	0.0	0.2
	Reach-3	0.0	0.2	0.3
	Reach-4	1.0	0.0	0.0
	Reach-5	0.0	0.0	0.0
	Reach-6	0.5	0.0	0.4
K_2 (day^{-1})	Reach-1	3.1	3.0	3.0
	Reach-2	13.1	9.3	10.0
	Reach-3	26.1	53.1	64.9
	Reach-4	0.8	0.5	0.4
	Reach-5	11.0	34.4	26.4
	Reach-6	2.8	2.2	2.6

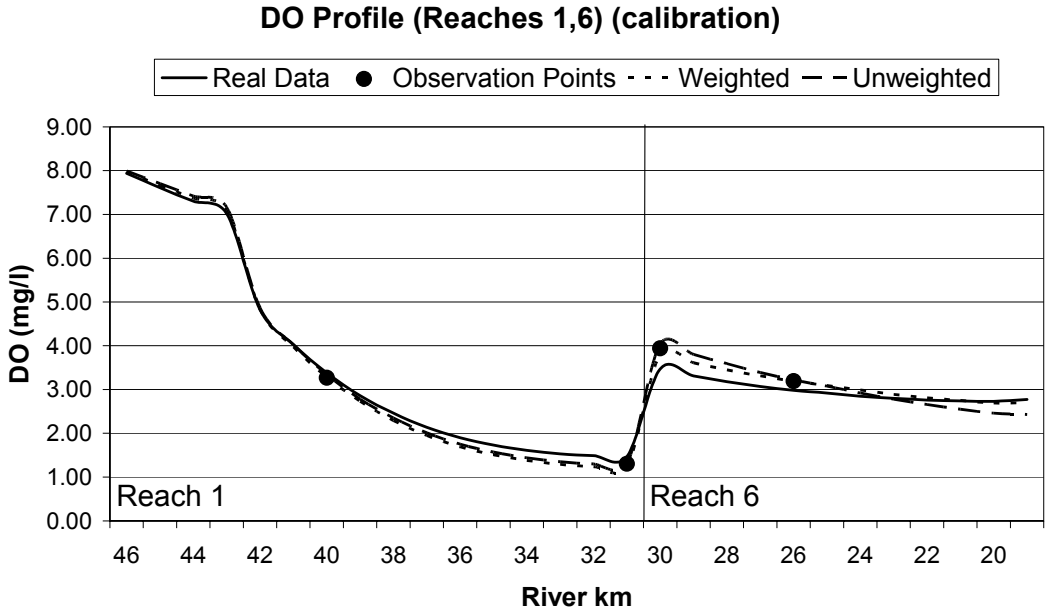

Figure 4.19: DO profile along the main river for the calibration conditions. (Estimated coefficients: K_1, K_3, K_4 and K_2 for biased observation data)

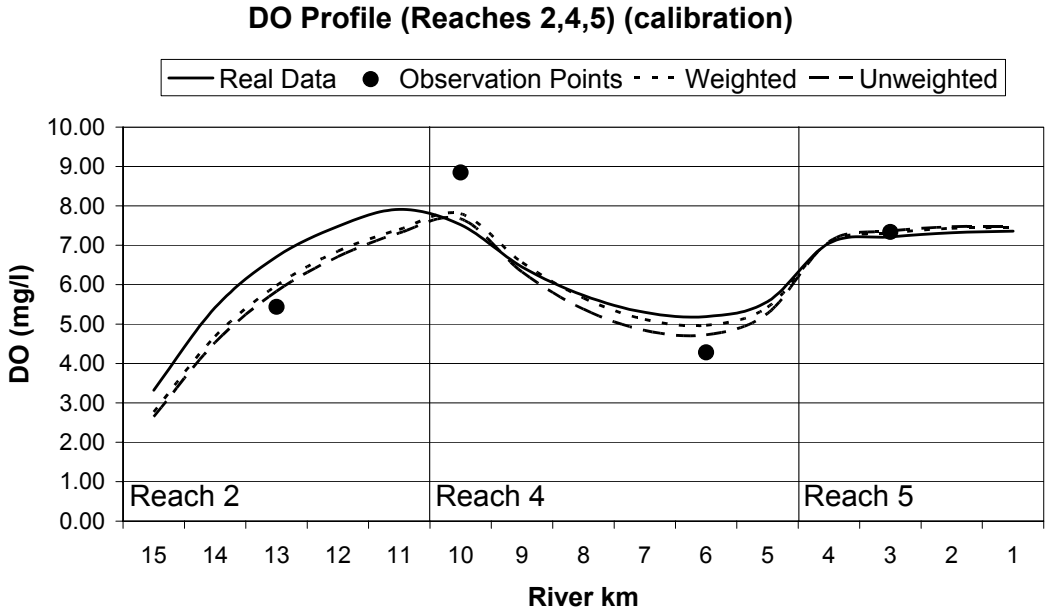

Figure 4.20: DO profile along the tributary for the calibration conditions. (Estimated coefficients: K_1, K_3, K_4 and K_2 for biased observation data)

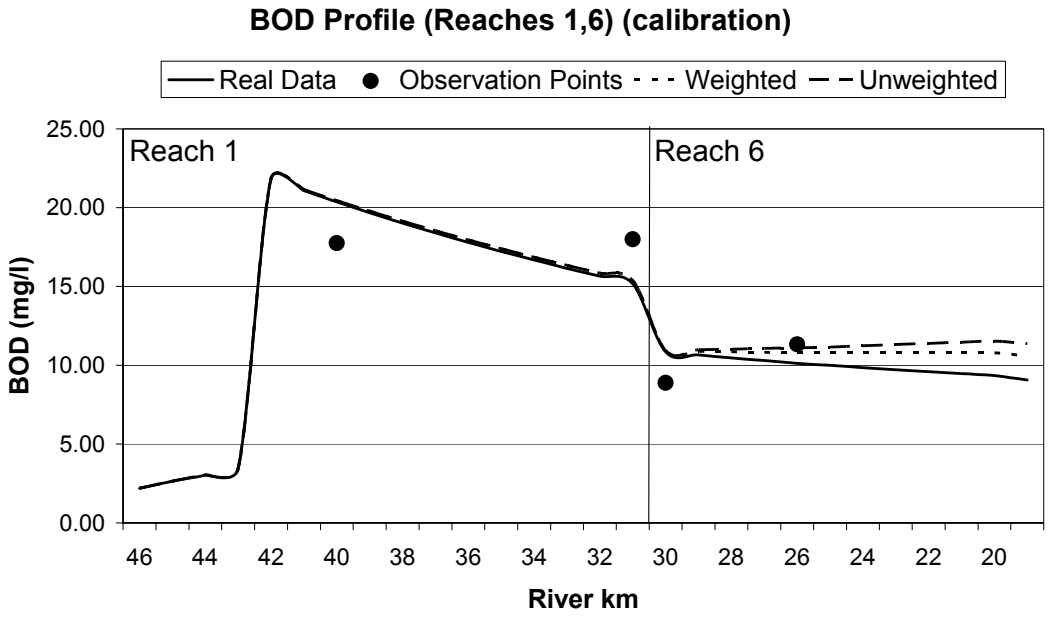

Figure 4.21: BOD profile along the main river for the calibration conditions. (Estimated coefficients: K_1, K_3, K_4 and K_2 for biased observation data)

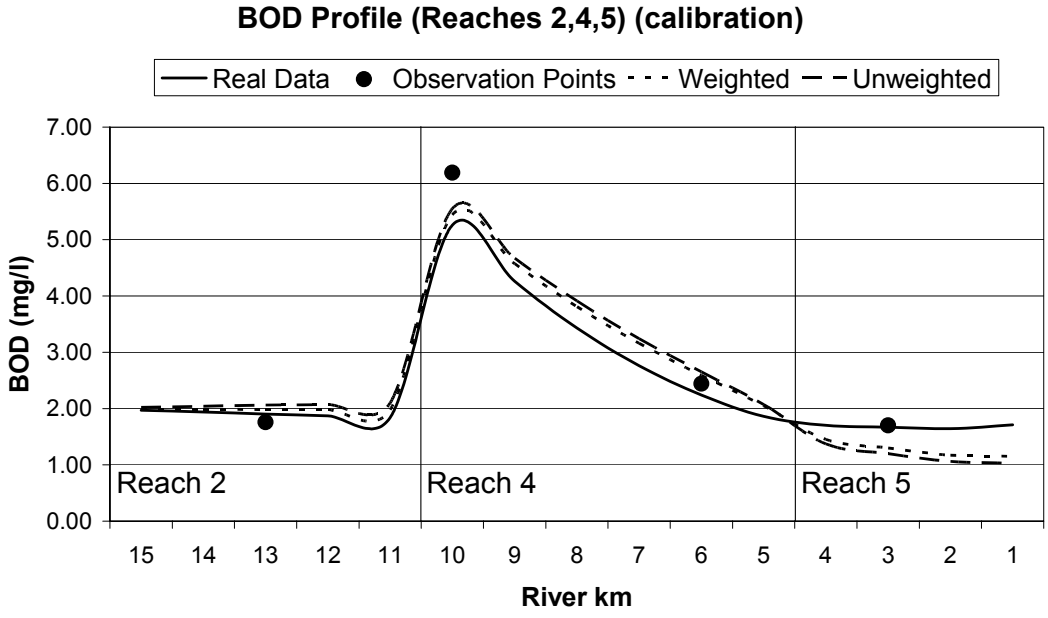

Figure 4.22: BOD profile along the tributary for the calibration conditions. (Estimated coefficients: K_1, K_3, K_4 and K_2 for biased observation data)

The estimated coefficient values are different from the real values of the coefficients. This result is parallel to the results of the previous tests. When the DO and BOD profiles are analyzed, a distortion caused by the biased observation data is easily seen. However, the profiles plotted according to the optimization results lie between the biased observation data points and the real profiles. Therefore, for the cases studied, the errors at the sampling points in the resultant profiles are smaller than the error in the observation data. Although the inherent error in the observed BOD data is similar in relative magnitude to the error in observed DO data, the resultant BOD profiles deviate less from the real profiles when compared with the deviation of the resultant DO profiles from the real profiles. When the profiles obtained from the results of 'weighted' and 'unweighted' runs are compared, it is seen that the estimates of the 'weighted' runs give better profiles with respect to the deviation from the real profiles.

4.4.3 Complex Problem II

In this test, the concentration of the nitrogen (N) species that are simulated by QUAL2E were added to the observed data set in addition to the DO and BOD values. Then, the complexity of the calibration problem was further increased by adding the QUAL2E kinetic coefficients related with the N-cycle. The resultant problem had 9 kinetic coefficients to estimate, namely; BOD decay rate coefficient (K_1), BOD settling rate (K_3), sediment oxygen demand (K_4), reaeration coefficient (K_2), organic-N hydrolysis rate (B_3), organic-N settling rate (S_4), ammonia oxidation rate (B_1), benthos source rate for ammonia (S_3) and nitrite oxidation rate (B_2). All the coefficients were assumed to be reach-variable resulting in an optimization problem with 54 decision variables. Information about the string mapping is given in Table 4.18. The search ranges of the coefficients were determined by referring to the QUAL2E documentation (Brown and Barnwell, 1987), except that of the sediment oxygen demand and the benthos source rate for ammonia-N.

Table 4.18: Information on decision variable encoding in GA strings for Complex Problem II.

DO and BOD coefficients				
Calibrated Parameter	**BOD decay rate coefficient (K_1)**	**BOD Settling rate (K_3)**	**Sediment oxygen demand (K_4)**	**Reaeration coefficient (K_2)**
Range	0.02 – 3.4 day^{-1}	-0.36 – 0.36 day^{-1}	0.0 – 1.0 g/m^2-day	0.0 – 100 day^{-1}
Required accuracy	0.01 day^{-1}	0.01 day^{-1}	0.1 g/m^2-day	0.1day^{-1}

N coefficients					
Calibrated Parameter	**Org. N hydrolysis rate (B_3)**	**Org. N settling rate (S_4)**	**NH_3 oxidation rate (B_1)**	**NH_3-N benthos source rate (S_3)**	**NO_2 oxidation rate (B_2)**
Range	0.02 – 0.4 day^{-1}	0.001 - 0.1 day^{-1}	0.1 – 2.0 day^{-1}	0.0 – 1.0 mg/m^2-day	0.2 – 2.0 day^{-1}
Required accuracy	0.01 day^{-1}	0.001 day^{-1}	0.01 day^{-1}	0.1 mg/m^2-day	0.01 day^{-1}

General	
Total number of state variables	54
Total string length	378 bits
Population size	1150

The original values of the coeficients to be estimated are given in Table 4.19. These quantities were used to obtain the observation data that is composed of DO, BOD and N-cycle constitiuents. The tests were performed for both perfect and biased observation data assumptions using the weighted and unweighted objective functions.

Table 4.19: Original values of the decision variables for Complex Problem II. (Values that were used for preparing the observation data and that are the solutions to the optimization problem.)

DO and BOD coefficients				
Reach Number	**K_1 (day^{-1})**	**K_3 (day^{-1})**	**K_4 (g/m^2-day)**	**K_2 (day^{-1})**
1	0.6	0.0	0.5	3.1
2	0.6	0.0	0.0	13.1
3	0.6	0.0	0.0	26.1
4	0.6	0.1	1.0	0.8
5	0.6	0.0	0.0	11.0
6	0.6	0.0	0.5	2.8

N coefficients					
Reach Number	**B_3 (day^{-1})**	**S_4 (day^{-1})**	**B_1 (day^{-1})**	**S_3 (mg/m^2-day)**	**B_2 (day^{-1})**
1	0.4	0.00	0.15	0.00	1.0
2	0.4	0.00	0.15	0.00	1.0
3	0.4	0.00	0.15	0.00	1.0
4	0.4	0.05	0.15	0.05	1.0
5	0.4	0.00	0.15	0.00	1.0
6	0.4	0.00	0.15	0.00	1.0

4.4.3.1 Perfect Data Case

The perfect observation data used in this test is given in Tables 4.20 ad 4.21 for calibration and verification conditions, respectively. The values of the weights assigned to the error functions of the simulated constituents are given in Appendix D (Table D.2). The DO, BOD, organic N, ammonia-N (NH_3-N), nitrite-N (NO_2-N), and nitrate-N (NO_3-N) profiles plotted using the optimized coefficients for both weighted and unweighted runs are given in Figures 4.23 to 4.34, for the main river and the triburatory.

Table 4.20: The perfect observation data used in Complex Problem II for the calibration conditions. (concentrations in mg/l).

Reach Number	Element Number	DO	BOD	Org.N	NH_3-N	NO_2-N	NO_3N
1	7	3.38	20.37	2.06	8.69	0.21	0.49
1	16	1.47	15.16	1.48	7.48	0.54	0.59
2	3	6.71	1.90	0.09	0.20	0.00	0.20
3	2	7.23	17.51	0.18	0.49	0.01	0.98
4	1	7.52	5.26	0.10	0.27	0.01	0.41
4	5	5.19	2.24	0.06	0.23	0.03	0.43
5	2	7.21	1.67	0.05	0.21	0.04	0.44
6	1	3.47	10.85	1.01	5.15	0.40	0.55
6	5	2.98	10.12	0.90	4.94	0.49	0.64

Table 4.21: The perfect observation data used in Complex Problem II for the verification conditions. (concentrations in mg/l).

Reach Number	Element Number	DO	BOD	Org.N	NH_3-N	NO_2-N	NO_3N
1	7	3.78	20.55	2.07	8.71	0.18	0.48
1	16	2.3	15.94	1.53	7.51	0.48	0.57
2	3	6.86	1.91	0.09	0.2	0	0.2
3	2	7.69	17.74	0.18	0.49	0.01	0.98
4	1	8.11	5.43	0.11	0.27	0.01	0.41
4	5	6.2	2.67	0.07	0.25	0.03	0.42
5	2	8.08	2.12	0.06	0.24	0.03	0.43
6	1	4.33	11.56	1.05	5.19	0.35	0.53
6	5	4.04	11.03	0.95	5.01	0.42	0.59

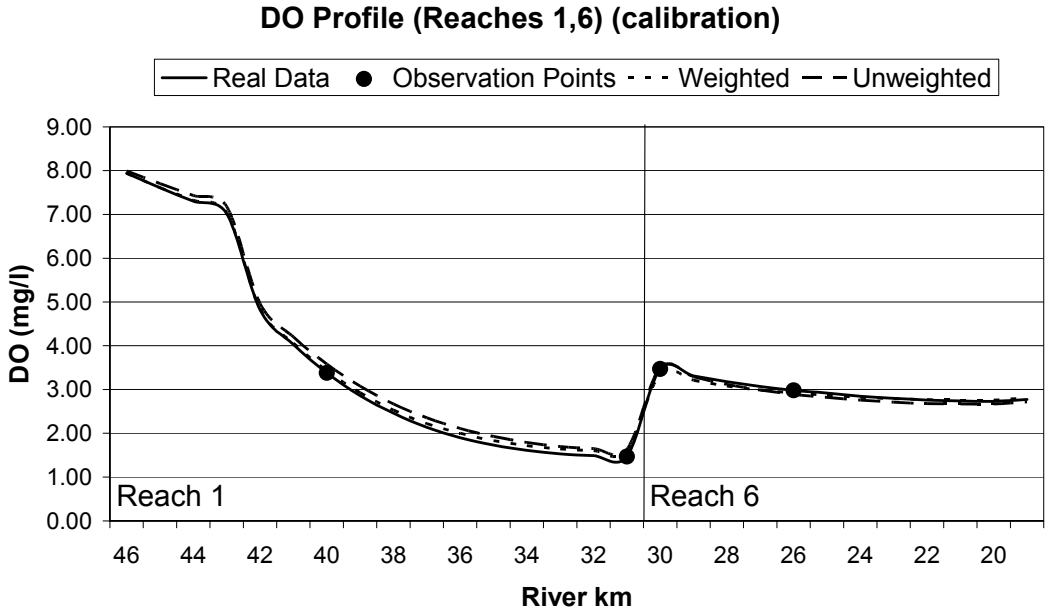

Figure 4.23: DO profile along the main river for Complex Problem II for the calibration conditions. (Perfect observation data)

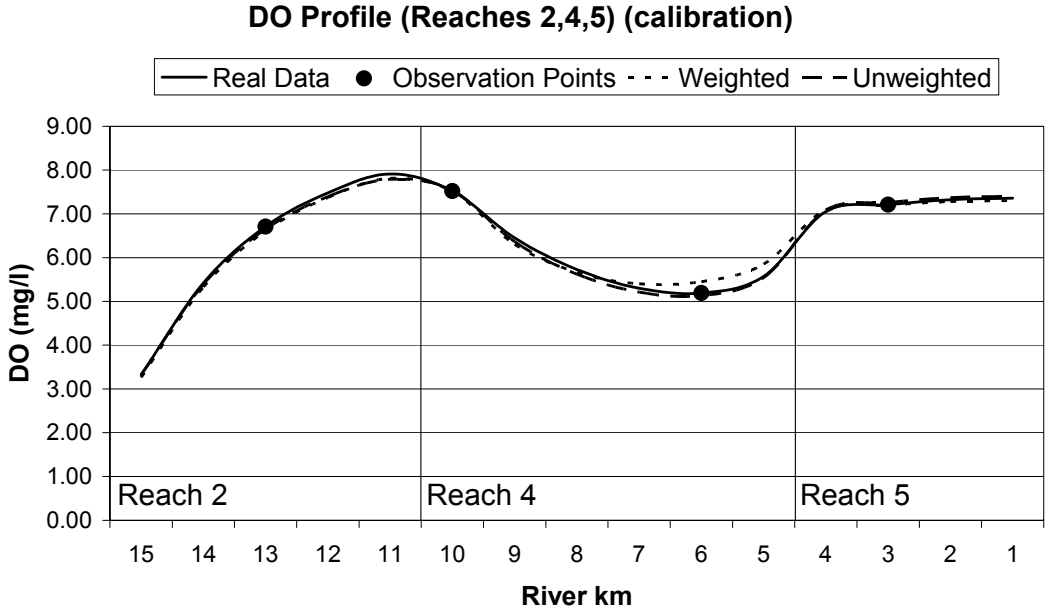

Figure 4.24: DO profile along the tributary for Complex Problem II for the calibration conditions. (Perfect observation data)

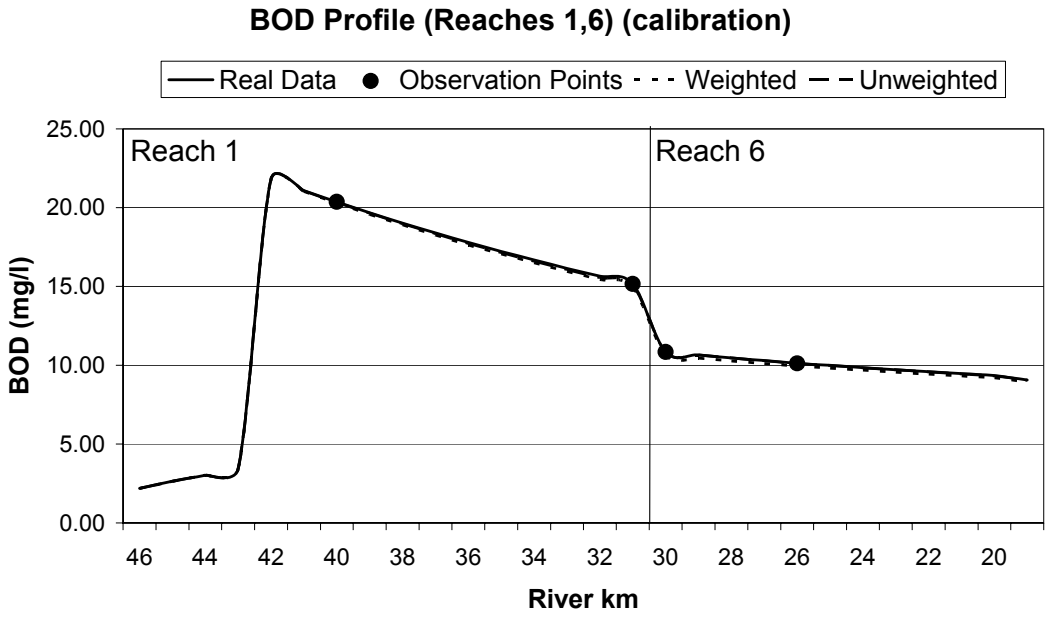

Figure 4.25: BOD profile along the main river for Complex Problem II for the calibration conditions. (Perfect observation data)

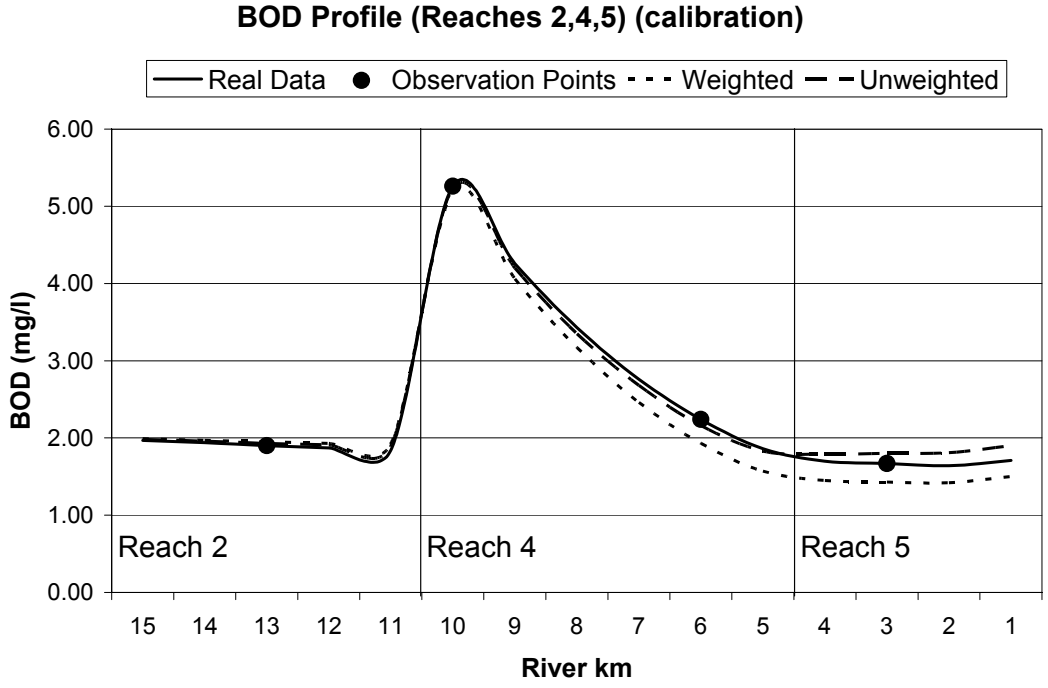

Figure 4.26: BOD profile along the tributary for Complex Problem II for the calibration conditions. (Perfect observation data)

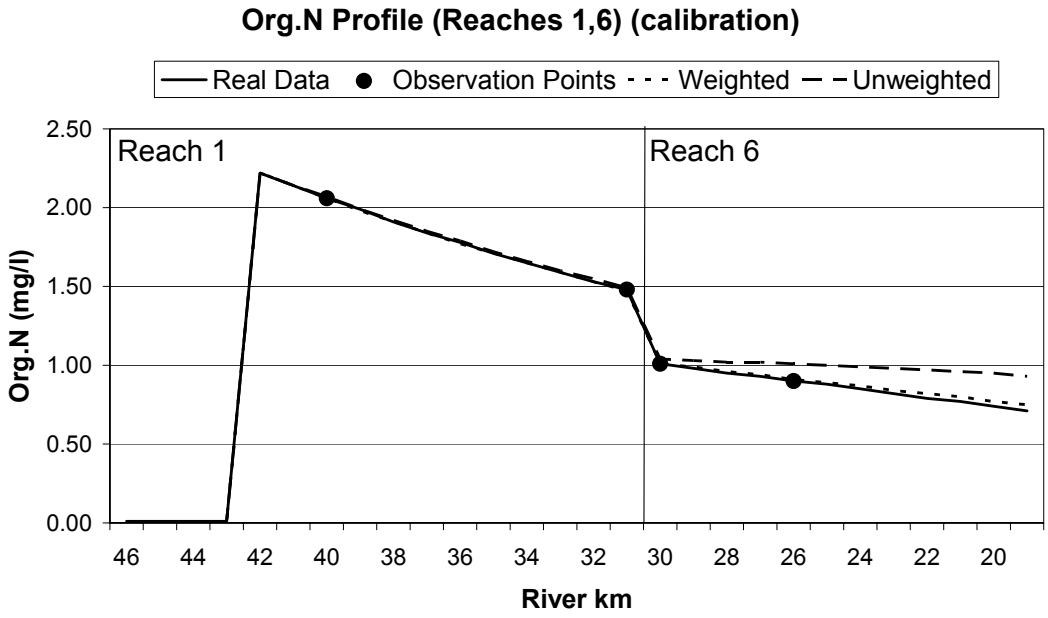

Figure 4.27: Organic N profile along the main river for Complex Problem II for the calibration conditions. (Perfect observation data)

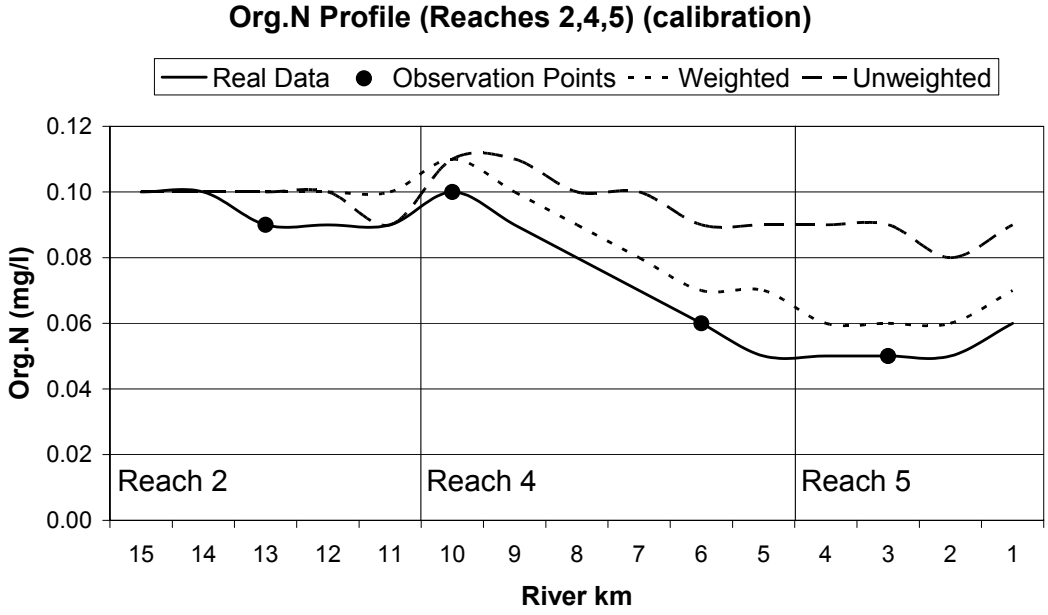

Figure 4.28: Organic N profile along the tributary for Complex Problem II for the calibration conditions. (Perfect observation data)

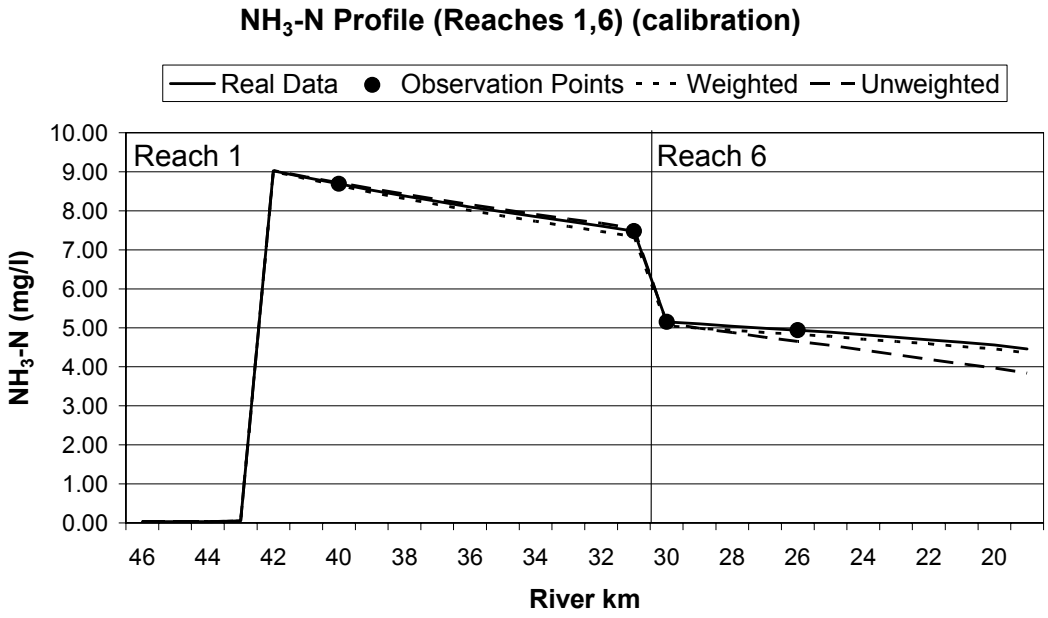

Figure 4.29: NH_3-N profile along the main river for Complex Problem II for the calibration conditions. (Perfect observation data)

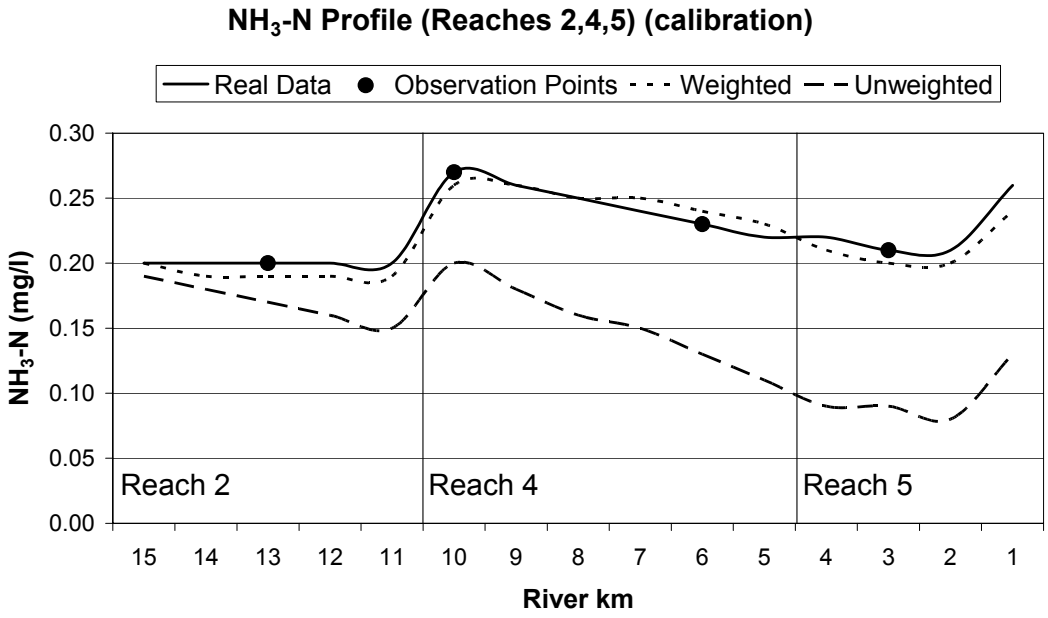

Figure 4.30: NH_3-N profile along the tributary for Complex Problem II for the calibration conditions. (Perfect observation data)

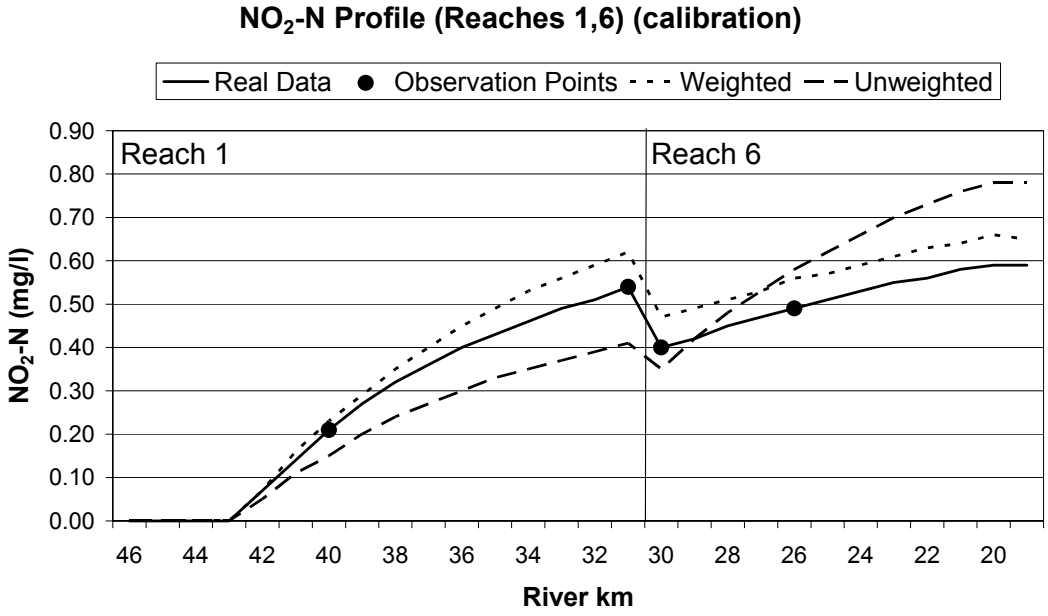

Figure 4.31: NO_2-N profile along the main river for Complex Problem II for the calibration conditions. (Perfect observation data)

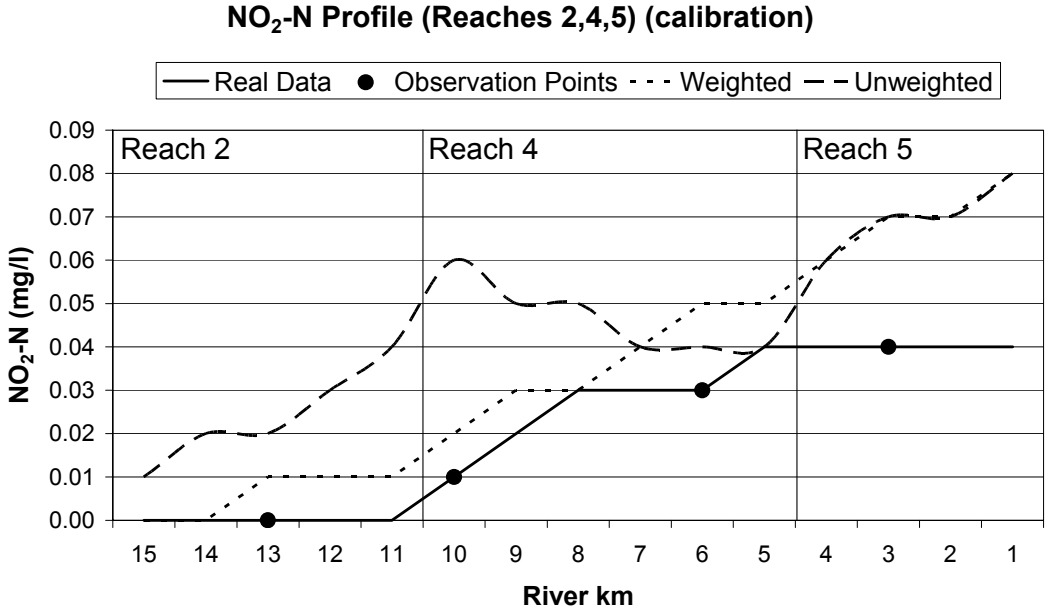

Figure 4.32: NO_2-N profile along the tributary for Complex Problem II for the calibration conditions. (Perfect observation data)

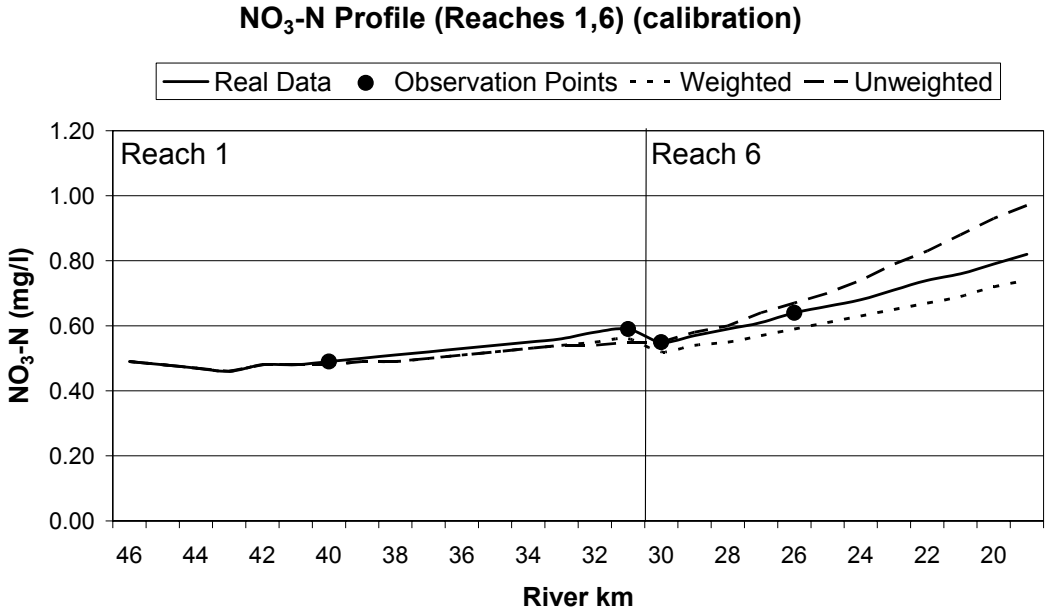

Figure 4.33: NO_3-N profile along the main river for Complex Problem II for the calibration conditions. (Perfect observation data)

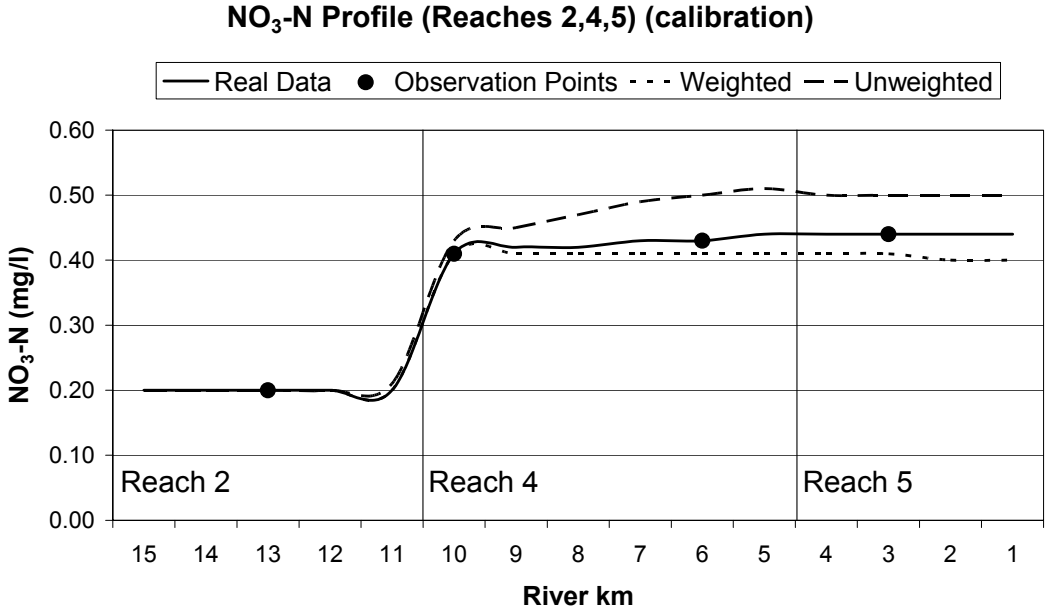

Figure 4.34: NO_3-N profile along the tributary for Complex Problem II for the calibration conditions. (Perfect observation data)

Although it was not possible to obtain the exact coefficient values, resulting water quality profiles were satisfactory except for the organic-N, and nitrite-N. However, in general, the plots show that GA optimization can give satisfactory results in complex calibration problems with a large number of decision variables. The difference between the results of the 'weighted' and 'unweighted' runs were more significant for this problem with respect to Complex Problem I. Weighted formulation gave better results especially for the members of the N-cycle, compared to the unweighted formulation. Since the possible instream concentration values of the N species was generally lower than that of BOD and DO, the resultant error values and the error function values was also relatively small. So, the weight values assigned to the error functions of the N species can be higher than that of DO and BOD. The error seen in the profiles of organic-N and nitrite-N, suggests that assigned weight values were not high enough.

4.4.3.2 Biased Data Case

The biased observation data used in this test is given in Tables 4.22 and 4.23 for calibration and verification conditions, respectively. The values of the weights assigned to the error functions of the simulated constituents are given in Appendix D (Table D.2). The DO, BOD, organic N, ammonia-N, nitrite-N, and nitrate-N profiles plotted using the optimized coefficients for both weighted and unweighted runs are given in Figures 4.35 to 4.46, for the main river and the triburatory, for the calibration conditions. The profiles for the verification conditions through the main river are given in Appendix C (Figures C.6 to C.11).

The water quality profiles plotted through the river reaches show that the resultant profiles were negatively affected from the errors in the observation data. The negative effects were more significant especially for the N-species. It should also be noted that the probable insufficient weight assignment might also have caused problems for N-species as was the case for the perfect observation data assumption. However, it should be noted that these errors are not large in absolute values but

may represent a drawback since the concentrations of these N-species are not high in natural waters.

The performance differenciation between the unweighted and weighted formulations did not follow the same trend in all profiles, such that for NO_2 and NO_3, the profile for the unweighted formulation was closer to the real profile compared to that of the weighted formulation. Simulated values of NH_3-N and NO_2-N were significantly different from the real values.

Table 4.22: The biased observation data used in Complex Problem II for the calibration conditions. (Concentrations in mg/l).

Reach Number	Element Number	DO	BOD	Org.N	NH_3-N	NO_2-N	NO_3N
1	7	3.27	17.75	2.23	8.45	0.21	0.45
1	16	1.30	18.00	1.56	7.76	0.53	0.70
2	3	5.44	1.76	0.09	0.19	0.00	0.19
3	2	6.51	19.49	0.21	0.50	0.01	0.88
4	1	8.85	6.19	0.09	0.27	0.01	0.33
4	5	4.28	2.44	0.07	0.23	0.04	0.41
5	2	7.34	1.70	0.05	0.20	0.05	0.51
6	1	3.94	8.88	0.83	5.09	0.42	0.45
6	5	3.19	11.33	0.93	4.63	0.51	0.71

Table 4.23: The biased observation data used in Complex Problem II for the verification conditions. (Concentrations in mg/l).

Reach Number	Element Number	DO	BOD	Org.N	NH_3-N	NO_2-N	NO_3N
1	7	4.30	23.97	2.44	9.00	0.15	0.48
1	16	2.65	15.54	1.38	8.15	0.46	0.67
2	3	6.61	2.04	0.09	0.21	0.00	0.17
3	2	8.87	14.27	0.18	0.43	0.01	0.81
4	1	6.81	4.40	0.11	0.22	0.01	0.36
4	5	7.37	2.18	0.06	0.21	0.03	0.43
5	2	6.58	2.37	0.05	0.27	0.03	0.50
6	1	4.82	11.54	1.21	4.62	0.31	0.53
6	5	3.85	11.54	0.82	5.94	0.42	0.53

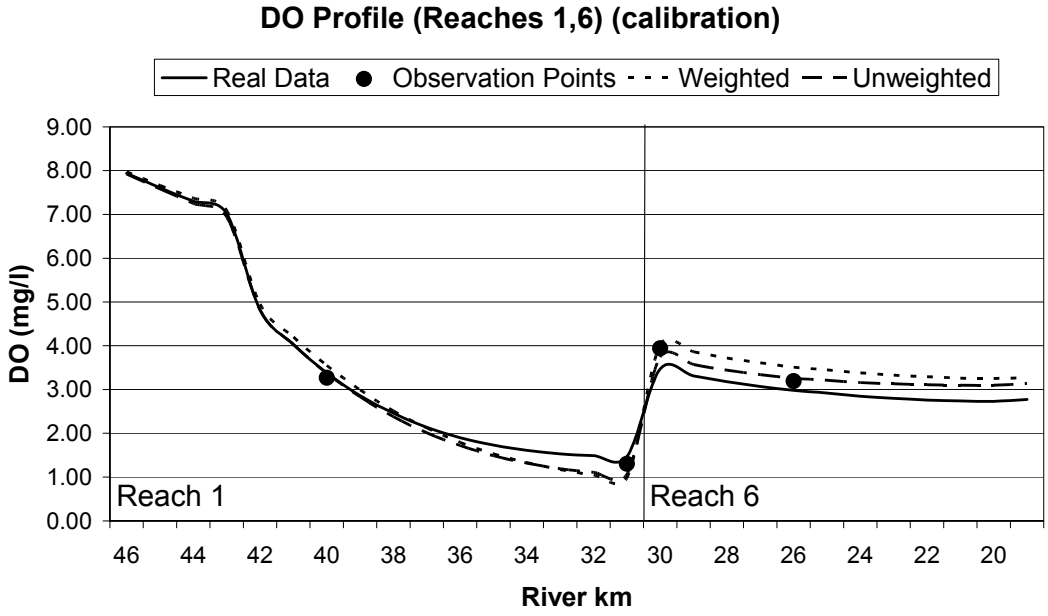

Figure 4.35: DO profile along the main river for Complex Problem II for the calibration conditions. (Biased observation data)

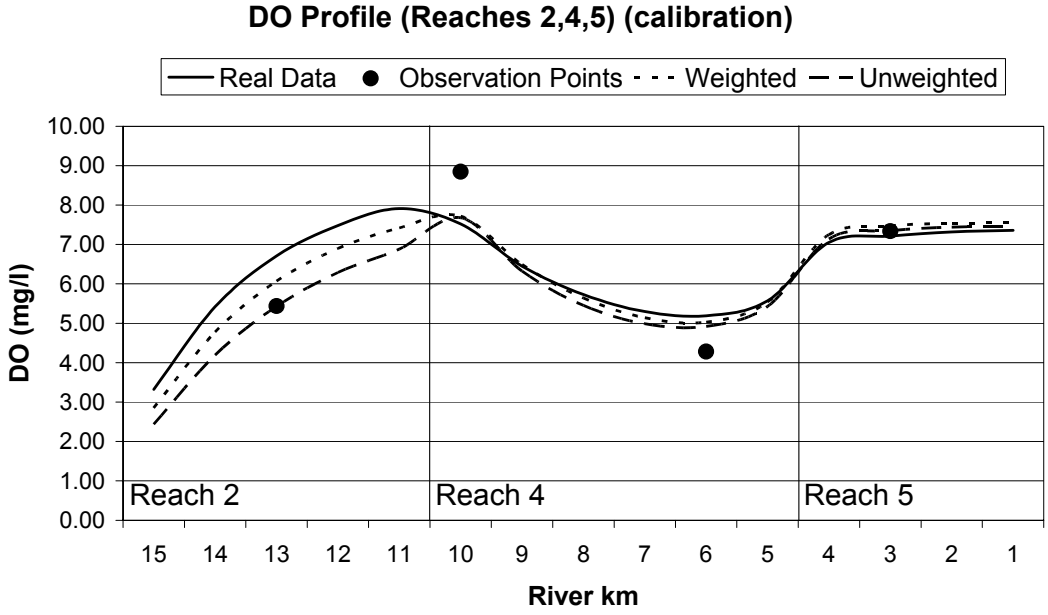

Figure 4.36: DO profile along the tributary for Complex Problem II for the calibration conditions. (Biased observation data)

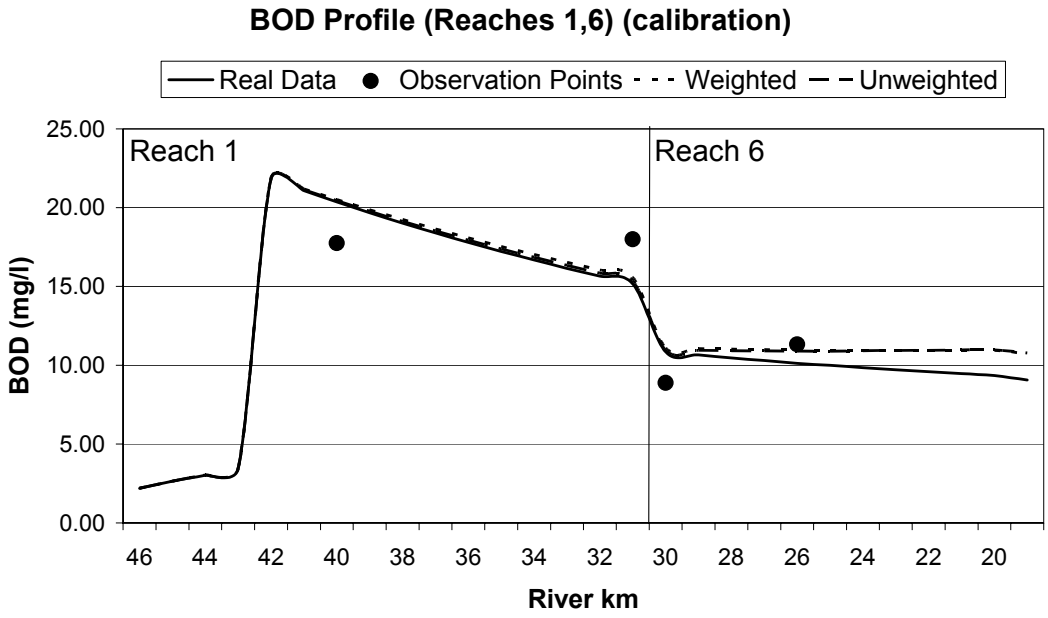

Figure 4.37: BOD profile along the main river for Complex Problem II for the calibration conditions. (Biased observation data)

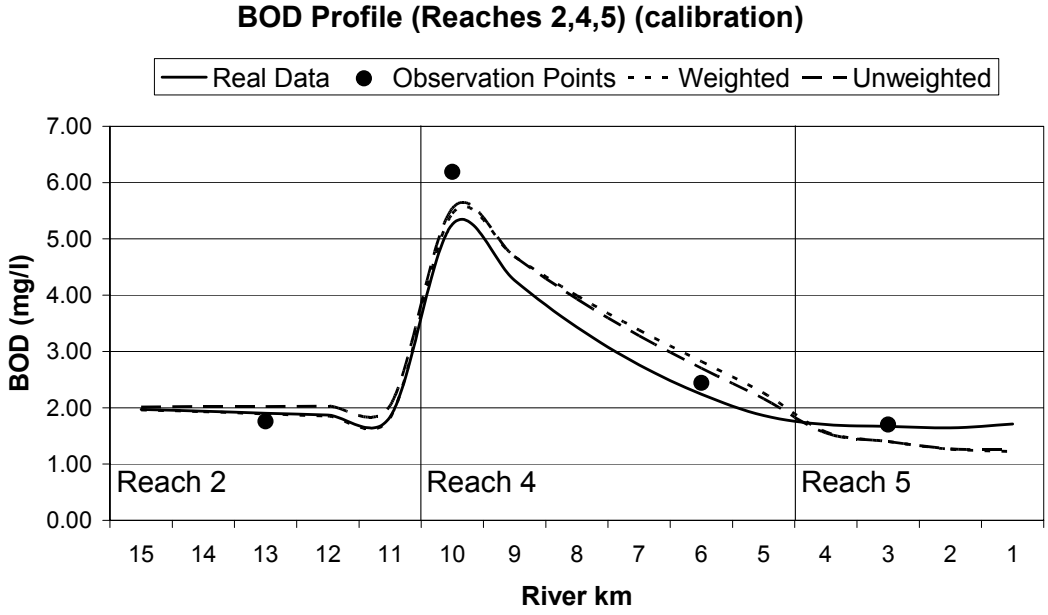

Figure 4.38: BOD profile along the tributary for Complex Problem II for the calibration conditions. (Biased observation data)

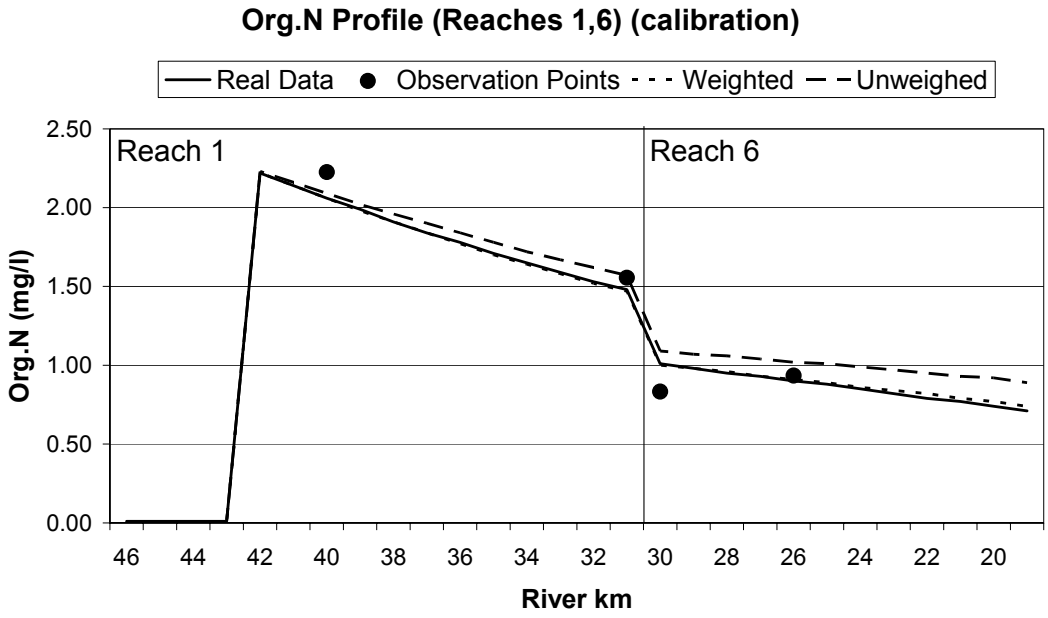

Figure 4.39: Organic N profile along the main river for Complex Problem II for the calibration conditions. (Biased observation data)

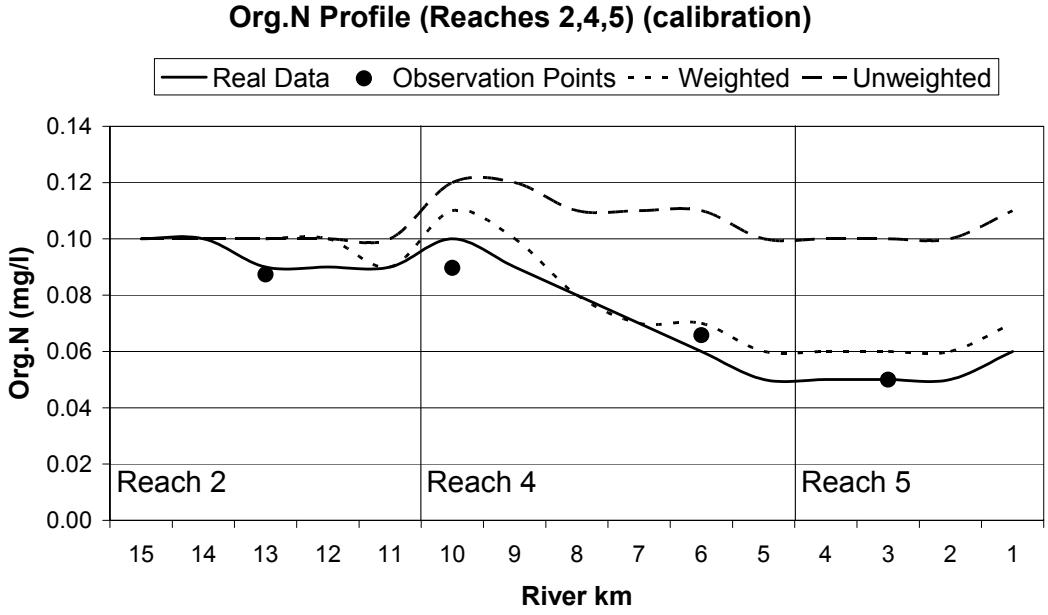

Figure 4.40: Organic N profile along the tributary for Complex Problem II for the calibration conditions. (Biased observation data)

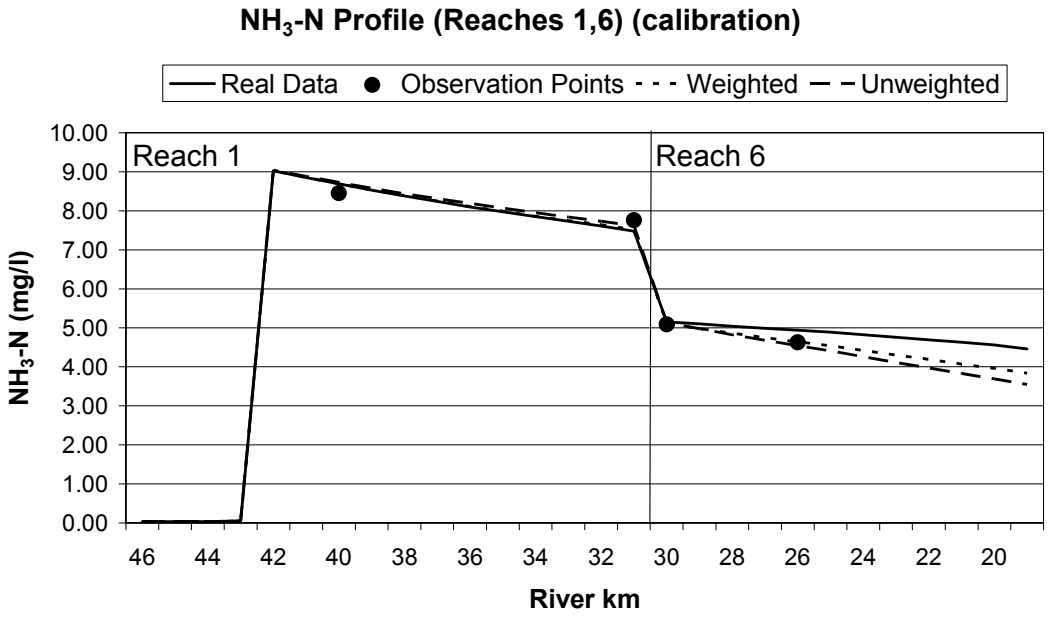

Figure 4.41: NH_3-N profile along the main river for Complex Problem II for the calibration conditions. (Biased observation data)

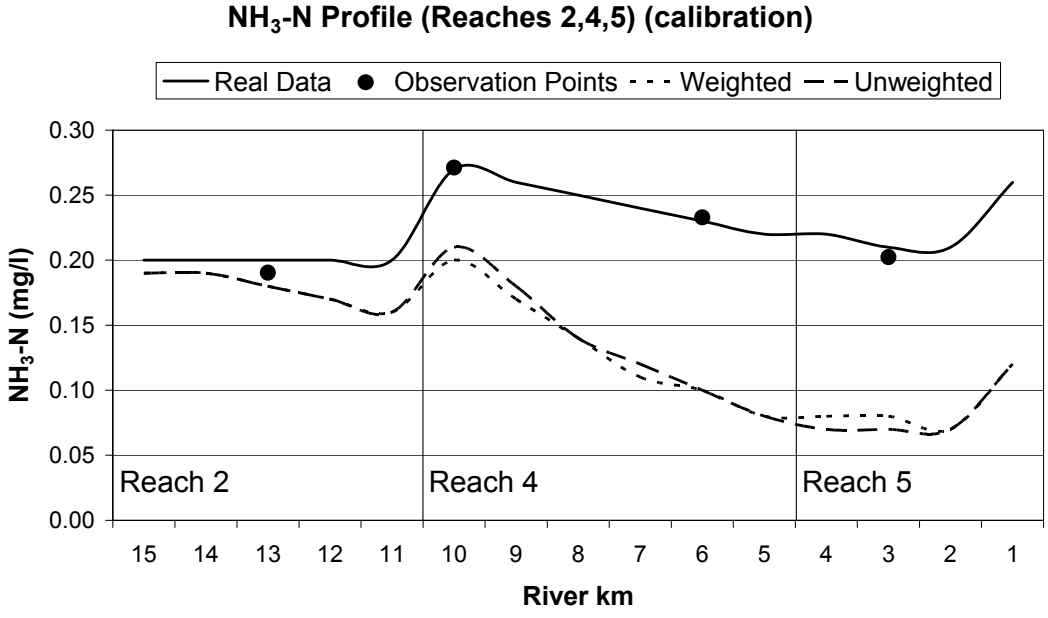

Figure 4.42: NH_3-N profile along the tributary for Complex Problem II for the calibration conditions. (Biased observation data)

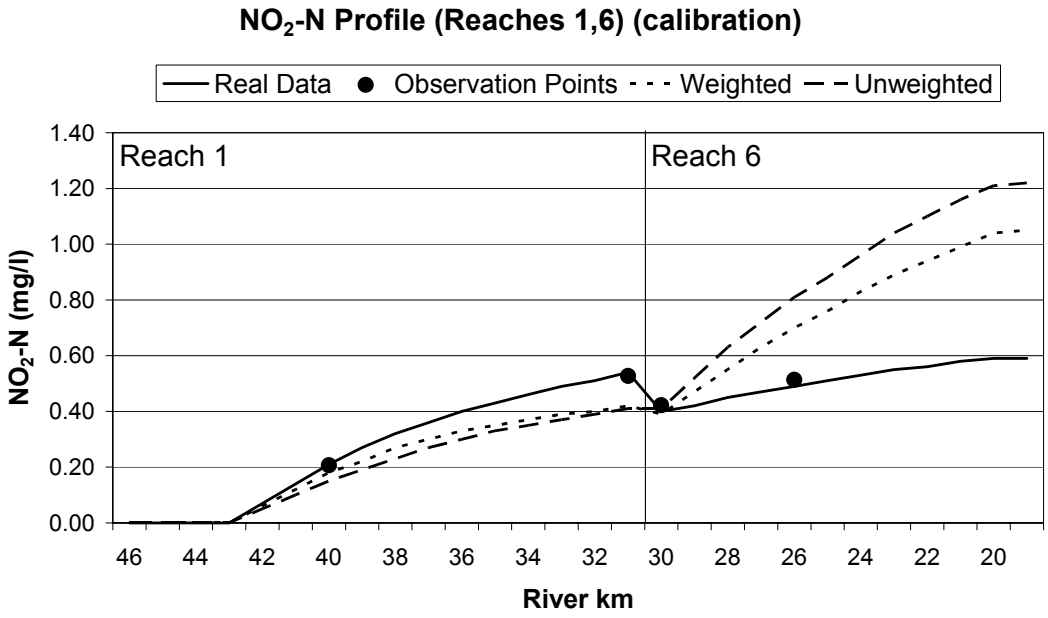

Figure 4.43: NO_2-N profile along the main river for Complex Problem II for the calibration conditions. (Biased observation data)

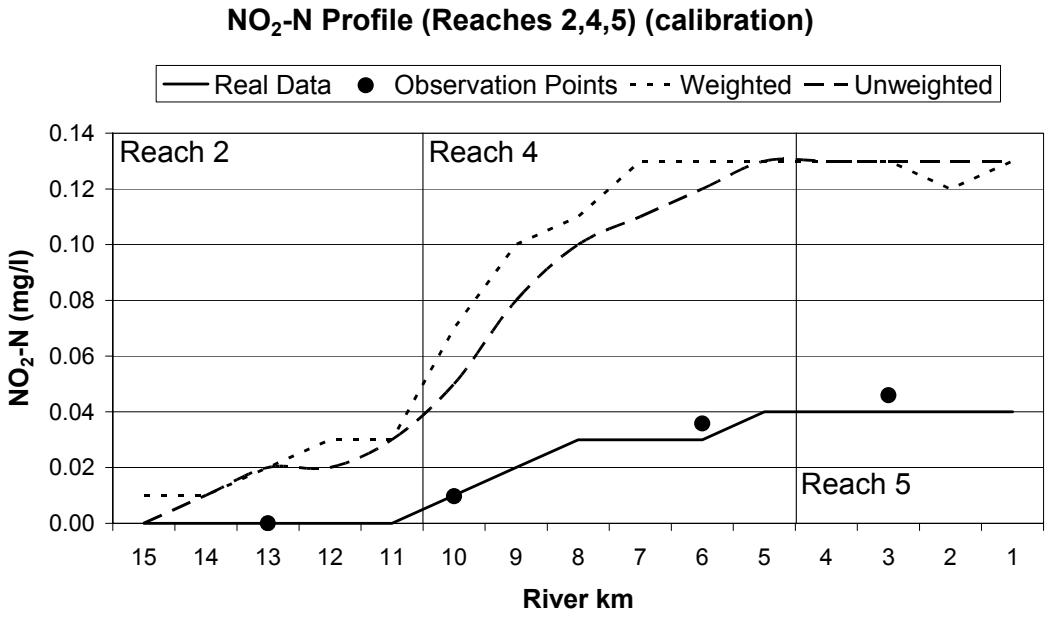

Figure 4.44: NO_2-N profile along the tributary for Complex Problem II for the calibration conditions. (Biased observation data)

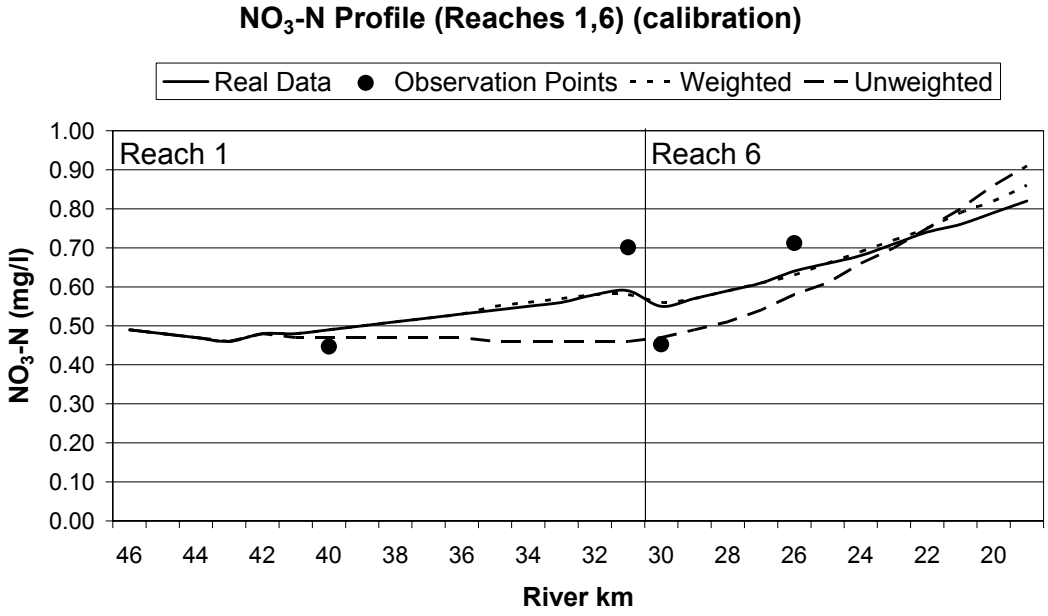

Figure 4.45: NO_3-N profile along the main river for Complex Problem II for the calibration conditions. (Biased observation data)

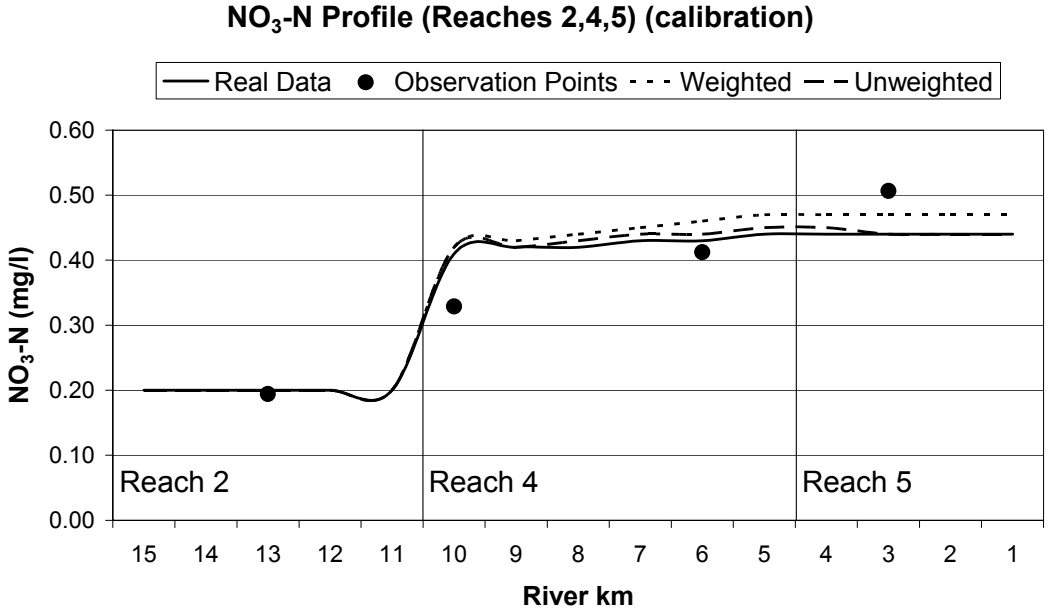

Figure 4.46: NO_3-N profile along the main river for Complex Problem II for the calibration conditions. (Biased observation data)

4.4.4 Complex Problem III

In this set of tests, the optimization runs of Complex Problem II were repeated with the addition of a second set of verification data. Therefore, two different verification sets representing different physical conditions were used. All the other assumptions were the same as for Complex Problem II. The second set of synthetic verification data was generated by increasing the headwater flow rates by 20%. The headwater flow rate values for the calibration and the second verification conditions are given in Table 4.24.

Table-4.24: Headwater flow rate values for the calibration and verification conditions.

Reach number	**Headwater flowrate for the calibration and the first verification conditions (m^3/s)**	**Headwater flowrate for the second verification condition (m^3/s)**
1	0.50	0.60
2	0.38	0.46
3	0.14	0.17

The tests for Complex Problem III were also repeated for perfect and biased observed data set conditions. The objective function values of the individual water quality variables were weighted with the previously described method. For these runs only the weighted optimization model formulation was used.

4.4.4.1 Perfect Data Case

For the calibration and the first verification data sets, the observation data presented in Tables 4.20 and 4.21 were used, respectively. The observation data used for the second verification data set is given in Table 4.25. The values of the weights assigned to the error functions of the simulated constituents are given in Appendix D (Table D.3). The water quality profiles plotted using the optimization results for

the calibration conditions are given in Figures 4.47 to 4.58 for the main river and the tributary.

The water quality profiles plotted through the river reaches according to the results of the optimization runs with two verification data sets are similar to ones in the previous tests (Complex Problem II, perfect observation data) that used a single verification data set. In general, no improvement was observed in any of the water quality profiles with the use of an additional verification data set. In fact, the deviation in the simulated NH_3 and NO_2 profiles were higher compared to the case where only a single verification data was used (Complex Problem II). However, it should be noted that these deviations are small in terms of absolute magnitude and may be hard to tune by the current algorithm used. This problem may be solved by readjusting the weights used for the N-species.

Table 4.25: The perfect observation data used in Complex Problem III for the second verification conditions. (Concentrations in mg/l)

Reach Number	Element Number	DO	BOD	Org.N	NH_3-N	NO_2-N	NO_3N
1	7	3.80	18.83	1.89	7.96	0.19	0.49
1	16	1.86	14.27	1.38	6.89	0.51	0.60
2	3	6.57	1.91	0.10	0.20	0.00	0.20
3	2	7.27	17.71	0.18	0.49	0.01	0.98
4	1	7.59	5.44	0.11	0.27	0.01	0.41
4	5	5.37	2.66	0.07	0.25	0.03	0.43
5	2	7.25	2.07	0.06	0.23	0.04	0.43
6	1	3.85	10.09	0.92	4.61	0.37	0.55
6	5	3.39	9.47	0.82	4.43	0.45	0.63

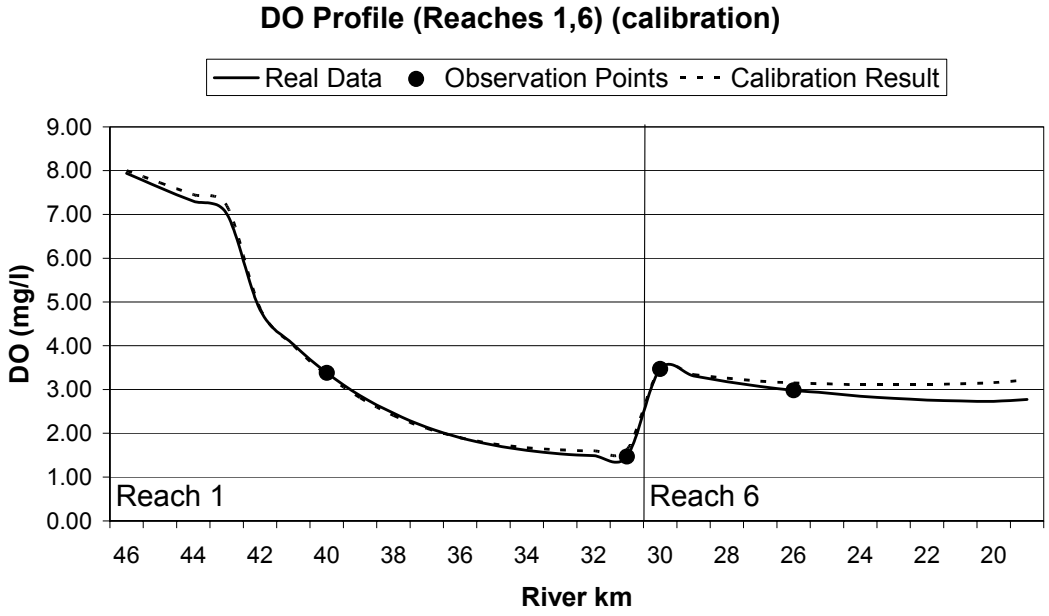

Figure 4.47: DO profile along the main river for Complex Problem III for the calibration conditions. (Perfect observation data)

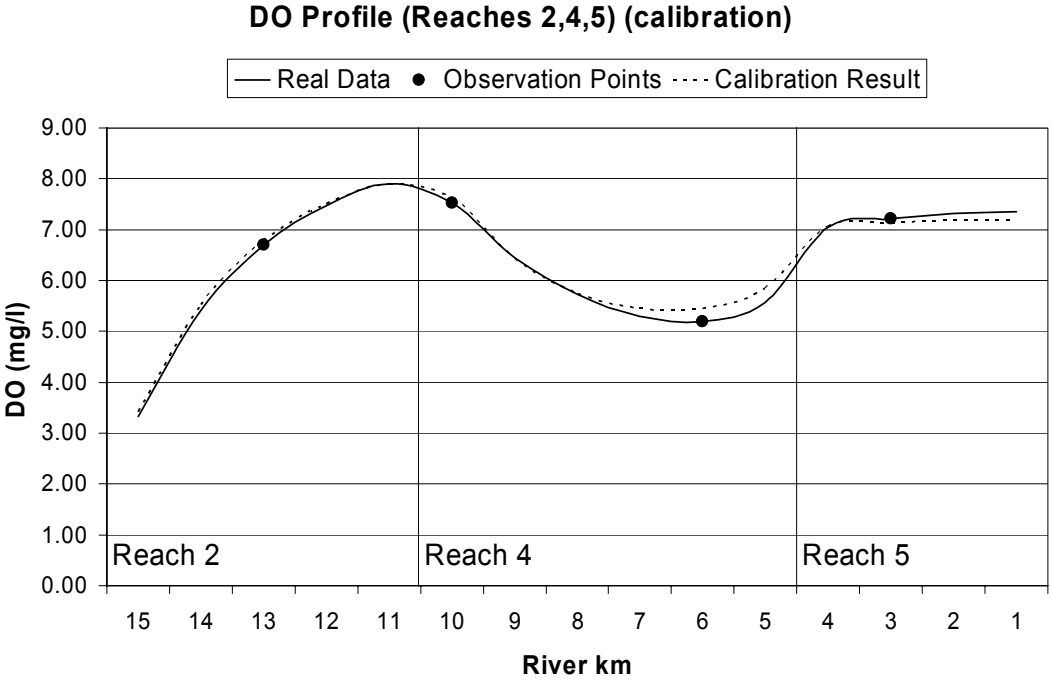

Figure 4.48: DO profile along the tributary for Complex Problem III for the calibration conditions. (Perfect observation data)

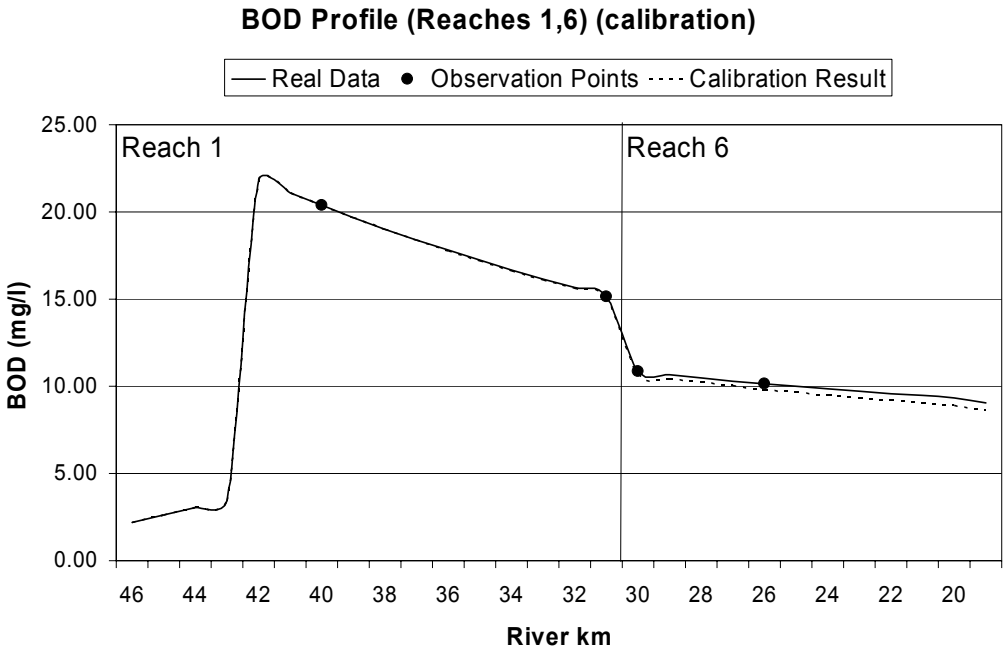

Figure 4.49: BOD profile along the main river for Complex Problem III for the calibration conditions. (Perfect observation data)

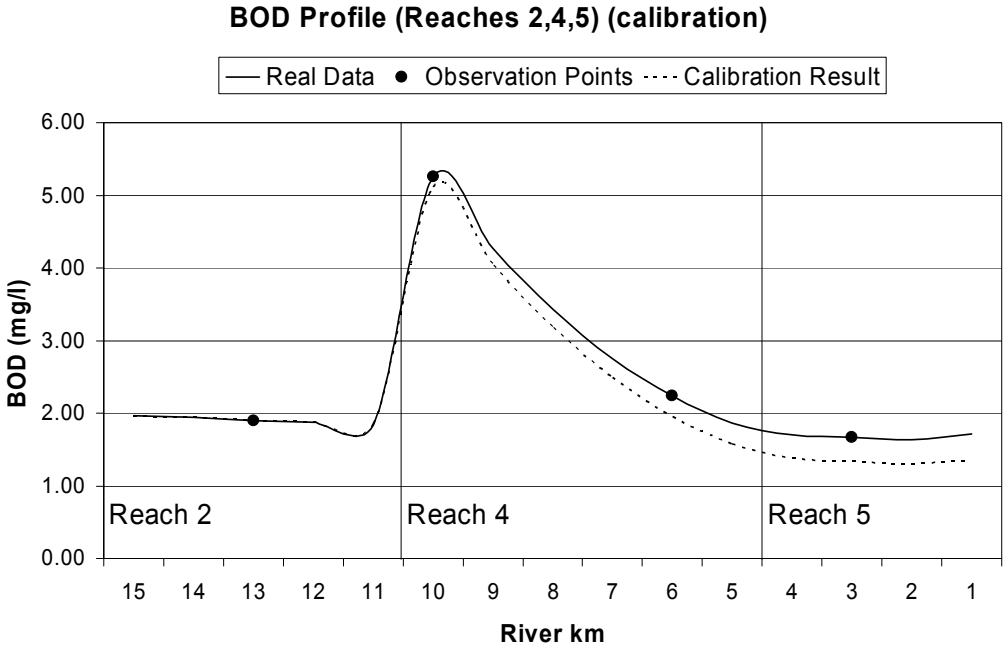

Figure 4.50: BOD profile along the tributary for Complex Problem III for the calibration conditions. (Perfect observation data)

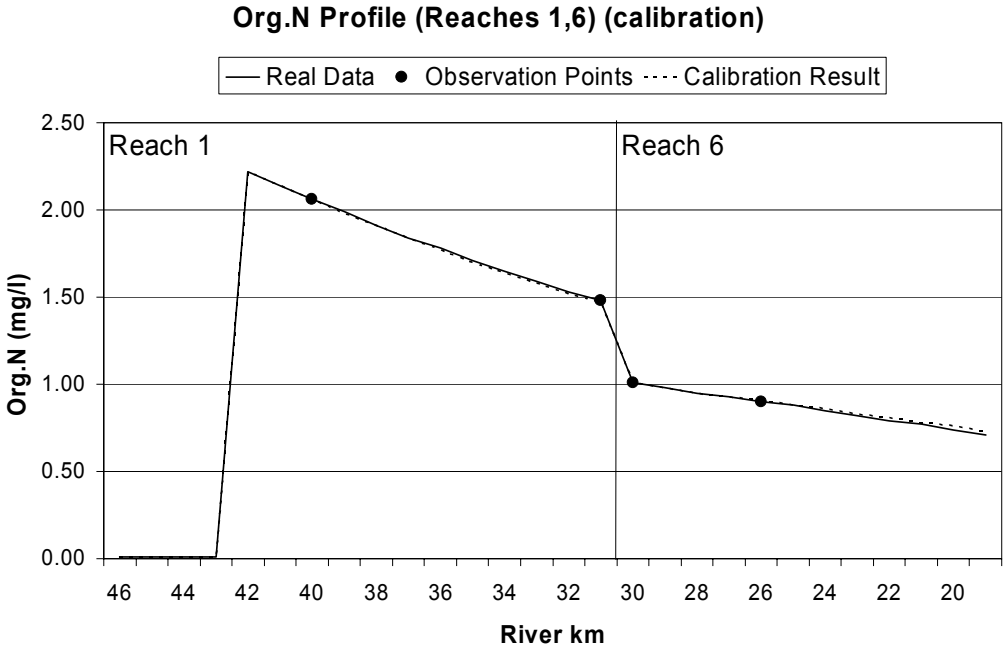

Figure 4.51: Organic N profile along the main river for Complex Problem III for the calibration conditions. (Perfect observation data)

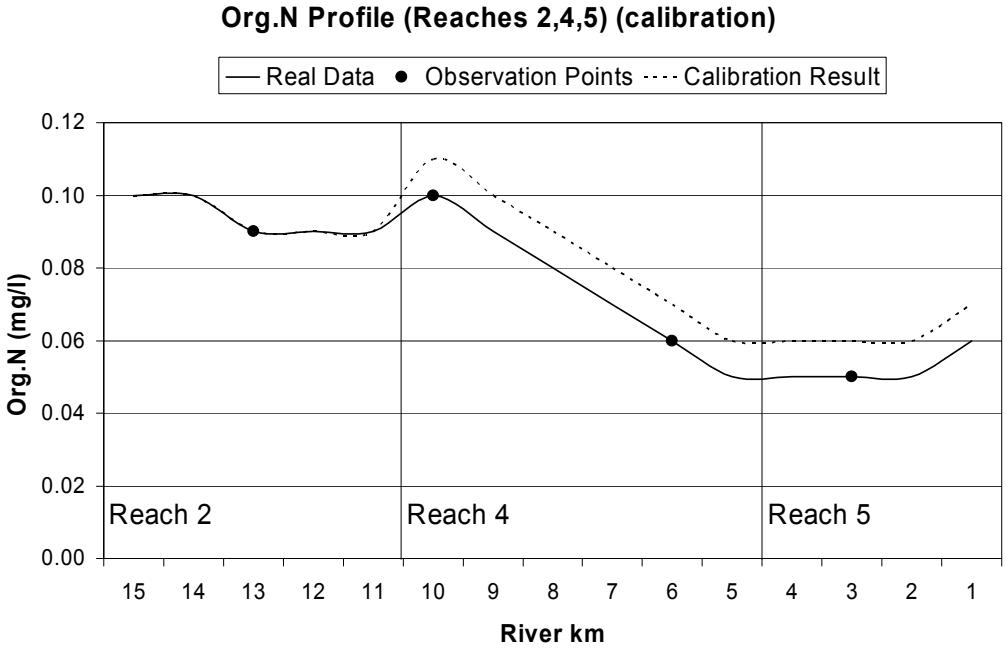

Figure 4.52: Organic N profile along the tributary for Complex Problem III for the calibration conditions. (Perfect observation data)

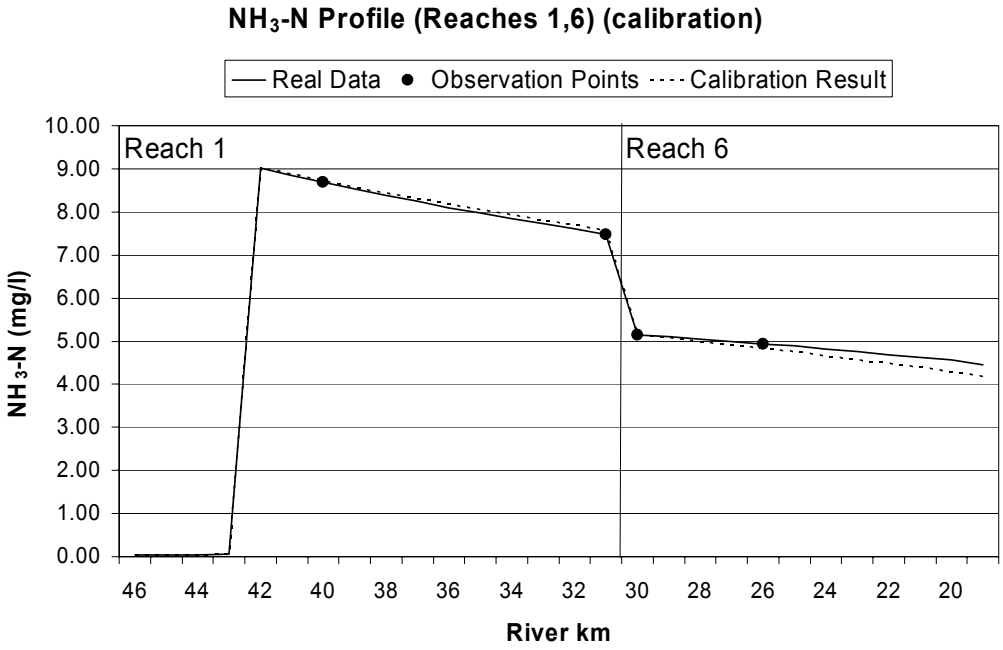

Figure 4.53: NH_3-N profile along the main river for Complex Problem III for the calibration conditions. (Perfect observation data)

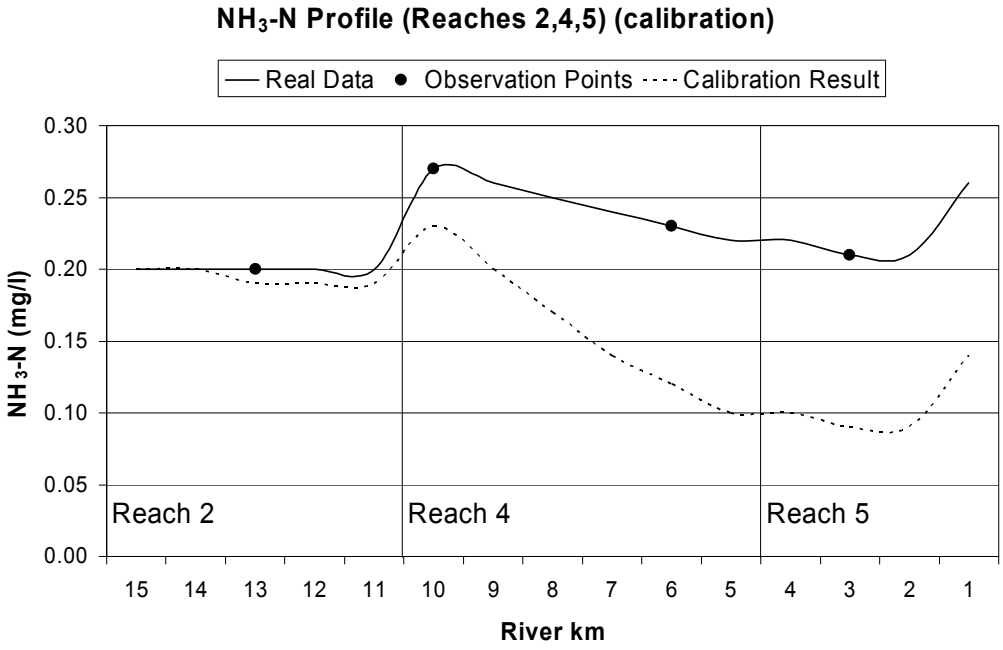

Figure 4.54: NH_3-N profile along the tributary for Complex Problem III for the calibration conditions. (Perfect observation data)

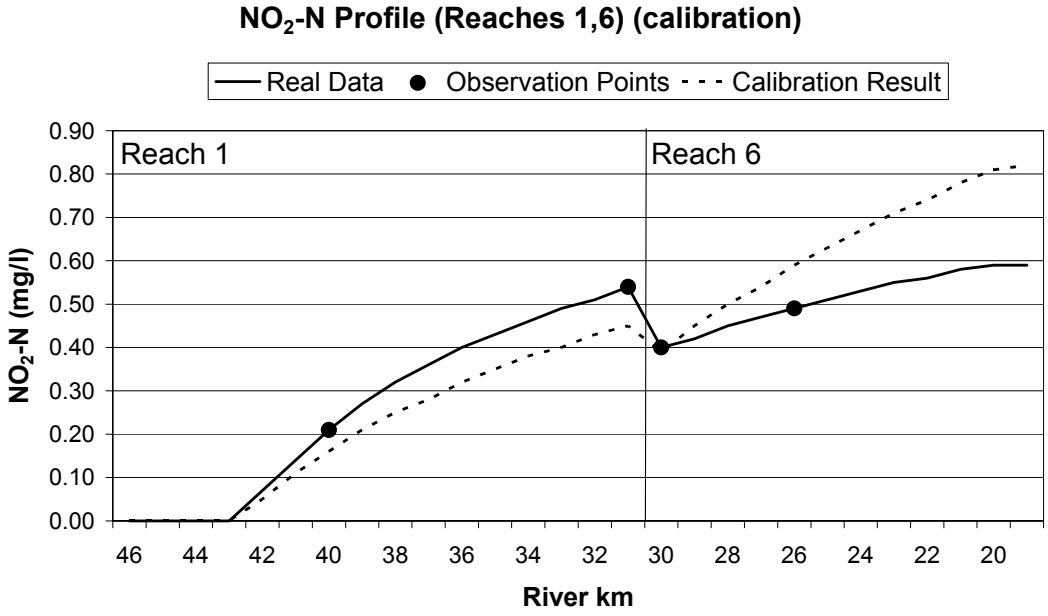

Figure 4.55: NO_2-N profile along the main river for Complex Problem III for the calibration conditions. (Perfect observation data)

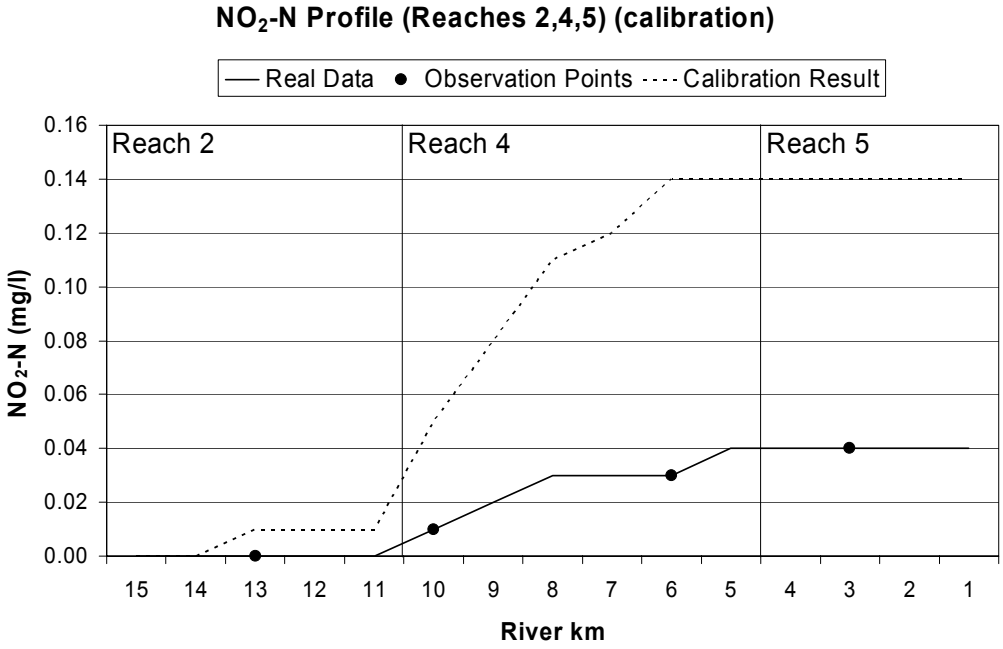

Figure 4.56: NO_2-N profile along the tributary for Complex Problem III for the calibration conditions. (Perfect observation data)

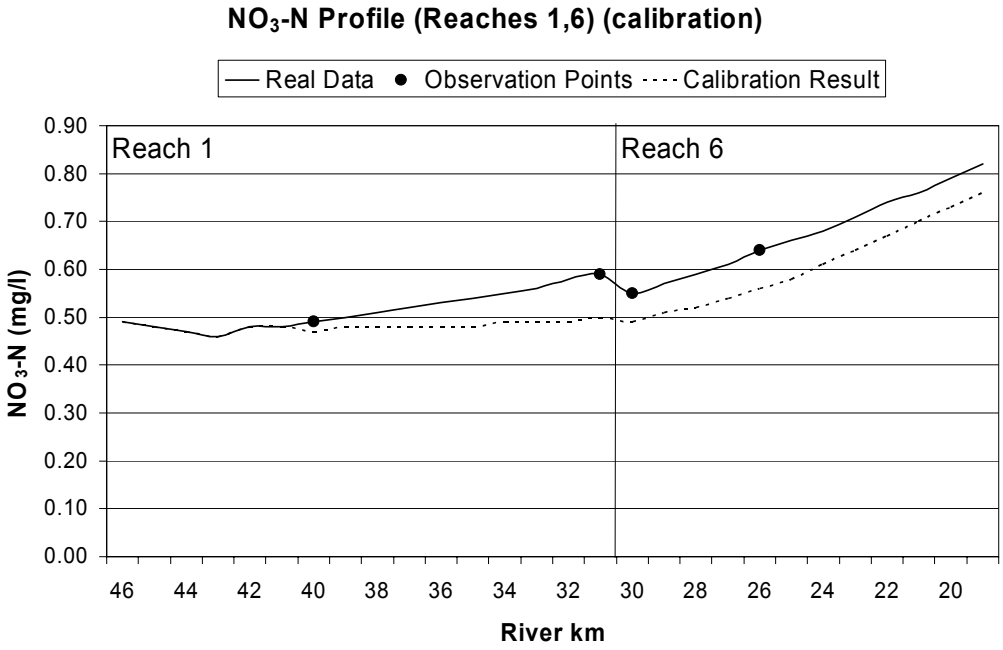

Figure 4.57: NO_3-N profile along the main river for Complex Problem III for the calibration conditions. (Perfect observation data)

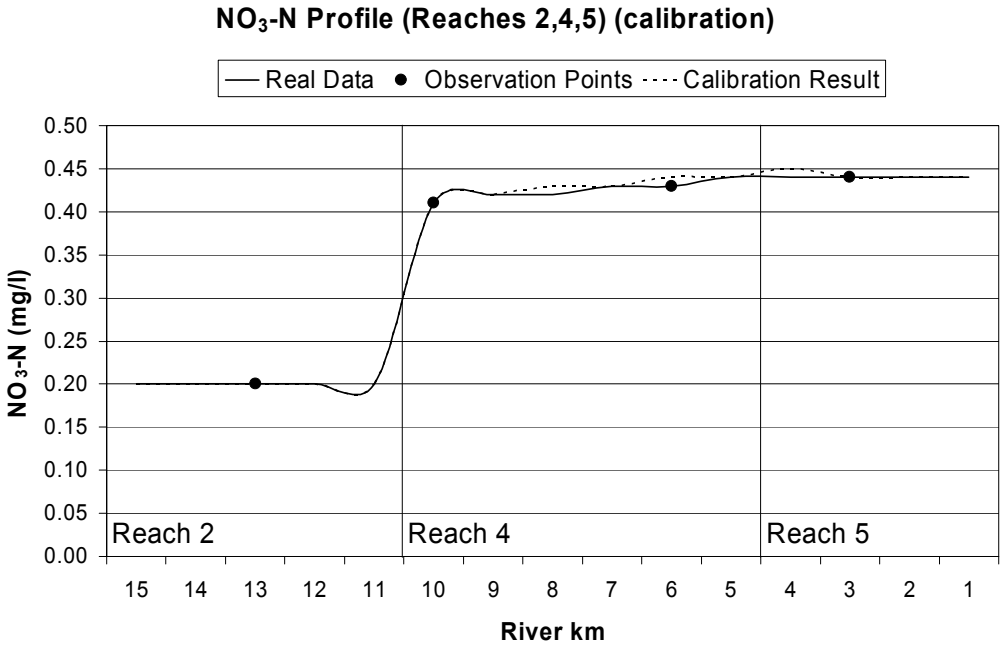

Figure 4.58: NO_3-N profile along the tributary for Complex Problem III for the calibration conditions. (Perfect observation data)

4.4.4.2 Biased Data Case

In addition to the observation data also used for the biased data of Complex Problem II, an additional verification data set presented in Table 4.26 was employed for the optimization runs. The values of the weights assigned to the error functions of the simulated constituents are given in Appendix D (Table D.3). The water quality profiles obtained from the results of the weighted optimization model formulation are depicted in Figures 4.59 to 4.70 for the main river and the tributary, for the calibration conditions. The profiles for the verification conditions through the main river are given in Appendix C (Figures C.12 to C.23). As was the case for the perfect data assumption cases, no significant improvement was observable with the addition of a second verification set.

The water quality profiles plotted according to the optimization results show the negative impact of the error in the observation data. However, the high error in the simulated values of NH_3 and NO_2 through reaches 4 and 5 (and reach 6 for NO_2) cannot be explained by the biased observed data since the quantity of resultant error is much higher than the error in the observed data. The main reason for the variations from the real profiles can be associated with the insufficient weight values used for the the N-species.

Table 4.26: The biased observation data used in Complex Problem III for the second verification conditions. (Concentrations in mg/l)

Reach Number	Element Number	DO	BOD	Org.N	NH_3-N	NO_2-N	NO_3N
1	7	4.54	20.04	2.18	8.33	0.22	0.56
1	16	1.96	16.22	1.63	6.55	0.50	0.53
2	3	6.42	2.15	0.10	0.22	0.00	0.19
3	2	7.89	20.58	0.18	0.45	0.01	1.09
4	1	7.30	5.04	0.12	0.32	0.01	0.45
4	5	5.15	2.72	0.08	0.25	0.03	0.42
5	2	6.12	2.12	0.07	0.26	0.04	0.35
6	1	3.74	10.61	0.75	3.88	0.40	0.63
6	5	3.87	7.63	0.94	3.99	0.44	0.53

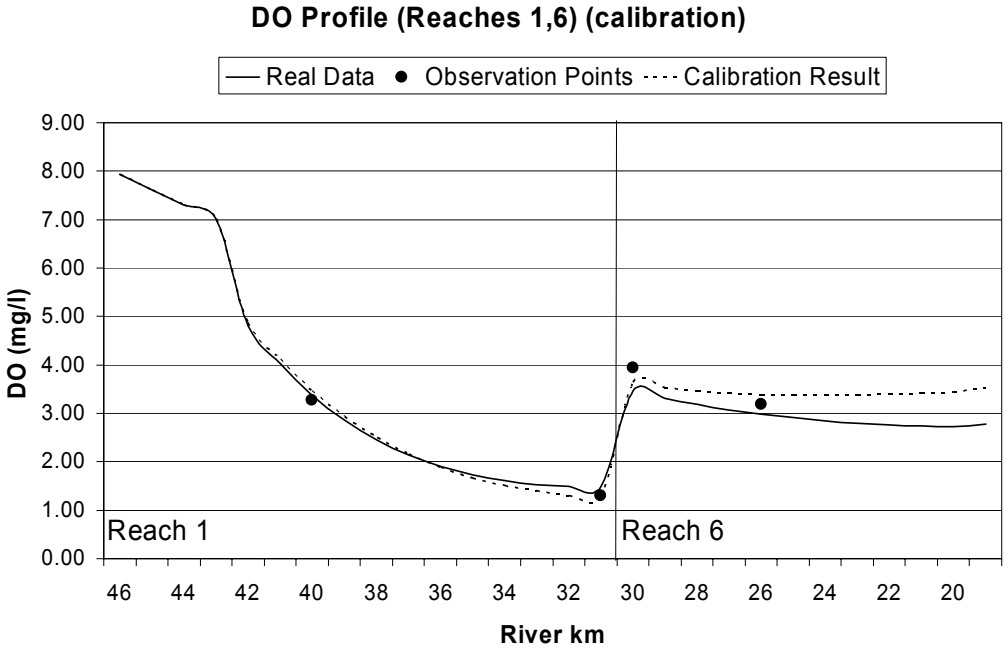

Figure 4.59: DO profile along the main river for Complex Problem III for the calibration conditions. (Biased observation data)

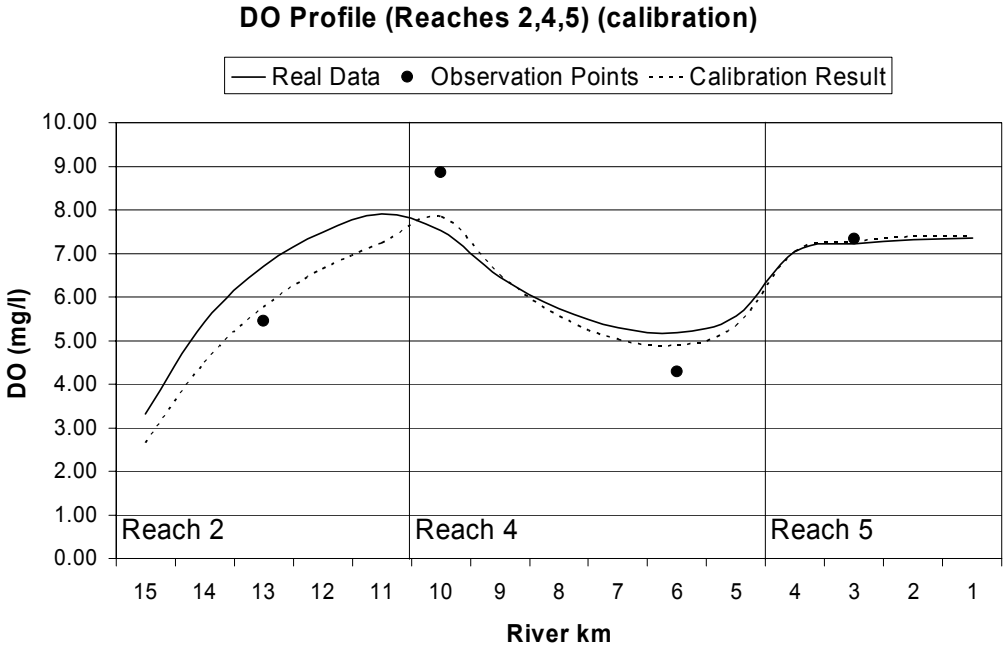

Figure 4.60: DO profile along the tributary for Complex Problem III for the calibration conditions. (Biased observation data)

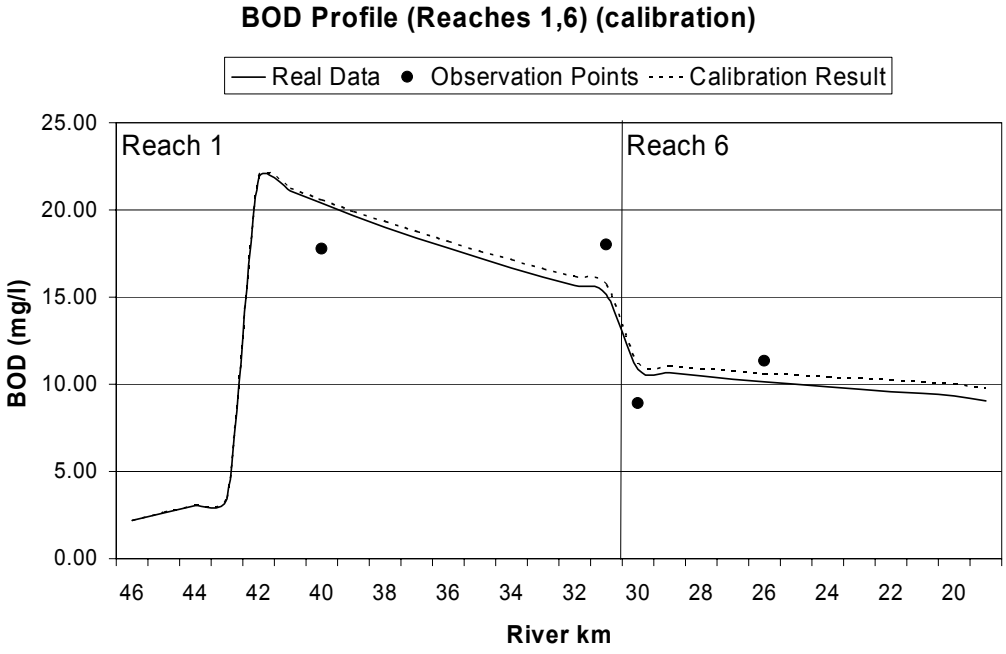

Figure 4.61: BOD profile along the main river for Complex Problem III for the calibration conditions. (Biased observation data)

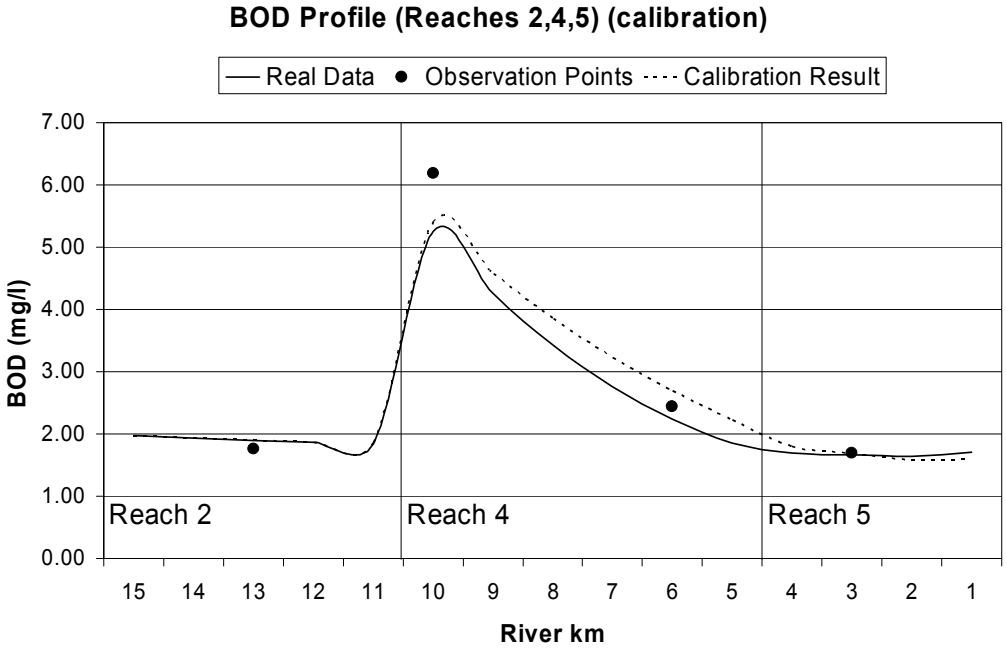

Figure 4.62: BOD profile along the tributary for Complex Problem III for the calibration conditions. (Biased observation data)

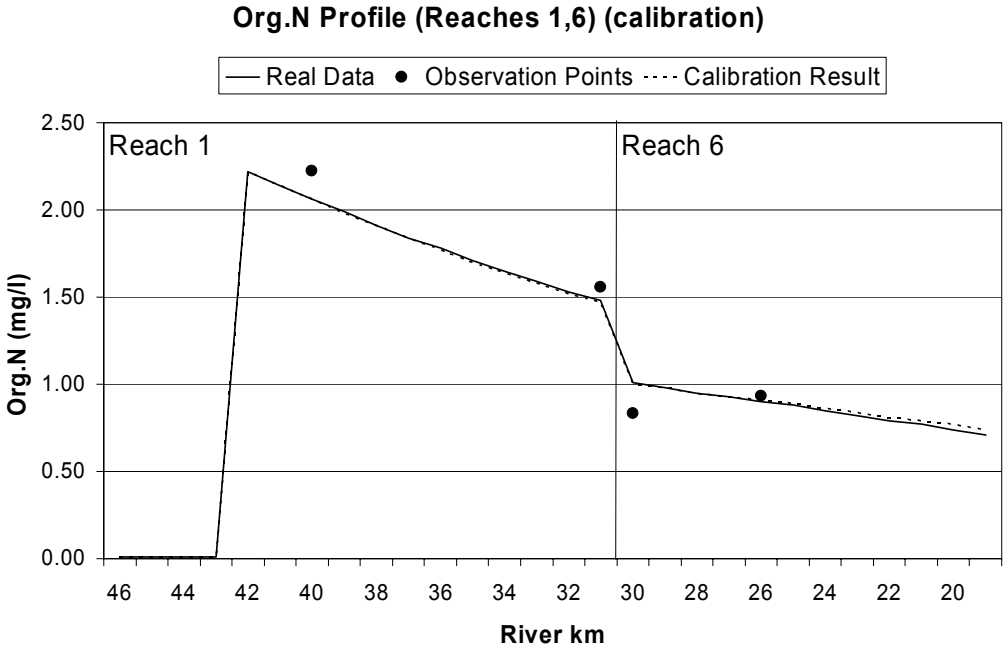

Figure 4.63: Organic N profile along the main river for Complex Problem III for the calibration conditions. (Biased observation data)

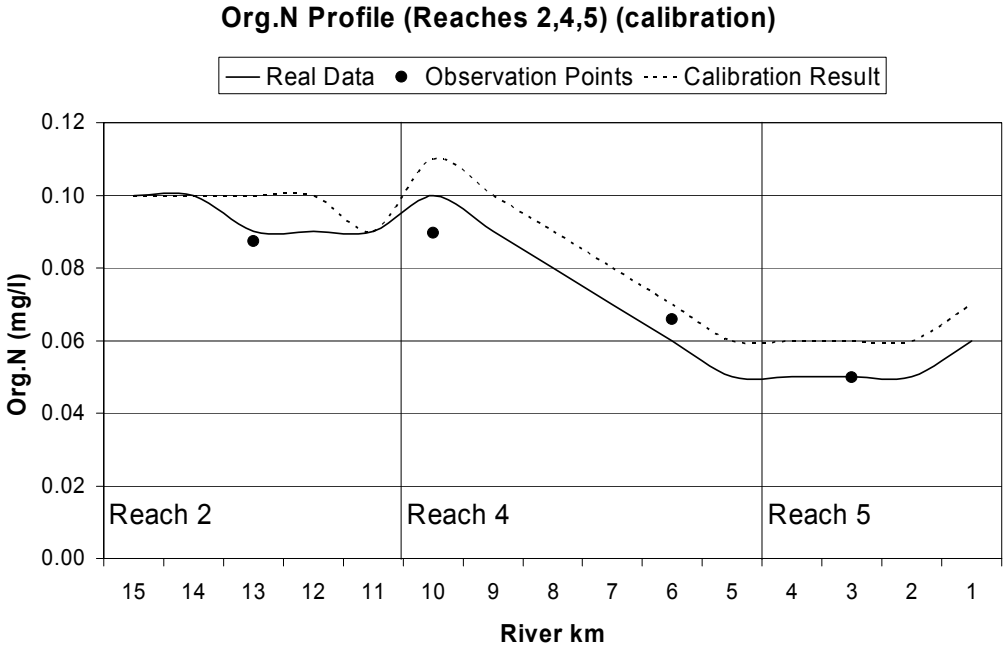

Figure 4.64: Organic N profile along the tributary for Complex Problem III for the calibration conditions. (Biased observation data)

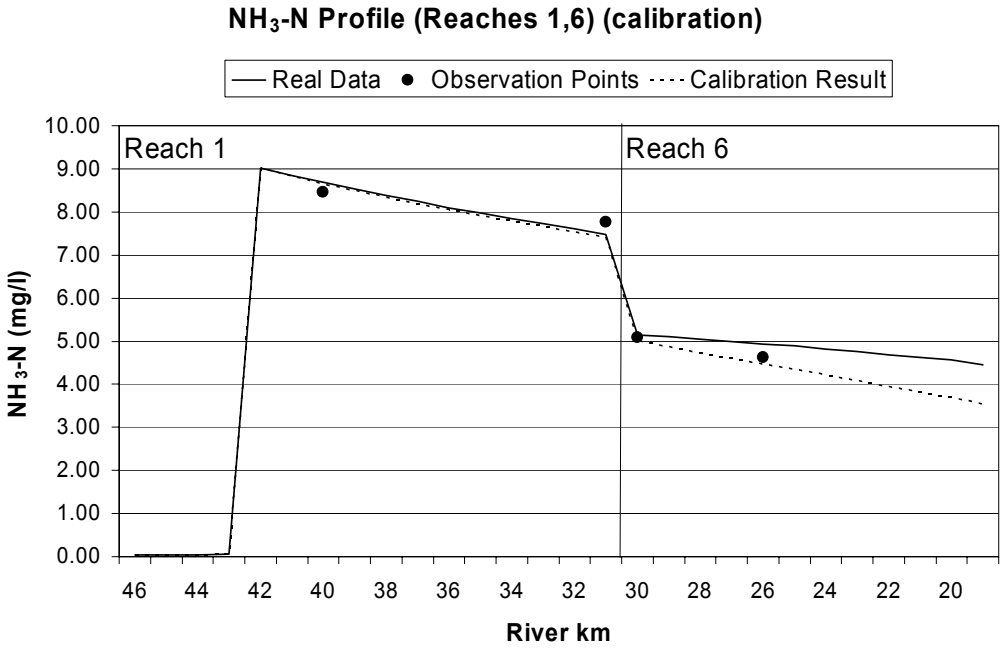

Figure 4.65: NH_3-N profile along the main river for Complex Problem III for the calibration conditions. (Biased observation data)

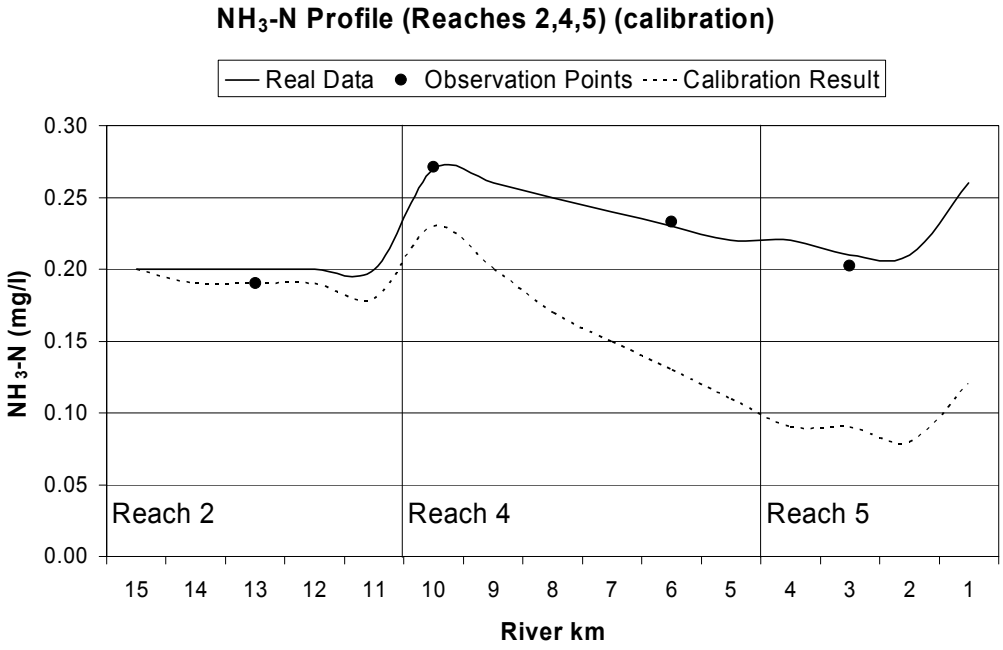

Figure 4.66: NH_3-N profile along the tributary for Complex Problem III for the calibration conditions. (Biased observation data)

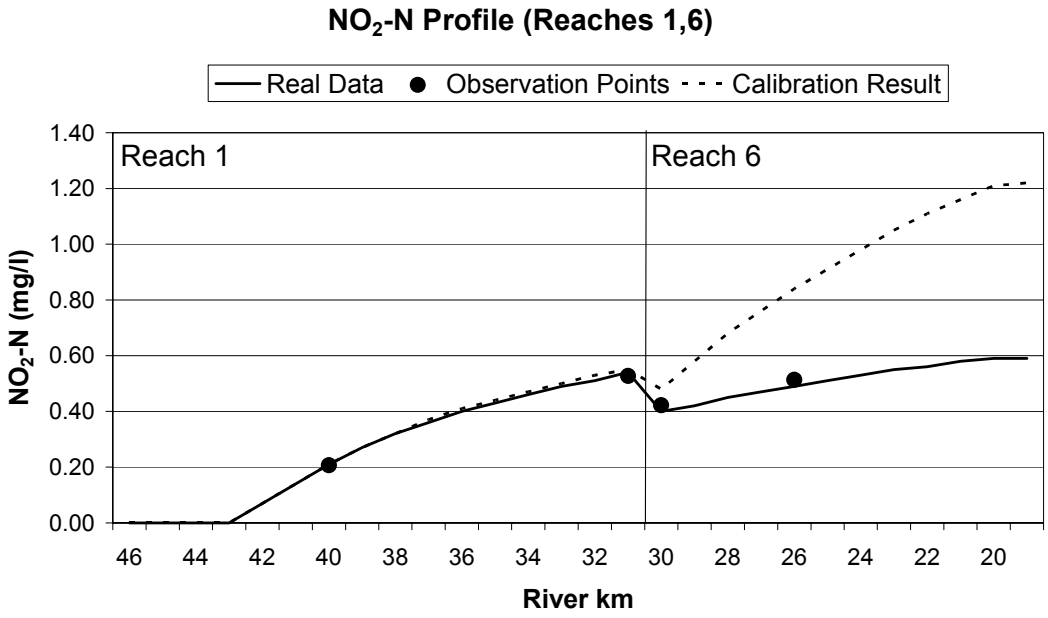

Figure 4.67: NO_2-N profile along the main river for Complex Problem III for the calibration conditions. (Biased observation data)

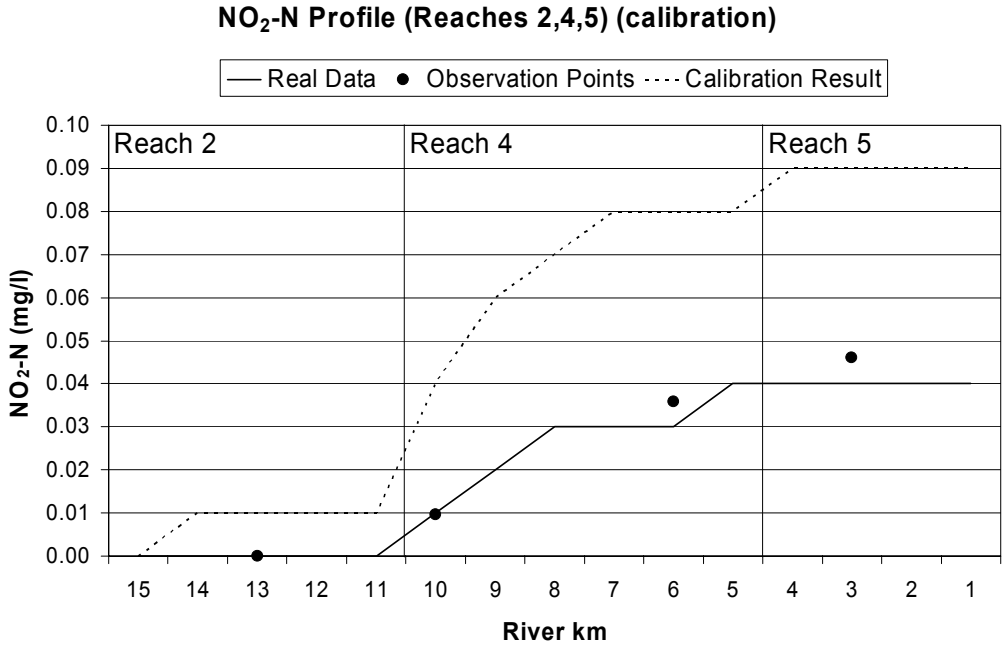

Figure 4.68: NO_2-N profile along the tributary for Complex Problem III for the calibration conditions. (Biased observation data)

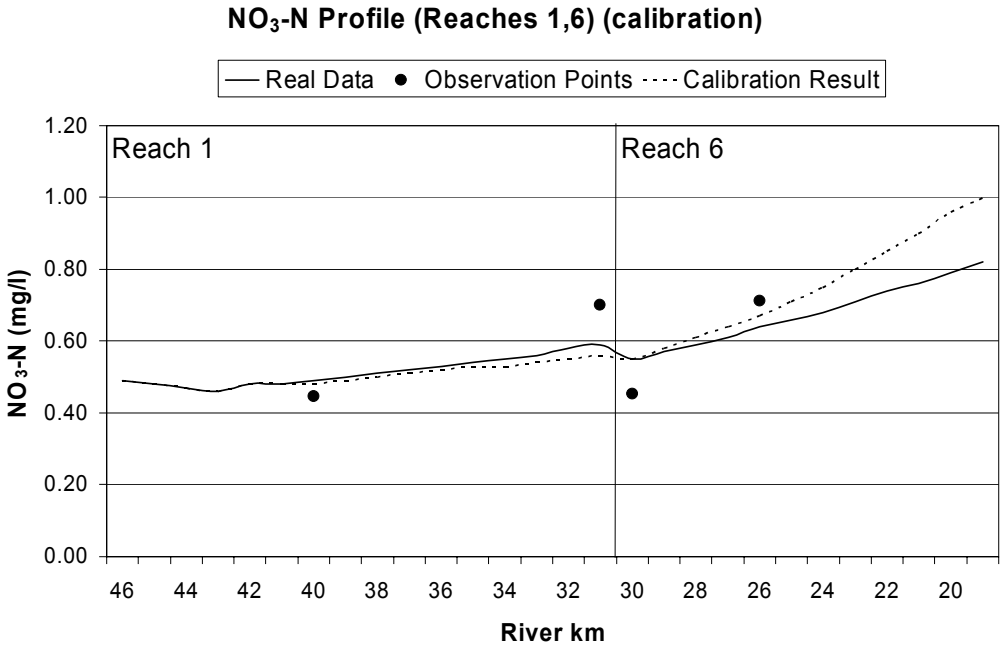

Figure 4.69: NO_3-N profile along the main river for Complex Problem III for the calibration conditions. (Biased observation data)

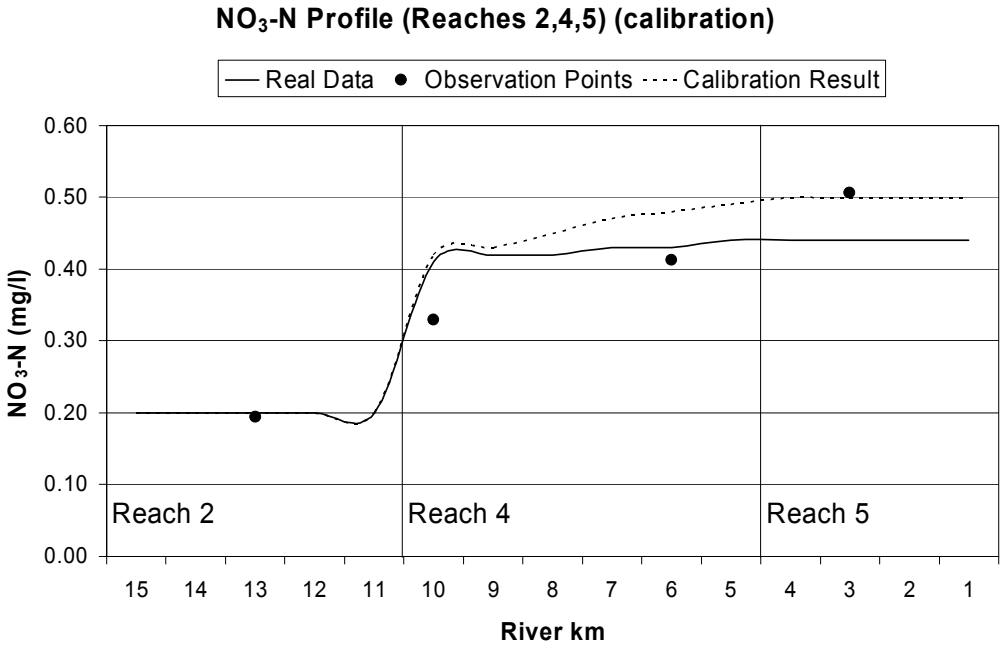

Figure 4.70: NO_3-N profile along the tributary for Complex Problem III for the calibration conditions. (Biased observation data)

CHAPTER 5

CASE STUDY

5.1 General Information

The Lower Seyhan River basin is a part of the Seyhan River basin which is situated in the southern part of Turkey. The Lower Seyhan River starts from the downstream of the Seyhan Dam, runs through the Çukurova plain and ends at the Mediterranean Sea. The catchment area of the Lower Seyhan River is only about 10% of the entire Seyhan River basin which has an area of about 20,731 km^2 (Onur et al., 1999). However, the lower section of the river is much more critical from the water quality management perspective. The Lower Seyhan River is subjected to domestic and industrial pollution loads from the city of Adana located downstream of the Seyhan Dam. Moreover, a significant amount of agricultural pollution load originating from the Çukurova plain is carried by the return waters which have been withdrawn from the river itself, as well.

The M.S. Thesis by Onur (1996) includes a comprehensive modeling study of the Lower Seyhan River. Onur combined the required data (hydrological, meteorological, hydraulic, and water quality data) from related sources and developed a modeling approach to simulate the water quality of the river by using the QUAL2E model. In her study, the calibration of the kinetic coefficients of the developed model was performed on a trial-and-error basis by visual comparison of the observed data and the simulated water quality profiles.

In this section of the study, the developed genetic algorithm (GA) optimization model is used as a tool for calibration of the QUAL2E model for the Lower Seyhan River. The QUAL2E input file was generated by extracting the information in

Onur's work. For the lacking input data, QUAL2E documentation (Brown and Barnwell, 1987) was referred to and the default or the recommended typical values were used. When both sources were inadequate in providing an exact value, personal judgment was incorporated. The latter method was used for only a few number of input parameters.

5.2 The Lower Seyhan River and the Modeling Approach

The Lower Seyhan River runs for 84 km between the Seyhan Dam and the Mediterranean Sea. In the modeling study, the river was divided into 6 consecutive reaches based on the slopes, cross-sectional areas and the hydraulic properties along the river (Onur et al., 1997). The schematic representation of the Lower Seyhan River as divided into reaches is given in Figure 5.1.

There are four major point discharges to the Lower Seyhan River. Sarıcam Creek, two industrial point sources (referred to as A and B in Figure 5.1), and a drainage canal (referred to as TDO in Figure 5.1). Sarıcam Creek confluences to the main river at 4 km downstream from the Seyhan I Regulator which is located just below the Seyhan Dam. The creek carries the loads of several small industries and residential areas and represented as a point source in the river system. Two point sources, A and B, carrying mainly industrial wastewaters, discharge to the river at the 6^{th} and 7^{th} km, respectively. TDO carries the agricultural return waters from irrigation and discharges into the river at the 79^{th} km from the dam structure. The load arising from agricultural activities was simulated separately and the output was considered as a point load to the main river. The dissolved oxygen (DO) improvement due to the 4.5 m waterfall at the Seyhan II Regulator and Hydroelectric Power Plant located just upstream of inputs A and B was also incorporated into the model calculations (Onur, 1996; Onur et al., 1997; Onur et al., 1999).

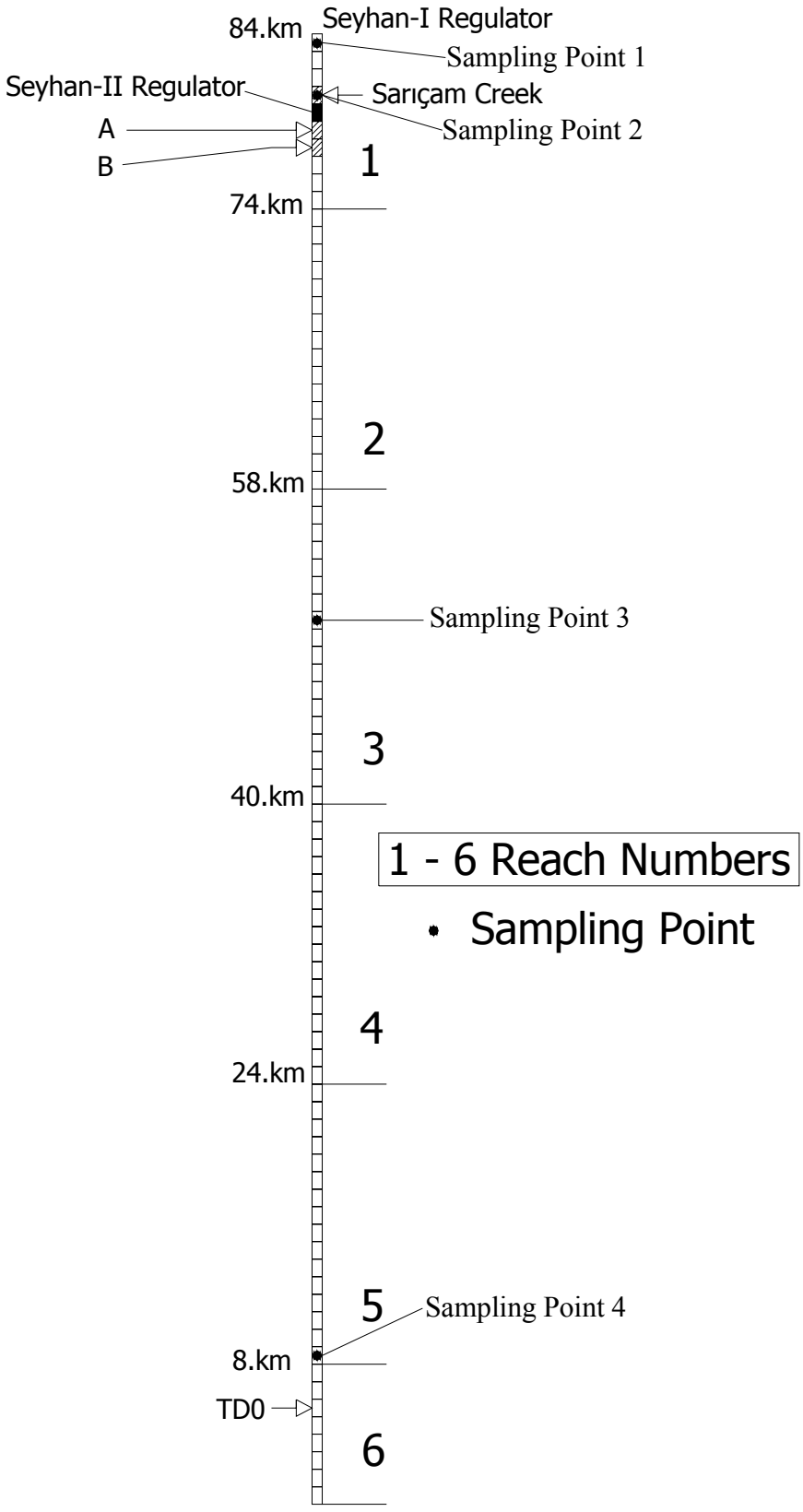

Figure 5.1: Schematic representation of the Lower Seyhan River.

Onur (1996) modeled the river system separately for four distinct hydrological periods. These were the wet (December – May), dry (July, August, September), November, and irrigation (June – November) periods. In this study, the calibration and verification were applied only for the wet and dry seasons that comprise the major portion of the hydrological period. Also, Onur (1996) had used the same kinetic coefficient values for the dry, November, and irrigation periods. Therefore, it was assumed that the river system behaves similarly in these three hydrological periods.

The simulated water quality constituents in Onur's study (1996) were dissolved oxygen, 5-day biochemical oxygen demand (BOD_5), nitrogen cycle, phosphorus cycle, algae and temperature. The same set of constituents was selected to be simulated in this study, as well. In describing the hydraulic characteristics of the river system, Onur (1996) selected to use the 'functional representation' option and set the discharge coefficients representing the relationship among the velocity, depth, and flow as input. These coefficients were calculated by studying 40 different cross-sections throughout the river.

Onur (1996) used the water quality measurement values from the four sampling stations operated by the State Hydraulic Works (DSİ) for the calibration and the verification of the model. The data of the year 1991 was used for calibration and two data sets which were representing the years 1992 and 1993 were used for verification of the calibration results. The same approach was adopted in this study, as well.

The QUAL2E input files used in this study were not exactly identical to the original input files that Onur have used since not all the details related with the specific input files were present in the original work. However, it is believed that sufficient effort was put forward in order to have similar input data by using the information in Onur's work (1996), typical values used in QUAL2E, and personal judgement. As a result of unavoidable differences between the input data used in this study and

the Onur's study, there were also differences in the QUAL2E output reports although the same values were used for the reaction coefficients. However, initial assessment has revealed that the water quality profiles plotted using the generated input data and Onur's profiles were in agreement with each other with only negligible deviations in the wet period. However, for the dry period, the DO profiles were significantly different. This fact should be recognized in the following analysis.

5.3 The Calibration Problem

The set of estimated kinetic coefficients include BOD decay rate coefficient (K_1), BOD settling rate (K_3), sediment oxygen demand (K_4), reaeration coefficient (K_2), ammonia oxidation rate (B_1), nitrite oxidation rate (B_2) and algal settling rate (S_1). All the coefficients were assumed to be reach variable. BOD settling rate was calibrated only for the reaches 1 and 2. It was assumed to be a negligible process for the other reaches. The resultant optimization problem had 38 decision variables. The information about the problem set is given in Table 5.1. The search ranges of the kinetic coefficients were determined based on the QUAL2E documentation (Brown and Barnwell, 1987). The exceptions were the BOD settling rate (K_3) and the sediment oxygen demand (K_4). The lower limit for the possible values of the BOD settling rate was selected as 0 day^{-1} omitting the negative values that reflected resuspension. In fact, Onur (1996) had considered only the positive BOD settling rate values throughout the river system as well and neglect it for some of the reaches. For the sediment oxygen demand (K_4) values, there is no typical range in the QUAL2E documentation and the search range was determined so as to cover the calibrated values in Onur's study (1996).

The calibration problem for the Lower Seyhan River is very similar to Complex Problem III. The objective function (Equation 4.4) used in Complex Problem III allows the use of multiple water quality variables. Also, this formulation can also make use of two different observed data sets for the verification phase. As a result, the optimization model derived for Complex Problem III was used with minor

modifications that will be discussed in the next sections. The water quality variables in the observation data set were DO, BOD_5, N species (organic-N, NH_3-N, NO_2-N, NO_3-N), and dissolved phosphorus (P).

In adapting the optimum calibration model to the case study, the most critical problem arisen was due to imperfect and incomplete observation data set (Tables 5.2 and 5.3 for the dry and wet periods, respectively). Often, but not always, the observed data set contains three different concentration values for a single water quality variable. These different concentration values represent the measurement results at different points in time within the specified period. However, there were missing data. Moreover, the water quality data obtained at the four sampling stations (Figure 5.1) do not always include the concentration values of all state variables.

Table 5.1: Information on decision variable encoding in GA strings for the case study.

DO and BOD coefficients				
Calibrated Parameter	**BOD decay rate coefficient (K_1)**	**BOD settling rate (K_3)**	**Sediment oxygen demand (K_4)**	**Reaeration coefficient (K_2)**
Range	0.02 – 3.4 day^{-1}	0.00 – 0.36 day^{-1}	0.0 – 1.0 g/m^2-day	0.0 – 100 day^{-1}
Required accuracy	0.01 day^{-1}	0.005 day^{-1}	0.1 g/m^2-day	0.1day^{-1}

	N coefficients		**Algal coefficient**
Calibrated Parameter	**NH_3 oxidation rate (B_1)**	**NO_2 oxidation rate (B_2)**	**Algal settling rate (S_1)**
Range	0.10 – 1.00 day^{-1}	0.20 – 2.00 day^{-1}	0.5 – 6.0 m/day
Required accuracy	0.01 day^{-1}	0.01 day^{-1}	0.1 m/day

General	
Total number of decision variables	38
Total string length	278 bits
Population size	850

Table 5.2: Observation data for the 'dry period'.

Sampling Point		1			2			3			4		
Data of the year 1991	DO				8.2	7.2		7.2	7.7		7.1	7.2	
	BOD_5	0.5	0.4		1.8	1.7		2.3	2.2		2.3	2.6	
	Org. N	0.27	0.11		0.34	0.22		0.29			0.38	0.29	
	NH_3-N	0.46	0.15		0.79	0.25		0.15	1.20		0.00	0.33	
	NO_2-N	0.004	0.025		0.190	0.015		0.044	0.074		0.007	0.081	
	NO_3-N	0.54	0.34		0.35	0.36		0.10	0.30		0.12	0.40	
	Dis. P	0.000	0.000		0.000	0.007		0.000	0.007		0.023	0.023	
Data of the year 1992	DO				8.2	5.9		7.2	7.2		8.0	7.5	
	BOD_5	0.8	1.2		2.1	3.0		3.6	3.0		6.4	2.9	
	Org. N	0.04	0.27										
	NH_3-N	0.08	0.16		0.28	0.85		0.76	1.52		0.00	0.10	
	NO_2-N	0.010	0.010		0.009	0.038		0.071	0.113		0.010	0.100	
	NO_3-N	0.54	0.19		0.28	0.80		0.21	0.42		0.09	0.88	
	Dis. P	0.000	0.000		0.037	0.027		0.140	0.150		0.000	0.060	
Data of the year 1993	DO				6.7	6.7		7.3	7.9		6.8	8.3	
	BOD_5	0.9	0.5		2.0	5.8		6.0	5.5		4.4	4.3	
	Org. N	0.25	0.13		0.36	0.67		0.18	0.69		0.38	0.49	
	NH_3-N	0.12	0.08		1.22	0.44		1.28	0.67		0.24	0.09	
	NO_2-N	0.012	0.005		0.022	0.031		0.124	0.003		0.052	0.005	
	NO_3-N	0.52	0.96		0.58	0.44		0.52			0.56	0.08	
	Dis. P	0.000	0.000		0.000	0.000		0.000	0.000		0.000	0.000	

Table 5.3: Observation data for the 'wet period'.

Sampling Point		1			2			3			4		
Data of the year 1991	DO					10.0	8.6		9.3	8.6		9.6	8.4
	BOD_5		1.0	0.7		2.1	1.8		1.5	1.7		1.3	1.2
	Org. N		0.18	0.11		0.22	0.57		0.20	0.09		0.16	0.35
	NH_3-N		0.04	0.16		0.41	0.05		0.08	0.05		0.06	0.00
	NO_2-N		0.006	0.029		0.006	0.017		0.014	0.016		0.013	0.005
	NO_3-N		0.90	0.92		1.06	0.80		0.96	0.84		0.92	0.84
	Dis. P		0.000	0.023		0.030	0.000		0.000	0.003		0.000	0.037
Data of the year 1992	DO				11.6	10.8	9.1	11.2	10.7	8.7	10.5	10.6	8.4
	BOD_5	1.3	0.6	0.6	2.6	1.3	1.5	2.6	2.5	1.3	3.0	2.8	0.9
	Org. N	0.09	0.08	0.22									
	NH_3-N	0.00	0.05	0.24	0.21	0.14	0.20	0.26	0.53	0.18	0.20	0.01	0.05
	NO_2-N	0.000	0.008	0.018	0.010	0.007	0.012	0.007	0.017	0.014	0.008	0.011	0.017
	NO_3-N	1.60	1.04	0.86	1.75	0.98	1.02	0.66	1.40	0.80	1.48	1.42	0.90
	Dis. P	0.000	0.000	0.000	0.000	0.000	0.023	0.000	0.040	0.000	0.000	0.040	0.020
Data of the year 1993	DO				10.6	11.0	8.2	10.8	10.4	9.2	10.7	11.0	8.5
	BOD_5		0.8	0.9	2.3	3.1	1.7	2.4	1.4	2.5	2.3	1.2	4.5
	Org. N	0.04	0.13	0.25	0.27	0.54	0.38	0.34	0.52	0.34	0.43	0.52	0.36
	NH_3-N	0.00	0.00	0.61	0.37	0.57	0.88	0.46	0.35	0.86	0.33	0.48	0.96
	NO_2-N	0.006	0.004	0.007	0.011	0.016	0.004	0.014	0.012	0.008	0.013	0.013	0.013
	NO_3-N	1.32	1.12	0.86	1.72	1.56	1.00	1.68	1.20	1.06	1.88	1.42	1.12
	Dis. P	0.000	0.000	0.006	0.007	0.020	0.000	0.013	0.010	0.000	0.007	0.023	0.013

The original objective function (Equation 4.4) includes the sum of the squared errors from the calibration phase. However, since the number of sampling points or the number of measurements are not constant for all the water quality variables, the magnitude of the sum of the squared errors does not give a correct information about the quality of the solution. The previous runs showed that GA performs well in the calibration of QUAL2E when the objective function is aimed at minimizing the maximum difference between the observed and simulated values. Depending on this observation, it was decided to use the maximum of the squared errors instead of the sum of them. Then, the objective function formulation is slightly modified as follows:

$$Min\,Z = \sum_{j=1}^{K} w_j \left\{ E\max_j \left[1 + w\left(D\max_{C_j}\right)\right]\left[1 + w\left(D\max_{V_j}\right)\right]\right\} \qquad (5.1)$$

where $E\max_j$ represents the maximum of the squared errors from the calibration phase for the water quality variable j. The squared error for a water quality variable is calculated for all of the available sampling points throughout the river and for all of the available measurements for that sampling point. Then the maximum value among these is assigned to $E\max_j$. Maximum deviations for the calibration and verification phases for a water quality variable ($D\max_{Cj}$ and $D\max_{Vj}$) are also determined in the same way.

5.4 Results

The calibration of the Lower Seyhan River was performed for both the dry and wet periods. The GA runs were repeated for three different initial populations as in the previous tests.

5.4.1 Dry Period

The dry period covers the months of July, August and September. This period is characterized by low headwater and incremental flows but high irrigation return flow from the canal TD0. The ambient temperature is high and the other meteorological conditions represent that of a summer season. The coefficient values

obtained through optimization are given in Table 5.4. The values of the assigned weights are given in Appendix D (Table D.4). The water quality profiles plotted using the optimization results are given in Figures 5.2 to 5.8, together with the observation data and the profiles from Onur's (1996) work for the same period.

Table 5.4: Optimization results for the dry period (Case Study).

		Optimization Results
K_1 (day^{-1})	Reach-1	2.43
	Reach-2	2.55
	Reach-3	2.16
	Reach-4	0.98
	Reach-5	0.15
	Reach-6	0.79
K_3 (day^{-1})	Reach-1	0.145
	Reach-2	0.080
K_4 (g/m^2-day)	Reach-1	0.4
	Reach-2	0.2
	Reach-3	0.7
	Reach-4	0.2
	Reach-5	0.1
	Reach-6	0.0
K_2 (day^{-1})	Reach-1	24.6
	Reach-2	27.6
	Reach-3	56.0
	Reach-4	33.8
	Reach-5	29.8
	Reach-6	48.1
B_1 (day^{-1})	Reach-1	0.41
	Reach-2	0.23
	Reach-3	0.11
	Reach-4	0.11
	Reach-5	0.12
	Reach-6	0.17
B_2 (day^{-1})	Reach-1	0.79
	Reach-2	0.81
	Reach-3	0.20
	Reach-4	0.22
	Reach-5	0.20
	Reach-6	0.28
S_1 (m/day)	Reach-1	1.8
	Reach-2	2.8
	Reach-3	1.6
	Reach-4	1.2
	Reach-5	1.0
	Reach-6	1.6

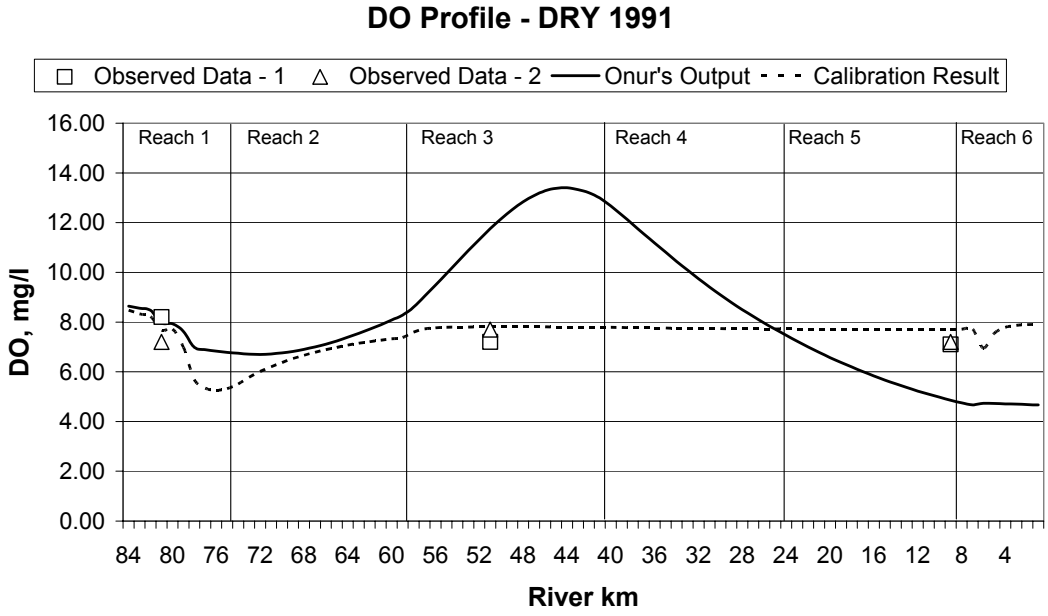

Figure 5.2: DO profile along the Lower Seyhan River for the calibration conditions. (Dry period)

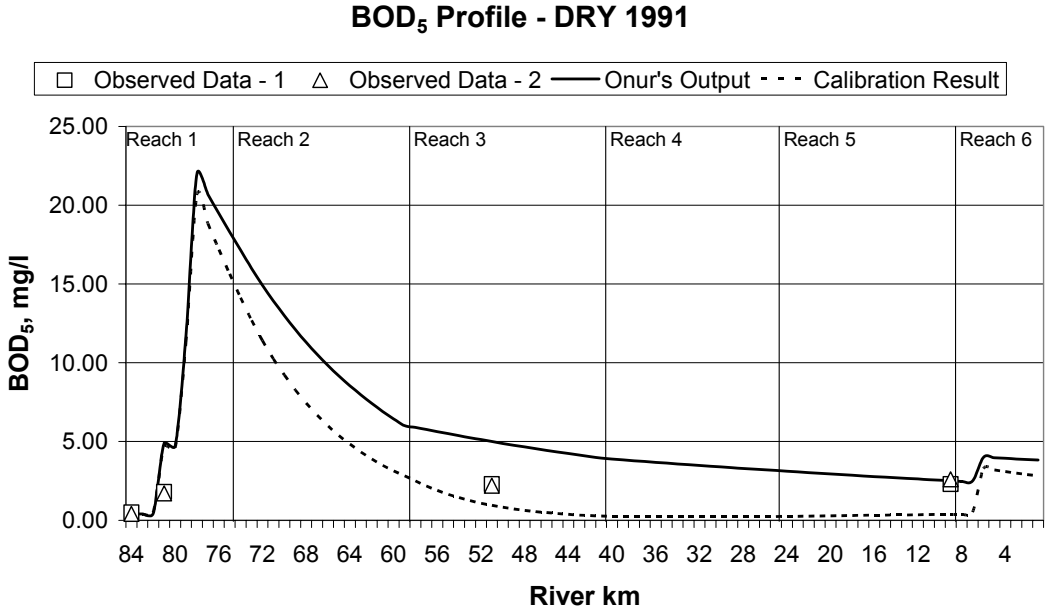

Figure 5.3: BOD_5 profile along the Lower Seyhan River for the calibration conditions. (Dry period)

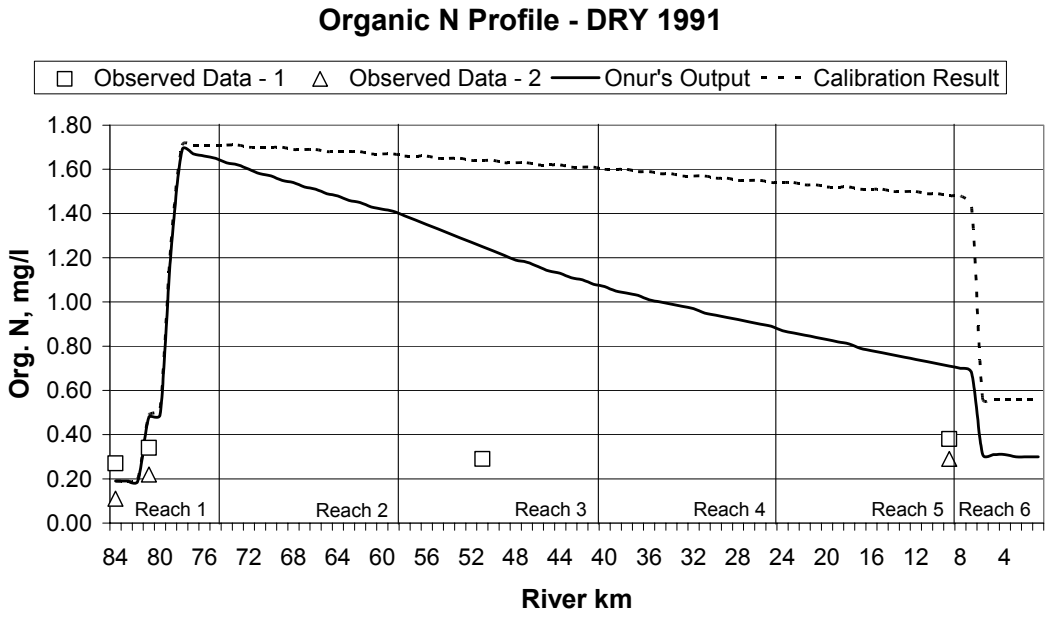

Figure 5.4: Organic N profile along the Lower Seyhan River for the calibration conditions. (Dry period)

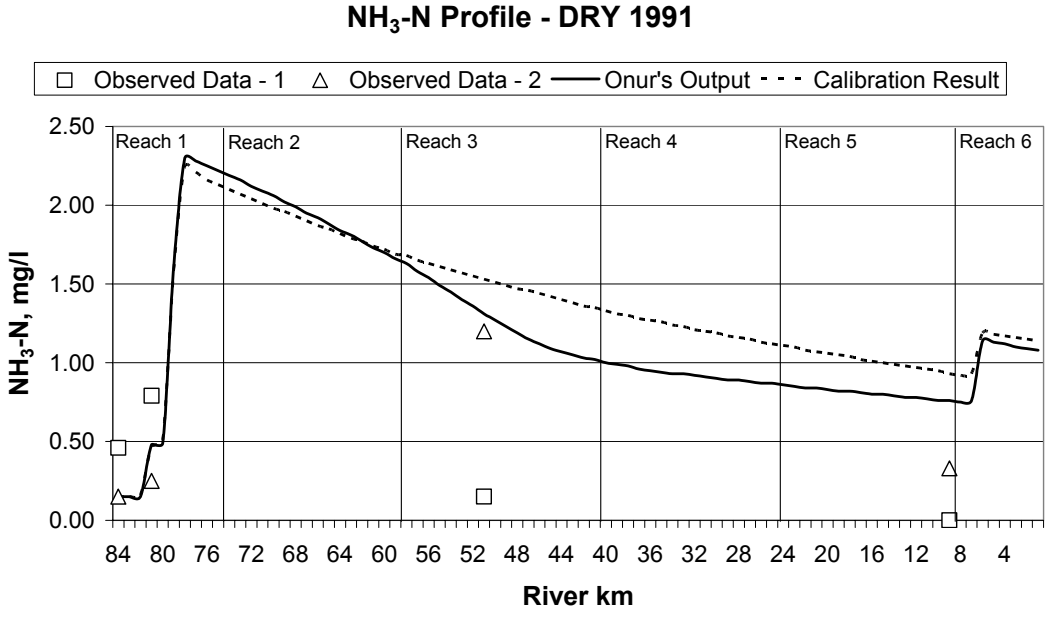

Figure 5.5: NH_3-N profile along the Lower Seyhan River for the calibration conditions. (Dry period)

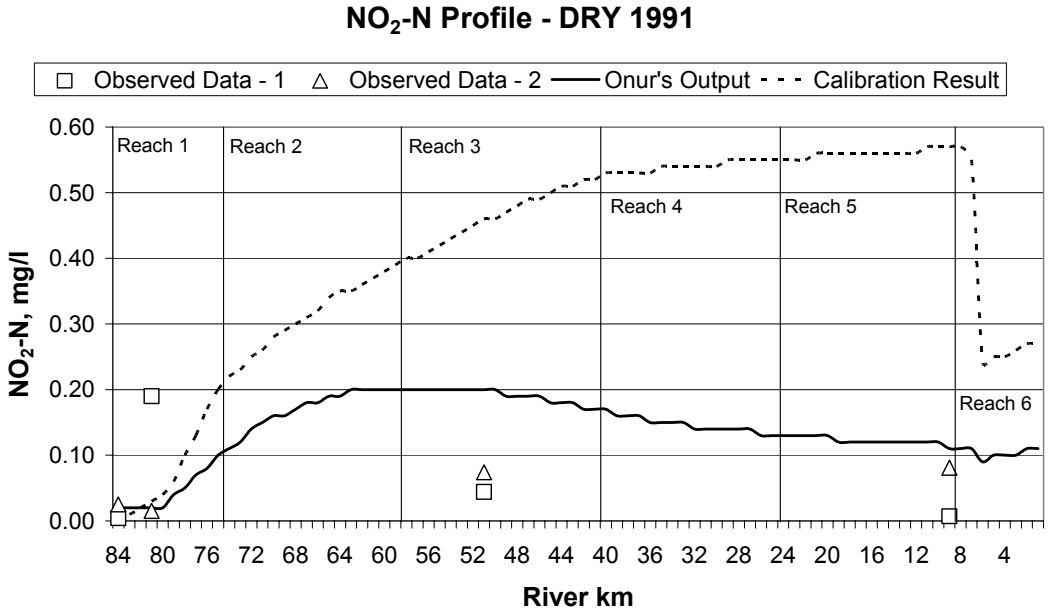

Figure 5.6: NO_2-N profile along the Lower Seyhan River for the calibration conditions. (Dry period)

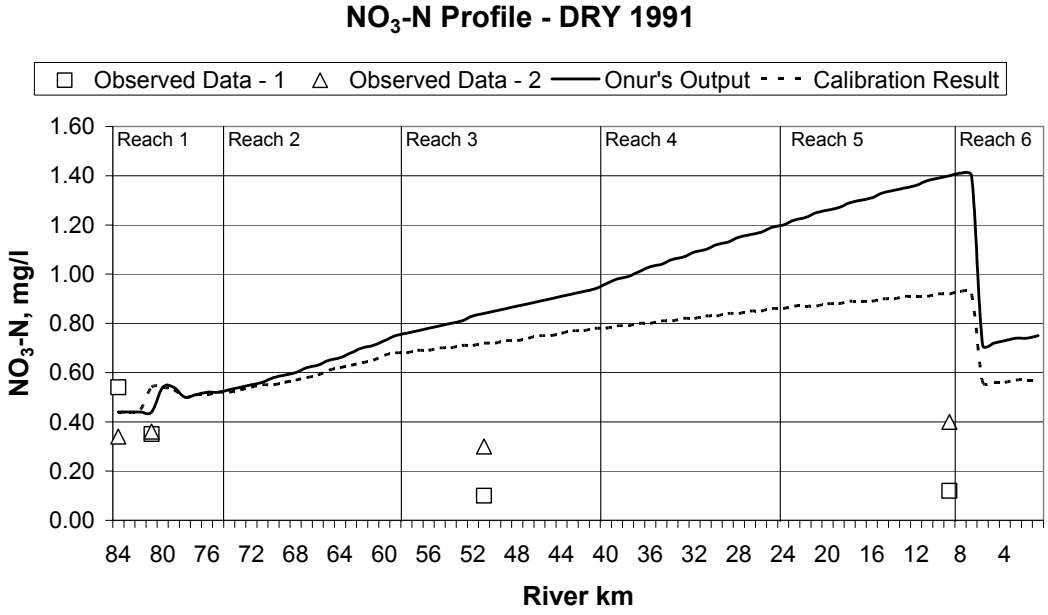

Figure 5.7: NO_3-N profile along the Lower Seyhan River for the calibration conditions. (Dry period)

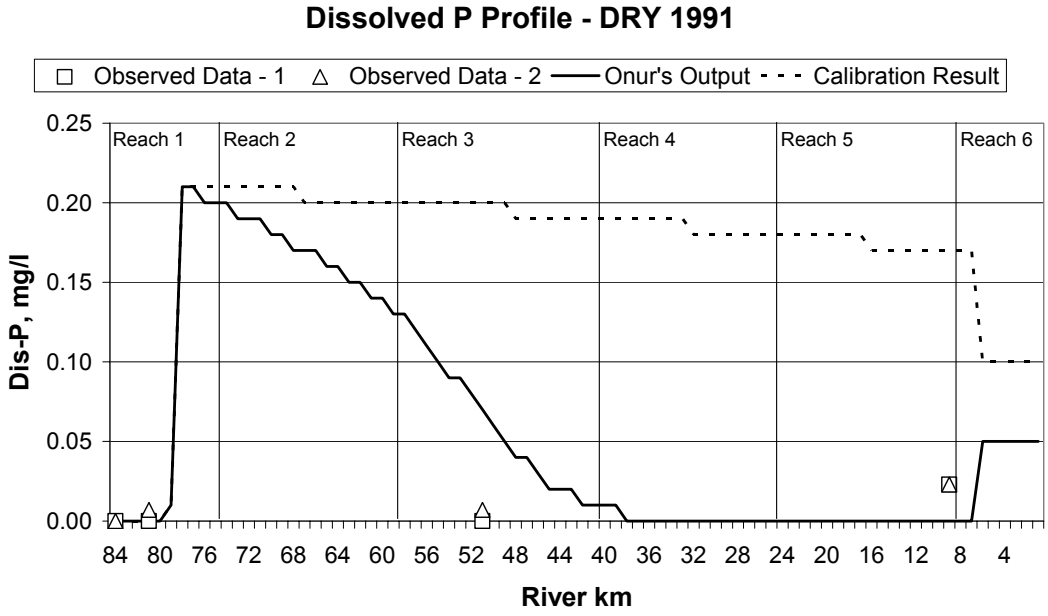

Figure 5.8: Dissolved P profile along the Lower Seyhan River for the calibration conditions. (Dry period)

When the water quality profiles are analyzed, the resultant profiles for DO (Figure 5.2) and BOD_5 (Figure 5.3) were satisfactory. The simulated DO and BOD_5 concentrations were close to the observation data at the sampling points. However, the water quality profiles for the other species resulted in a relatively poor fit.

When the resultant profiles and the profiles from the Onur's study are compared, it is seen that the GA calibration results produced a better DO profile when the deviations from the observation data are considered. Onur's DO profile shows a supersaturation condition in reaches 3 and 4. On the other hand, for the N species (except for NO_3) and dissolved P, Onur's profiles were better. However, it should be emphasized that it was not possible to fully replicate the input algal related parameters of the model due to the lack of sufficient information. Algal growth affects the concentrations of the N and P species as well as the DO levels. Therefore, comparisons made between the results obtained in this study and Onur's work are not necessarily on the same input data and similar conditions in terms of the algal related parameters.

5.4.2 Wet Period

The wet period represents the time of the year between the months of December to May. In this period, the headwater flow rates are much higher than that of the dry period. The incremental flow rate values are also higher. However, the irrigation return flow through canal TD0 is low. As a result, the concentration values of the water quality variables are lower in the wet period because of the dilution effect.

The results of the optimization runs for this period are given in Table 5.5. The values of the assigned weights are given in Appendix D (Table D.4). The water quality profiles plotted using the optimization results are given in Figures 5.9 to 5.15, together with the observation data and profiles from Onur's (1996) study. When the water quality profiles are analyzed, it is seen that the resultant profiles of all the simulated variables fit to the observation data sufficiently well. The profiles of DO and NO_3 can be labeled as the best fits as they lie between the lowest and the highest values of the observation data all through the river (Figures 5.9 and 5.14).

As in the results of the dry period, the simulated profiles of the N and P species have high error. However, the amount of errors in these profiles are lower than the errors observed in the dry period. The BOD_5 profile is out of the range limited by the multiple observation values, but the actual error amount is very low (less than 1 mg/l) (Figure 5.10). When the profiles from the GA calibration results and the profiles from the Onur's output are compared, it is seen that the GA calibration gave better fits to the observation data for all of the state variables except for BOD_5.

Table 5.5: Optimization results for the wet period (Case Study).

		Optimization Results
K_1 (day^{-1})	Reach-1	0.03
	Reach-2	0.17
	Reach-3	0.04
	Reach-4	0.12
	Reach-5	0.34
	Reach-6	1.00
K_3 (day^{-1})	Reach-1	0.000
	Reach-2	0.010
K_4 (g/m^2-day)	Reach-1	0.3
	Reach-2	0.6
	Reach-3	0.1
	Reach-4	0.1
	Reach-5	0.5
	Reach-6	0.1
K_2 (day^{-1})	Reach-1	2.3
	Reach-2	0.6
	Reach-3	0.1
	Reach-4	1.4
	Reach-5	0.8
	Reach-6	9.2
B_1 (day^{-1})	Reach-1	0.10
	Reach-2	0.12
	Reach-3	0.10
	Reach-4	0.14
	Reach-5	0.11
	Reach-6	0.35
B_2 (day^{-1})	Reach-1	1.37
	Reach-2	1.89
	Reach-3	1.33
	Reach-4	1.17
	Reach-5	1.30
	Reach-6	0.74
S_1 (m/day)	Reach-1	1.0
	Reach-2	2.0
	Reach-3	3.1
	Reach-4	1.1
	Reach-5	1.6
	Reach-6	5.2

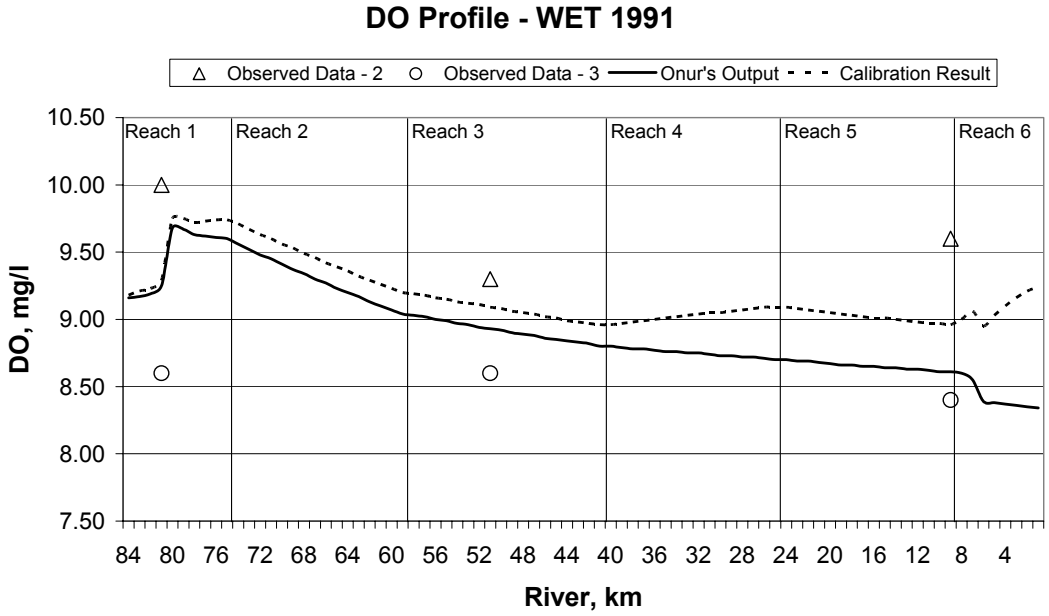

Figure 5.9: DO profile along the Lower Seyhan River for the calibration conditions. (Wet period)

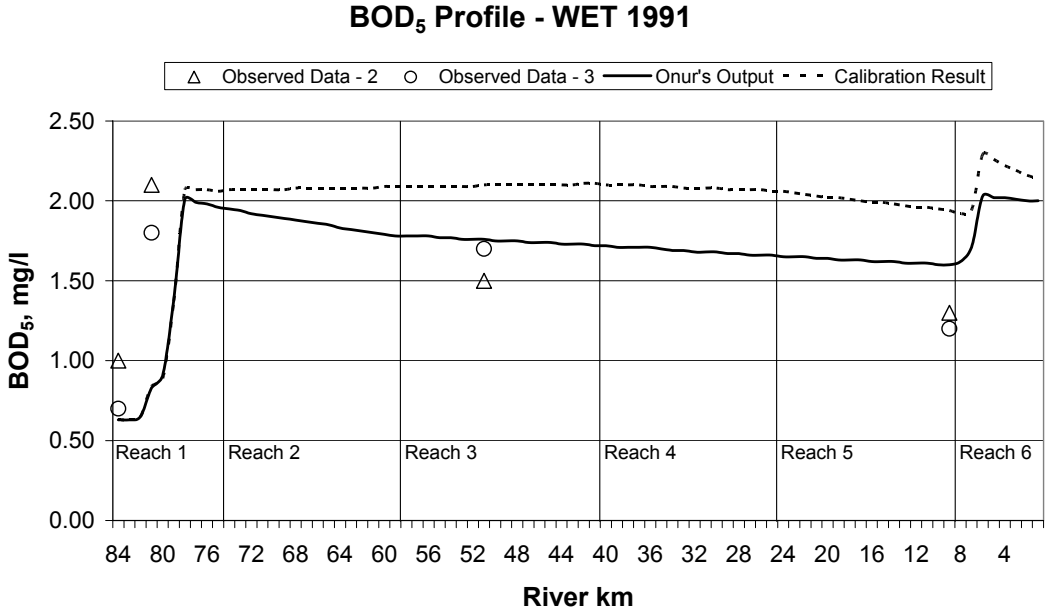

Figure 5.10: BOD_5 profile along the Lower Seyhan River for the calibration conditions. (Wet period)

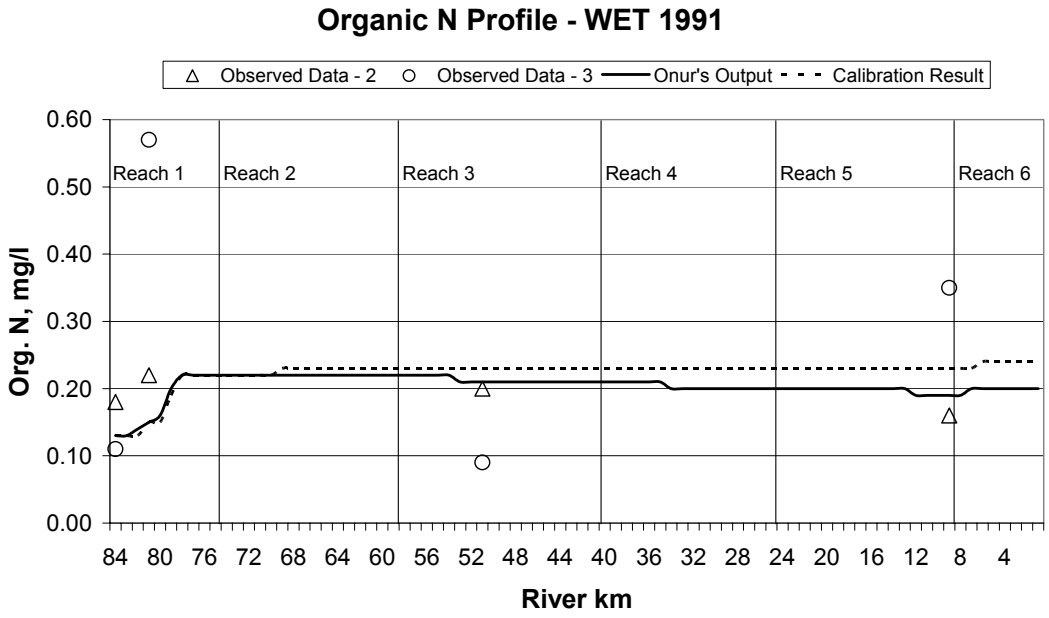

Figure 5.11: Organic N profile along the Lower Seyhan River for the calibration conditions. (Wet period)

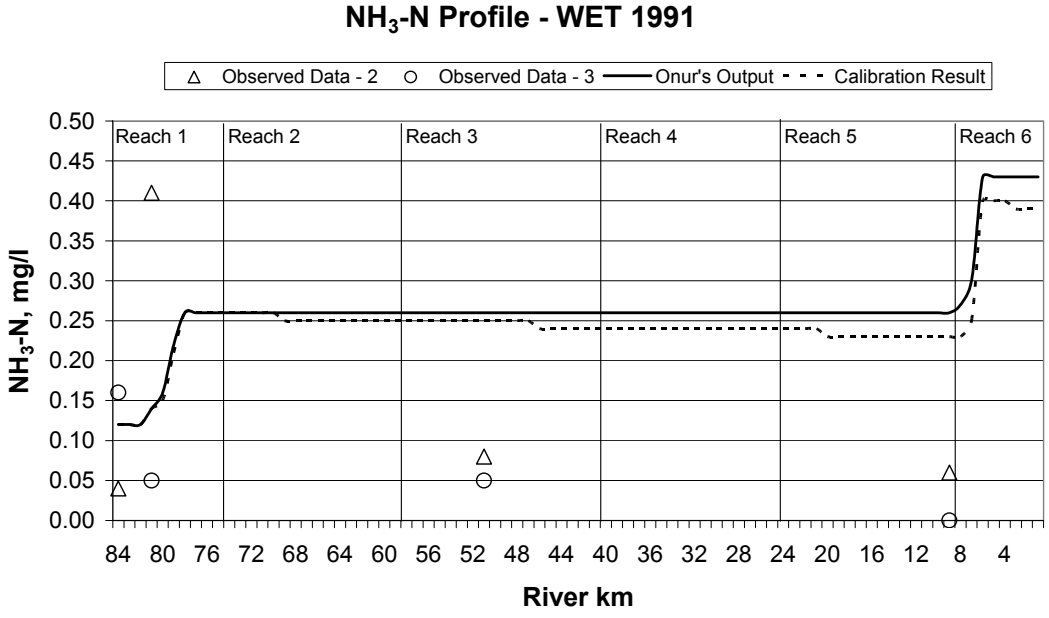

Figure 5.12: NH_3-N profile along the Lower Seyhan River for the calibration conditions. (Wet period)

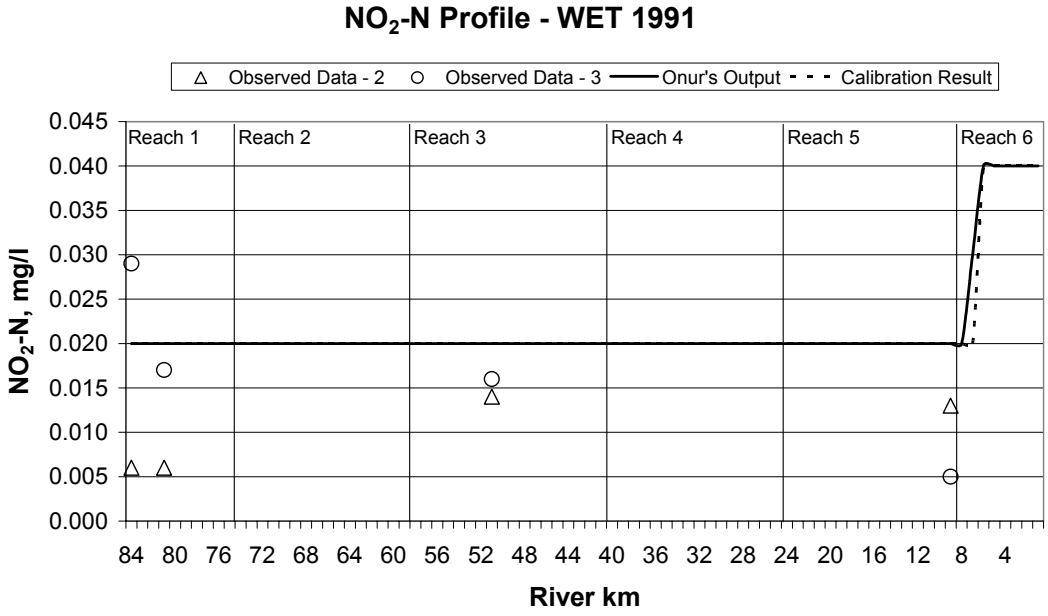

Figure 5.13: NO_2-N profile along the Lower Seyhan River for the calibration conditions. (Wet period)

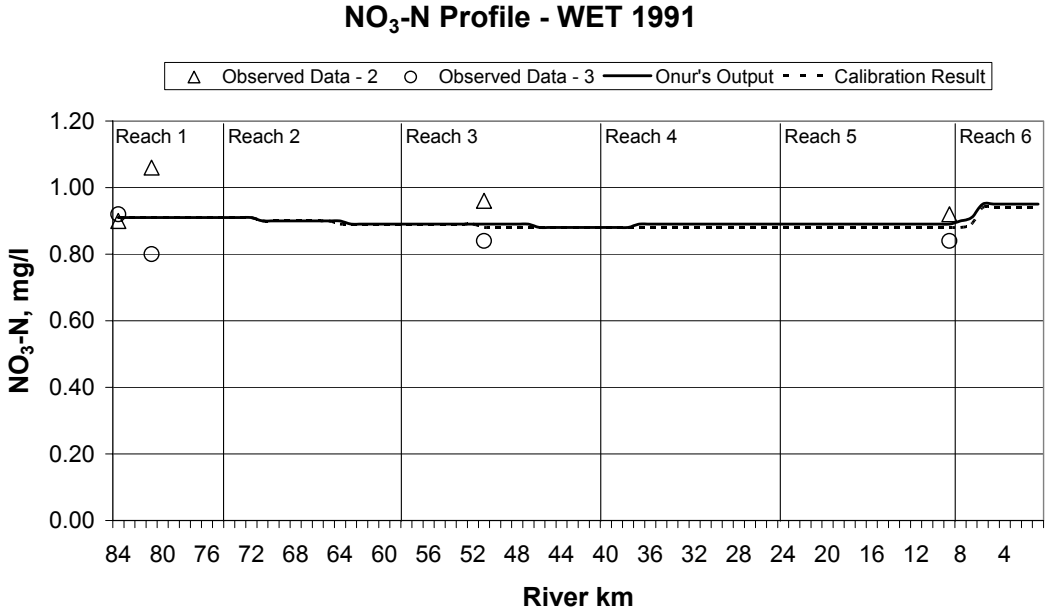

Figure 5.14: NO_3-N profile along the Lower Seyhan River for the calibration conditions. (Wet period)

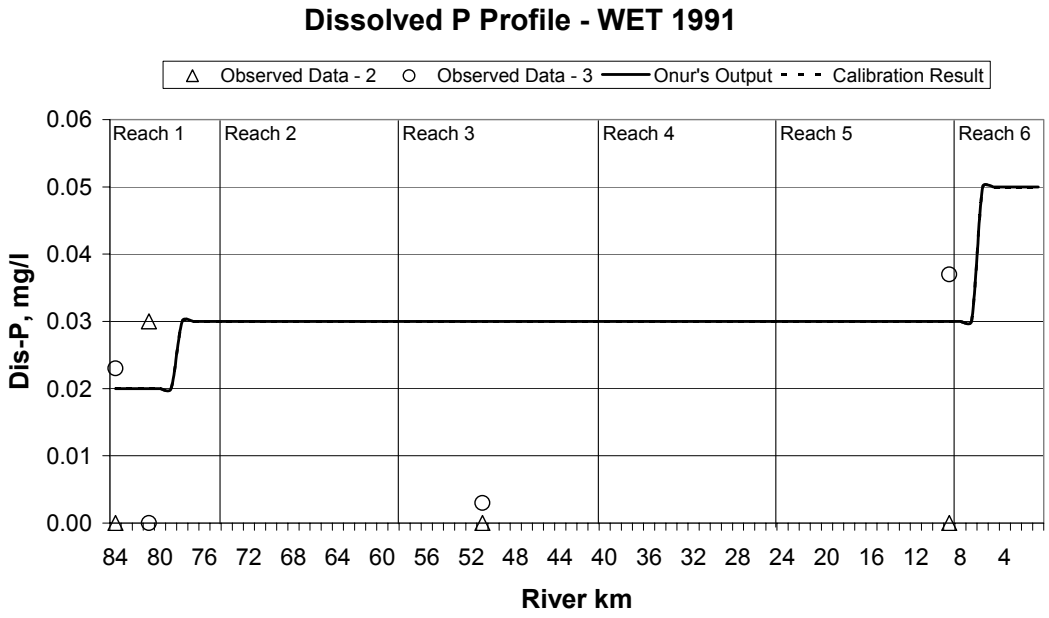

Figure 5.15: Dissolved P profile along the Lower Seyhan River for the calibration conditions. (Wet period)

5.5 Discussion of the Results

The results obtained for the dry period are similar to the results of the previously solved Complex Problems II and III which also gave poorer results for the N species. The N species in the previous problems and the N and P species in this problem both take very low concentration values when compared to the concentration values of DO and BOD_5. A weighting procedure was applied on the objective function in order to overcome the possible optimization problems that would arouse if these variabilities were ignored. However, as it was stated before in the discussion of the results for Complex Problems II and III, the weighting methodology may need further improvement for better results.

In order to adopt the objective function formulation to the available data set, the maximum values of the squared errors and the maximum differences between the simulated and observed values were used. By this way, all the available data was able to be used and the variables with lack of data were able to be treated equally

with the others. However, using the maximum error strategy with multiple data points at one location increases the complexity of the solution. The objective function value can considerably increase or decrease with even a slight change in the value of a decision variable when the reference sampling location for the maximum error has changed. As a result, outcome of the optimization can be sensitive to the extent of variation in sampling data at a location.

Another important issue to be pointed out is that the number of sampling point locations in the case study may be too few for better results. The 84 km. long Lower Seyhan River has only 4 sampling locations. Reach-1 has 2 of these 4 sampling points. In fact, the resultant water quality variable profiles fit to the observation data quite satisfactorily on this reach. In addition, the reliability of the available observation data may also be questionable since there is a very high variance in the observed concentration values at some of the sampling locations.

In Table 5.6, the maximum errors at the sampling points for the 1991 data of both dry and wet periods are given for each water quality variable. The maximum error amounts for Onur's profiles are also represented in the table. For the dry period, GA gave better DO and BOD_5 profiles when compared with the profiles of Onur's output. However, the error amounts were higher in the GA calibration results for the N and P species. For the wet period, the GA calibration results were better for DO and BOD_5. The profiles for the N and P species were very similar for both studies giving an equal amount of maximum error.

When the maximum error values for the GA calibration results for different periods are compared, it is seen that, in general, the amount of the maximum error in the wet period is much less than the amount of maximum error in the dry period for all of the water quality variables. In general, dry periods represent the conditions where reaction coefficient values have a pronounced impact on the water quality. Therefore, any bias introduced into the model by inaccurate model parameters (fixed ones) would have a negative impact on the results. As mentioned earlier,

some of the parameters related to the algal dynamics may not be set correctly, which may impact the overall outcome of the optimization runs.

Table 5.6: The maximum error amounts at the sampling points on the Lower Seyhan River for the data of the year 1991 according to the calibration results of Onur's study and the GA methodology.

Water quality variable	**Maximum error amount (mg/l)**			
	Dry period		**Wet period**	
	Onur's Output	GA results	Onur's Output	GA results
DO	4.55	0.61	0.99	0.71
BOD_5	3.20	3.01	1.27	1.26
Organic N	1.25	1.35	0.42	0.42
NH_3	1.16	1.38	0.27	0.27
NO_2	0.17	0.56	0.02	0.02
NO_3	1.28	0.80	0.15	0.15
Dissolved P	0.07	0.20	0.03	0.03

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.1.Conclusions

In this study, the Enhanced Stream Water Quality Model, QUAL2E, has been linked with a genetic algorithm (GA) library in order to apply optimization for the calibration and verification of the model. Optimization models were developed to perform simultaneous calibration and verification. In general, the optimization model uses an objective function that is formulated according to the sum-of-least squares approach, aiming at minimizing the difference between the observed and simulated conditions. In order to perform simultaneous calibration and verification, verification of the calibrated results was treated as a constraint and inserted into the objective function in the form of a penalty function.

The performance of the optimization model was tested for two kinds of observed data set quality. The perfect data set consisted of observed water quality values that were exactly reflecting the state of the water quality as it is. However, in the biased data set, the observed water quality values had an error.

The results of the tests showed that GA optimization can successfully be applied for the calibration and verification of the QUAL2E model. It is seen that the performance of the optimization model is generally sensitive to the error in the observed data sets and to the number and location of sampling points. Although, it was not possible to obtain the exact values of the kinetic coefficients for any of the tests performed in the study, the coefficient estimates were successful in reflecting the water quality variable profiles in the river. It should be noted that the amount of variation between the observed and the simulated water quality values were very small in the vicinity of the sampling points.

The runs with biased observation data indicated that since the optimization algorithm relies on the biased data, the results of the optimum calibration and verification can be negatively affected. This impact will be related to the extent of bias in the observation data. However, the amounts of error in the simulated water quality values were not higher than the amounts of error in the observed data, and often it was much less. Nevertheless, even with biased observation data, if there is sufficient number of representative sampling points, the optimization model can estimate the coefficient values that can simulate the water quality profiles with an acceptable amount of error.

The impact of different objective function formulations on the optimization model performance was also tested. In these tests, Objective Function – 3, which used the maximum error values instead of the sum-of-the-squared errors, seemed to perform better. However, for the biased observation data case, all the objective functions tested performed similarly. In this case, the error in the observation data dominated the performance of the optimization model rather than the formulation of the objective function.

The application of the optimization model to more complex problems that involve higher number of decision variables, water quality variables, and increased number of observation data sets required a weighting approach in the objective function. With this method, the error function values for each water quality variable were weighted and then summed up to define the value of the objective function. In order to check if this application is beneficial or not, runs were repeated for the unweighted objective function as well. The results pointed out that better performance can be established with the weighted objective function. This impact was more pronounced for the members of the nitrogen (N) cycle. In case where concentrations associated with the N cycle are lower than that of dissolved oxygen (DO) and biochemical oxygen demand (BOD), the contribution of the error from these constituents to the overall objective function value can be insignificant.

Weighting helps to minimize this situation. However, for successful outcomes, adequate weights should be assigned.

The final part of the study was the application of the optimization model for a case study. For this purpose, a previous modeling study conducted on the Lower Seyhan River by Onur (1996) using QUAL2E was utilized. Using the same modeling approach with Onur, two different QUAL2E input files representing the two major hydrological periods, namely dry and wet, were prepared for the Lower Seyhan River. Before running the optimization model on these input files, a modification was necessary to adapt the objective function formulation to the existing observation data. The observation data set had many data deficiencies, and as a result, a non-uniform distribution of data points among the observed water quality variables.

The calibration results for the case study were negatively affected from the inadequate number of sampling point locations throughout the river and from the unusual variability of observation values at some of the sampling locations. The calibration results for the dry period also suffered from the inadequacies of the input data in defining the river system. Nevertheless, the optimum calibration results for the wet period were superior to the calibration results of the original study which employed a trial-and-error approach.

In conclusion, GA optimization can be used as an efficient tool for calibration and verification of the QUAL2E model. With this approach, coefficient estimates and resulting water quality profiles can be obtained with a much less effort compared to the trial-and-error method.

6.2.Recommendations for Future Study

The results of this study showed that the objective function formulation can have a significant impact on the GA performance when calibrating QUAL2E. In addition, different ways of constraint handling can also result in a better performance.

Problem-specific objective functions that perform well for a particular system or data set may be developed by further studies. Objective functions that can handle observation data sets with non-uniform distribution of data points or data deficiencies (e.g. the observation data set used in the Case Study) may be developed for performance enhancements. Moreover, further studies may investigate the use of some additional constraints to formulate a 'supervised calibration'. This supervised calibration technique may especially be useful when the available data is insufficient to guide the GA in the search process. For example, for a river reach with no sampling points, a constraint that limits the value of a state variable or a decision variable according to other known values in the river system may be incorporated to the optimization model.

When using multiple response data in the automated calibration runs, weighting of the contributions of different responses is necessary. In this study, an objective method was developed for weighting and it was shown that the optimization runs with the weighted objective functions gave better results than the runs with the unweighted ones. A research on this subject with new objective weighting methodologies may improve the results. Further studies may explore the use of percentage errors in the objective function calculations so that an adaptive, dynamic weighting methodology can be developed.

It was shown that the number and locations of sampling points have a significant impact on the calibration results. Obtaining the required data to define the system in a most economical way by minimizing the required sampling is currently a popular research subject. Further study can be conducted on the Case Study to define the minimum number and location of representative sampling points.

An important part of this study was linking the QUAL2E model with a GA library. The experience gained in this step and the resultant program may be used in subsequent water quality management studies. The developed program can be

easily adapted to solve optimization-simulation problems in water quality management (e.g. waste load allocations, discharge permit determinations).

REFERENCES

Aksoy, A.; Culver, T. B. (2000) Effects of sorption assumptions on the aquifer remediation designs. **Ground Water**, 38 (2): 200-208.

Bowles, D.S.; Grenney, W.J. (1978) Steady state river quality modeling by sequential extended Kalman filters. **Water Resources Research**, 14 (1): 84-96.

Brown, L.C.; Barnwell, T.O. (1987) *The enhanced stream water quality models Qual2e and Qual2e-Uncas: Documentation and user manual.* EPA Office of Research and Development.

Carroll, D.L. (2001) FORTRAN Genetic Algorithm Driver, version 1.7a, http://cuaerospace.com/carroll/ga.html, last access: 04.08.2004.

Chan Hilton, A. B.; Culver T. B. (2000) Constraint-handling methods for genetic algorithms in optimal pump-and-treat design. **Journal of Water Resources Planning and Management**, ASCE, 126 (3): 128-137.

Chapra, S.C. (1997) *Surface Water-Quality Modeling*. McGraw-Hill, New York, N.Y.

Chaudhury, R.R.; Sobrinho, J.A.H.; Wright, R.M.; Sreenivas, M. (1998) Dissolved oxygen modeling of the Blackstone River (Northeastern United States). **Water Research**, 32 (8): 2400-2412.

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

Cooper, V.A.; Nguyen V.T.V.; Nicell, J.A. (1997) Evaluation of global optimization methods for conceptual rainfall-runoff model calibration. **Water Science and Technology**, 36 (5): 53-60.

Davis, L. (1991) *Handbook of genetic algorithms*, Van Nostrand Reinhold, New York.

Drolc, A.; Zagorc Končan, J. (1996) Water quality modeling of the river Sava, Slovenia. **Water Research**, 30 (11): 2587-2592.

Drolc, A.; Zagorc Končan, J. (1999) Calibration of QUAL2E Model for the River Sava (Slovenia). **Water Science and Technology**, 40 (10): 111-118.

Dussaillant, A.; Munoz, J.F. (1997) Water quality modeling of Mapocho river, Chile, using QUAL2E-UNCAS. Proceedings of the 4th International Conference on Modelling, Measuring and Prediction of Water Pollution, June 1997, Lake Bled, Slovenia; 567-576.

Goldberg, D.E. (1989) *Genetic algorithms in search, optimization and machine learning*. Massachusetts: Addison Wesley Longman.

Ghosh, N.C.; McBean E.A. (1998) Water quality modeling of the Kali River, India. **Water, Air and Soil Pollution**, 102 (1-2): 91-103.

Holland, J.H. (1975) *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. University of Michigan Press: Ann Arbor.

Holland, J.H. (1992) *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. Cambridge, Mass.: MIT Press.

Jaffe, P.R.; Paniconi, C.; Wood, E.F. (1988) Model calibration based on random environmental fluctuations. **Journal of Environmental Engineering**, 114 (5): 1136-1145

Levine, D. (1996) *Users Guide to the PGAPack Parallel Genetic Algorithm Library*. ANL-95/18; Argonne National Laboratory, Argonne, IL.

Little, K.W.; Williams, R.E. (1992) Least-squares calibration of QUAL2E. **Water Environment Research**, 64 (2): 179-185.

Lung. W.S.; Larson, C.E. (1995) Water quality modeling of upper Mississippi River and Lake Pepin. **Journal of Environmental Engineering**, 121 (10): 691-699.

McKinney, D.C.; M.D. Lin. (1994) Genetic algorithm solution of groundwater management models. **Water Resources Research**, 30(6): 1897-1906.

Mulligan, A.E.; Brown, L.C. (1998) Genetic Algorithms for Calibrating Water Quality Models. **Journal of Environmental Engineering**, 124 (3): 202-211.

Ng, A.W.M.; Perera, B.J.C. (2003) Selection of genetic algorithm operators for river water quality model calibration. **Engineering Applications of Artificial Intelligence**, 16: 529-541.

Ning, S.K.; Chang, N.; Yang, L.; Chen, H.W.; Hsu, H.Y. (2000) Assessing pollution prevention program by QUAL2E simulation analysis for the Kao-Ping River basin, Taiwan. **Journal of Environmental Management**, 61: 61-76.

Onur, A.K. (1996) *Behaviour of the Lower Seyhan River Under Different Pollution Control Strategies*. M.S. Thesis, Middle East Technical University.

Onur, A.K.; Soyupak, S.; Yurteri, C. (1997) Aşağı Seyhan Nehri'nde su kalitesi modellemesi çalışmaları. **Su Kirliliği Kontrolü Dergisi**, 7 (2): 41-52.

Onur, A.K.; Ekemen, E.; Soyupak, S.; Yurteri, C. (1999) Management strategies for the Lower Seyhan catchment. **Water Science and Technology**, 40 (10): 177-184.

van Griensven, A.; Bauwens, W. (2001) Integral water quality modeling of catchments. **Water Science and Technology**, 43 (7): 321-328.

van Griensven, A.; Francos, A.; Bauwens, W. (2002) Sensitivity analysis and auto-calibration of an integral dynamic model for river water quality. **Water Science and Technology**, 43 (7): 321-328.

Wood, D.M.; Houck, M.H.; Bell, J.M. (1990) Automated calibration and use of a stream-quality simulation model. **Journal of Environmental Engineering**, 116 (2): 236-249.

Yih, S.M.; Davidson, B. (1975) Identification in nonlinear, distributed parameter water quality models. **Water Resources Research**, 11 (5): 693-704.

APPENDIX A

AN EXAMPLE CODE

The following is the code of the subroutine 'func' which runs as a part of the evaluation phase of the genetic algorithm. The optimization algorithm for the calibration and verification of QUAL2E is coded in this subroutine. This subroutine transfers the parameter values associated with a string to QUAL2E and calls it for performing a water quality simulation. Following the QUAL2E simulation, the objective function value is evaluated using the quantities of the relavant water quality variables. Finally, the subroutine returns the objective function value.

The following code employs Objective Function – 1. The objective function coding should be changed for the runs for different objective function formulations. Also, an additional code segment is required for Complex Problem III and Case Study to make use of two verification data sets.

```
c##############################################################
c
      subroutine func(j,funcval)
c
      implicit real*8 (a-h,o-z)
      save
c
      include 'params.f'
c-----------------------------------
c mainga.var was included by Recep Kaya GÖKTAS, 30.06.2003
c needed to link GA with QUAL2E for calibration
c This file is the same with main.var, but SAVE statement is
c commented out.
c
     INCLUDE 'mainga.var'
c-------------------------------------------------------
      dimension parent(nparmax,indmax)
      dimension iparent(nchrmax,indmax)
c---------------------------------------
c HERE ARE THE DIMENSIONS OF THE ARRAYS USED IN THE OBJECTIVE
c FUNCTION
```

```
C RECEP KAYA GÖKTAS, 17.06.2003
C
      dimension rcoe(nparmax)
      dimension sim(MC,nobsp)
      dimension noloc(MC,3),obsc(MC,nobsp),obsv(MC,nobsp)
      dimension w(nobsp)
c error(1,i) = sum of squared errors in calibration for
c              simulated parameter i
c error(2,i) = max. deviation in calibration  for simulated
c              parameter i
c error(3,i) = sum of squared errors in verification for
c              simulated parameter i
c error(4,i) = max. deviation in verification for simulated
c              parameter i
      dimension error(4,nobsp)
c---------------------------------------
c
       common / ga2   / nparam,nchrome
       common / ga3   / parent,iparent
c
c---------------------------------------------------------------
c THIS COMMON BLOCK CONTAINS THE ARRAYS USED IN CALIBRATION
C an attempt to develop a more generic program for
C calibration optimization.
C RECEP KAYA GÖKTAS, 26.06.2003
c
       common /calibr/ noloc,obsc,obsv,w
       common    nconv
c---------------------------------------------------------------
c These 70 character strings are used in changing the
c contents of QUAL2E.SUP file
c Names of the QUAL2E input files for calibration and
c verification
c Recep Kaya GÖKTAS, 07.07.2003
c
      character*70 calfil,verfil,calout,verout
c
c-------------------------------------------------
c contents of the error array is initialized to 0
      do 5009 i=1,4
            do 5009 k=1,nobsp
                  error(i,k)=0
5009  continue
c---------------------------------------------------------------
c If the elements do not converge, QUAL2E should not stop the
c program.
c Instead it should return to GA
c Then assign a very low fitness value.
c Result of the convergence test is kept in the value of
c "nconv".
c Convergence test is made in SSCONV.FOR
c SSCONV.FOR runs under Q2EZ.FOR
```

```
c nconv = 1979 -> elements not converged
c nconv = else -> elements converged
c Assign the deafult value of nconv as 2004:
      nconv=2004
c-------------------------------------------------------------
c
c +++++++++++++++++
c     Calibration
c +++++++++++++++++
c
      do 5010 i=1,nloc
            do 5010 k=1,nobsp
                  sim(i,k)=0
5010  continue
      do 5015 i=1,nparmax
            rcoe(i)=0
5015  continue
       do 5050 i=1,nparam
            rcoe(i)=parent(i,j)
5050  continue
c
      nrkg=nloc
      if (nloc.lt.nparam) nrkg=nparam
c
      OPEN(UNIT=1979,FILE='rkg.txt',
     1     FORM='FORMATTED',ACCESS='SEQUENTIAL')
      REWIND(1979)
       write (1979,5060) nrkg
       do 5020 i=1,nrkg
            write(1979,5001) rcoe(i),(sim(i,k),k=1,nobsp)
5020  continue
      close (1979)
c
c     icalibr=1 -> Calibration phase, use calibration input
c                  file
c
      icalibr=1
c
c     Calculating the objective function values
c
      call Q2e3p1
c-------------------------------------------------------------
c If the elements do not converge, QUAL2E should not stop the
c program.
c Instead it should return to GA
c Then assign a very low fitness value.
c Result of the convergence test is kept in the value of
c "nconv".(main.var,mainga.var)
c Convergence test is made in SSCONV.FOR
c SSCONV.FOR runs under Q2EZ.FOR
c nconv = 1979 -> elements not converged
c nconv = else -> elements converged
```

```
c
      if (nconv.EQ.1979) go to 6000
c-------------------------------------------------------------
c
      OPEN(UNIT=1979,FILE='rkg.txt',status='old',
     1     FORM='FORMATTED',ACCESS='SEQUENTIAL')
      read (1979,5060) nrkg
      do 5030 i=1,nrkg
            read(1979,5000) rcoe(i),(sim(i,k),k=1,nobsp)
5030  continue
      close(1979)
c
c Calculate the sum of the squared errors and maximum
c deviations for each observed variable,
c error(1,i) and error(2,i)
      do 5039 i=1,nobsp
            do 5039 k=1,nloc
                  x=(sim(k,i)-obsc(k,i))**2
                  error(1,i)=error(1,i)+x
                  if (abs(sim(k,i)-obsc(k,i)).GE.error(2,i))
     1             then
                        error(2,i)=abs(sim(k,i)-obsc(k,i))
                  endif
5039  continue
c
c++++++++++++++++++
c     Verification
c++++++++++++++++++
c
      do 5090 i=1,nloc
            do 5090 k=1,nobsp
                  sim(i,k)=0
5090  continue
c
c     icalibr=0 -> Verification phase, use verification input
c                  file
c
      icalibr=0
c
c     Calculating the penalty function
c
      call Q2e3p1
c-------------------------------------------------------------
c If the elements do not converge, QUAL2E should not stop the
c program.
c Instead it should return to GA
c Then assign a very low fitness value.
c Result of the convergence test is kept in the value of
c "nconv".(main.var,mainga.var)
c Convergence test is made in SSCONV.FOR
c SSCONV.FOR runs under Q2EZ.FOR
c nconv = 1979 -> elements not converged
```

```
c nconv = else -> elements converged
c
      if (nconv.EQ.1979) go to 6000
c-------------------------------------------------------------
c
      OPEN(UNIT=1979,FILE='rkg.txt',status='old',
     1     FORM='FORMATTED',ACCESS='SEQUENTIAL')
      read (1979,5060) nrkg
      do 5100 i=1,nrkg
            read(1979,5000) rcoe(i),(sim(i,k),k=1,nobsp)
5100  continue
c
      close(1979)
c
c Calculate the sum of the squared errors and maximum
c deviations for each observed variable,
c error(3,i) and error(4,i)
      do 5109 i=1,nobsp
            do 5109 k=1,nloc
                  x=(sim(k,i)-obsv(k,i))**2
                  error(3,i)=error(3,i)+x
                  if (abs(sim(k,i)-obsv(k,i)) .GE. error(4,i))
     1             then
                        error(4,i)=abs(sim(k,i)-obsv(k,i))
                  endif
5109  continue
c
c ++++++++++++++++++++++++++++++++++++++++
c     Calculation of the objective function with the
c constraint
c ++++++++++++++++++++++++++++++++++++++++
c
c Recep Kaya GOKTAS, 22.05.2004
c
c F1  funcval=-z*(1+p)
c F2  funcval=-(z*(1+p)+p*(1+z))
c F3  funcval=-z*(1+100*dmaxc)*(1+100*dmaxv)
c
c F1 is used in this run:
c-------------------------------------------------------------
c If the elements do not converge, QUAL2E should not stop the
c program.
c Instead it should return to GA
c Then assign a very low fitness value.
c Result of the convergence test is kept in the value of
c "nconv".(main.var,mainga.var)
c Convergence test is made in SSCONV.FOR
c SSCONV.FOR runs under Q2EZ.FOR
c nconv = 1979 -> elements not converged
c nconv = else -> elements converged
c-----------------------------------------------
6000  continue
```

```
      if (nconv.EQ.1979) then
                  funcval=-((10.0)**11)
            else
                  funcval=0
                  do 5110 i=1,nobsp
            funcval=funcval-(w(i)*error(1,i)*(1+error(3,i)))
5110              continue
      endif
c
c     Format statements
c
5000  format (F11.0,6F7.0)
5001  format (F11.3,6F7.2)
5060  format (1X,I4)
c---------------------------------------
c
       return
       end
c#################################################################
```

APPENDIX B

INFORMATION ABOUT THE REQUIRED INPUT AND OUTPUT FILES

Information about the required input files:

- MSGFILE.DAT : contains the QUAL2E run-time messages.
- QUAL2E.SUP : run-time supervisor file that is used to answer interactive prompts issued during execution of the QUAL2E model.
- INPUT1.DAT : QUAL2E input file describing the river system with the conditions to be used in the calibration phase.
- INPUT2.DAT : QUAL2E input file describing the river system with the conditions to be used in the verification phase.
- GA.INP : contains the required genetic algorithm input parameters.
- CONS.TXT : contains the declaration of the kinetic parameters to be calibrated.
- OBSC.TXT : contains the observed data information; observation point locations, and the observed values of the monitored parameters for the calibration conditions. The weights associated with each monitored parameter are also declared in this file.
- OBSV.TXT : contains the observed data information; observation point locations, and the observed values of the monitored parameters for the verification conditions
- INPUTFIL.TXT : contains the names and locations of the QUAL2E input and output files to be used both in the calibration and verification phases (user defined names and locations for INPUT1.DAT, INPUT2.DAT, OUTPUT1.DAT, OUTPUT2.DAT.)

After running the program, calculation results are written to a set of text files. Some of these files can be omitted by commenting out the related lines in the code. The names of the output text files with their descriptions are given below:

- DATA.TXT: The scratch file used for information tranfer between QUAL2E and the genetic algorithm driver.
- OUTPUT1.DAT: QUAL2E output file for the run with the input file INPUT1.DAT.
- OUTPUT2.DAT: QUAL2E output file for the run with the input file INPUT2.DAT.
- GA.OUT : Output report of the genetic algorithm run.
- GA.RESTART : contains the population information to be used in a restart continuation of the genetic algorithm run.
- SUMMARY.TXT: The summary part of the genetic algorithm output report is also written in a separate file.
- BESTPARAMS.TXT: The best fitness value at the end of the last generation is recorded in this file. The file contains the generation number, best fitness value and the corresponding values of the parameters.

APPENDIX C

EXAMPLE WATER QUALITY PROFILES FOR THE VERIFICATION CONDITIONS

C.1 Results for the Calibration and Verification for Two Model Parameters

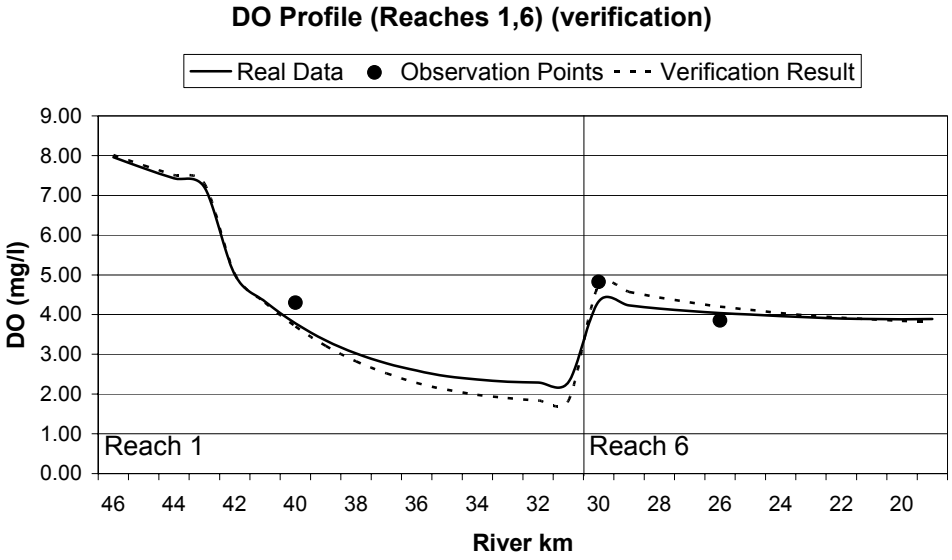

Figure C.1: DO profile along the main river for verification conditions. (Estimated coefficients: K_2 and K_4 for biased observation data) **(Corresponds to Figure 4.4)**

C.2 The Impact of the Number and Location of Sampling Points

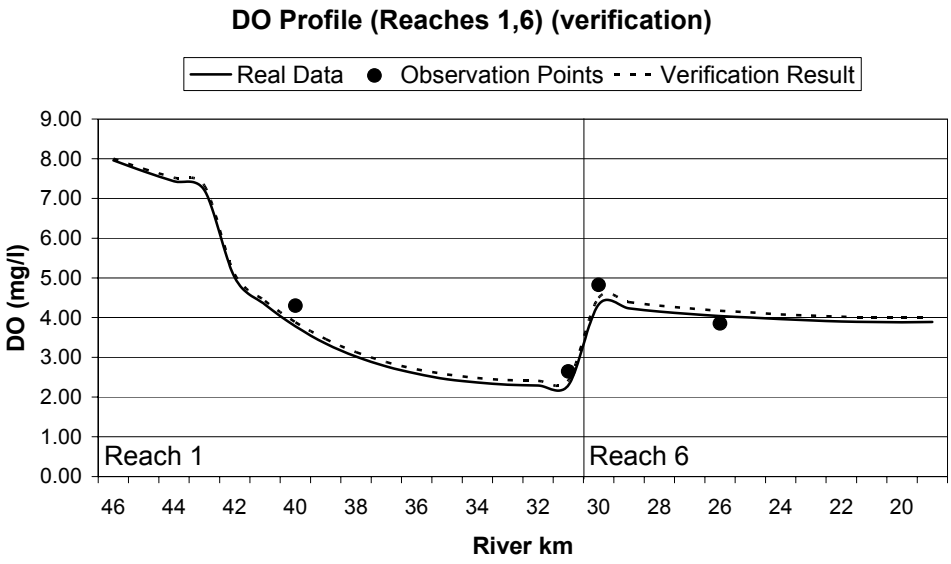

Figure C.2: DO profile along the main river for 9 sampling points case for the verification conditions. (Estimated coefficients: K_2 and K_4 for biased observation data) **(Corresponds to Figure 4.9)**

C.3 The Impact of Different Objective Function Formulations

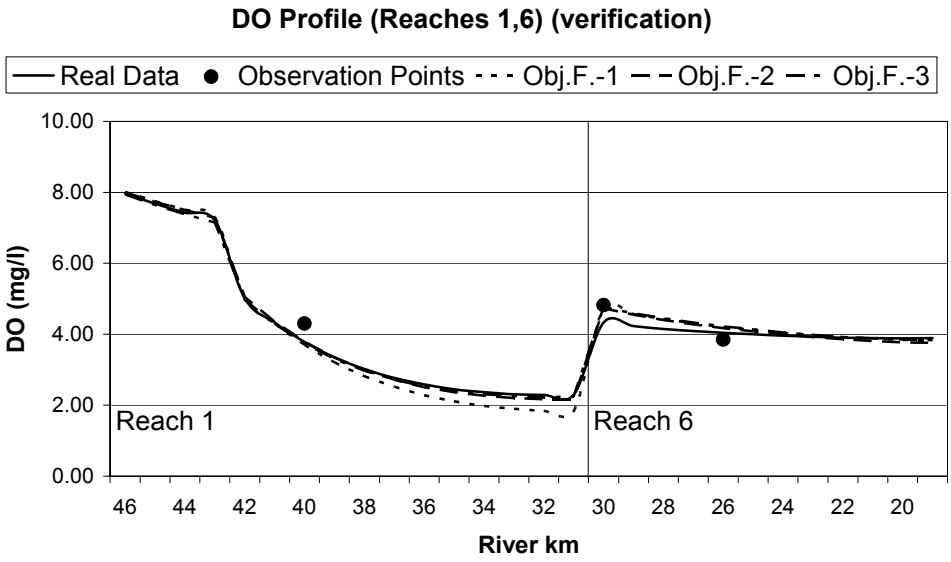

Figure C.3: DO profiles along the main river for three different objective function formulations for the verification conditions. (Estimated coefficients: K_2 and K_4 for biased observation data) **(Corresponds to Figure 4.13)**

C.4 Complex Problem I

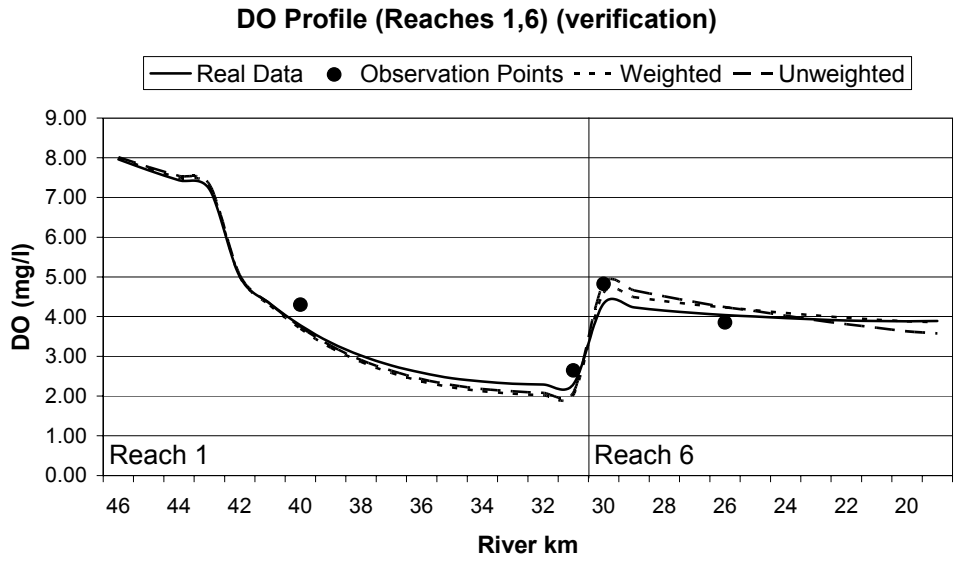

Figure C.4: DO profile along the main river for the verification conditions. (Estimated coefficients: K_1, K_3, K_4 and K_2 for biased observation data) **(Corresponds to Figure 4.19)**

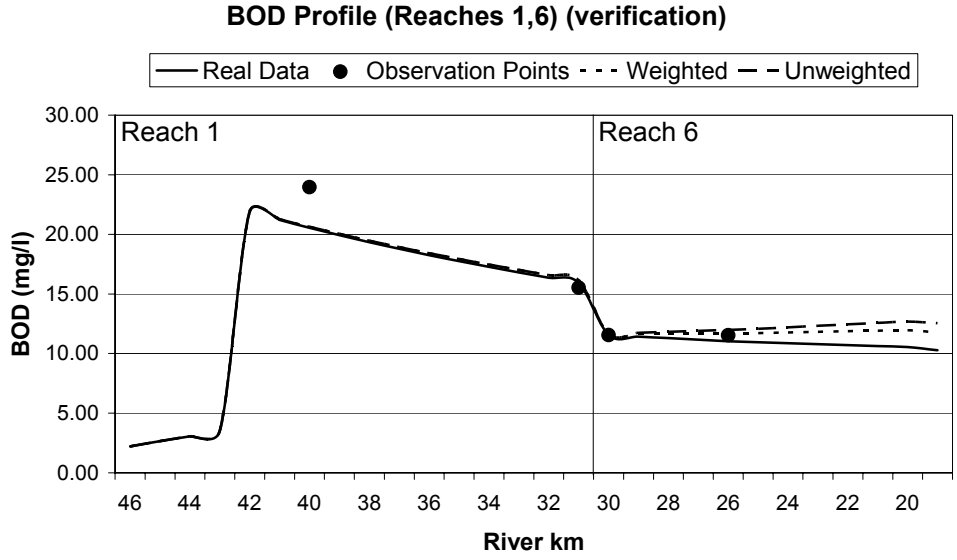

Figure C.5: BOD profile along the main river for the verification conditions. (Estimated coefficients: K_1, K_3, K_4 and K_2 for biased observation data) **(Corresponds to Figure 4.21)**

C.5 Complex Problem II

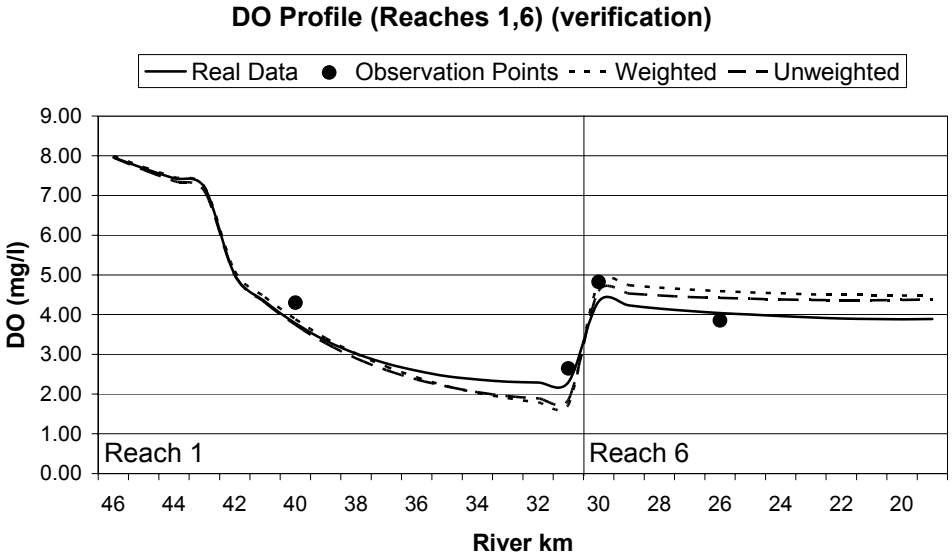

Figure C.6: DO profile along the main river for Complex Problem II for the verification conditions. (Biased observation data) **(Corresponds to Figure 4.35)**

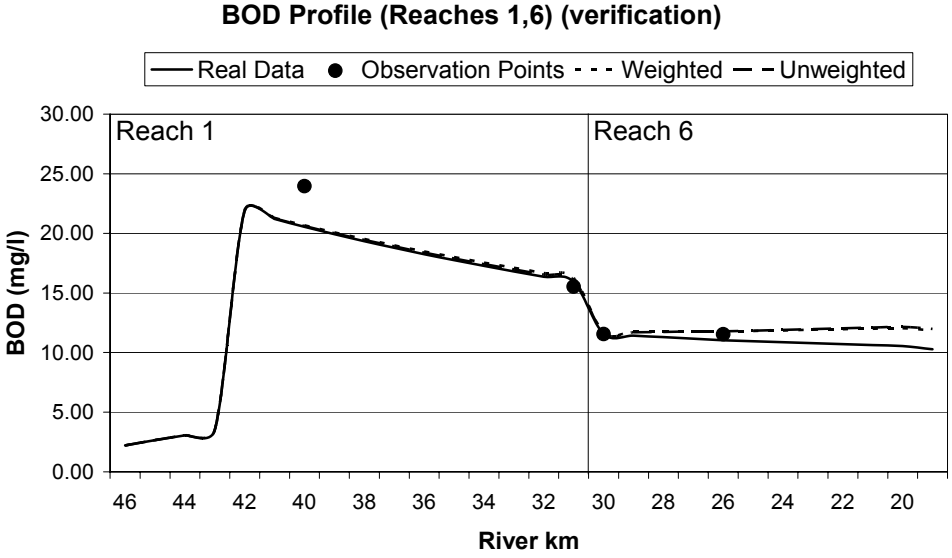

Figure C.7: BOD profile along the main river for Complex Problem II for the verification conditions. (Biased observation data) **(Corresponds to Figure 4.37)**

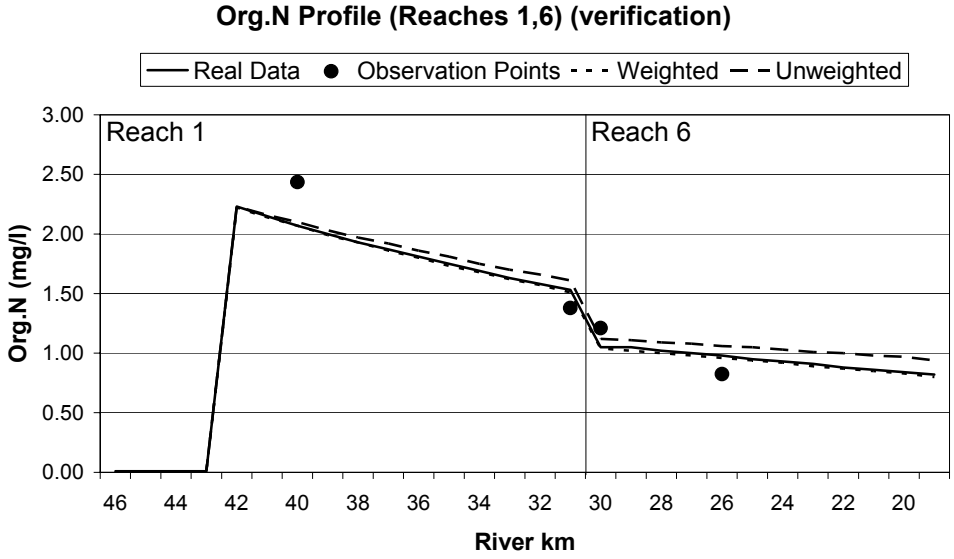

Figure C.8: Organic N profile along the main river for Complex Problem II for the verification conditions. (Biased observation data) **(Corresponds to Figure 4.39)**

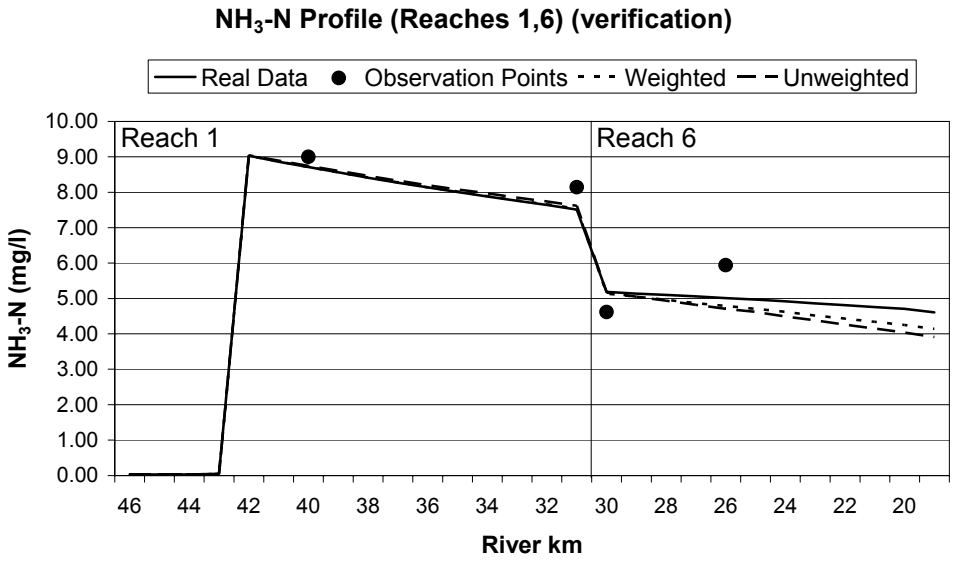

Figure C.9: NH_3-N profile along the main river for Complex Problem II for the verification conditions. (Biased observation data) **(Corresponds to Figure 4.41)**

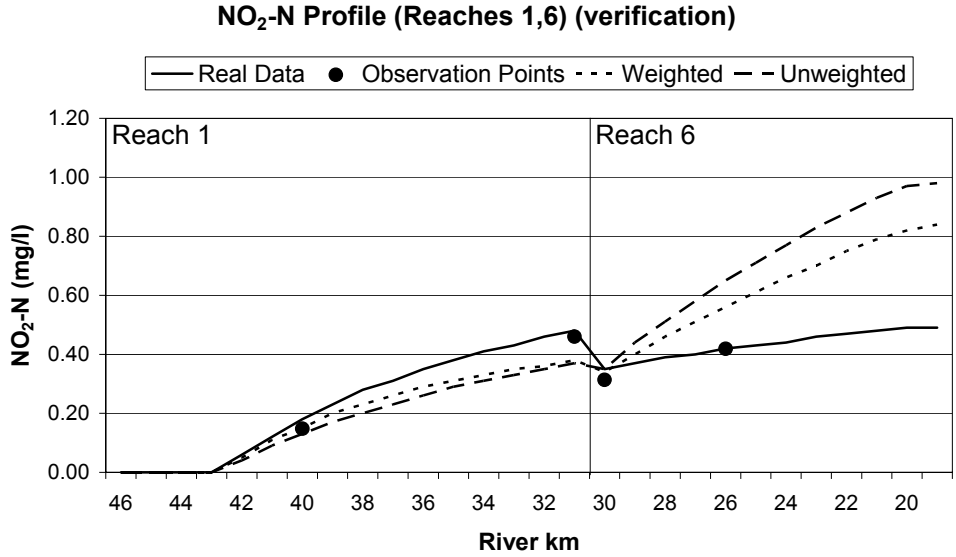

Figure C.10: NO_2-N profile along the main river for Complex Problem II for the verification conditions. (Biased observation data) **(Corresponds to Figure 4.43)**

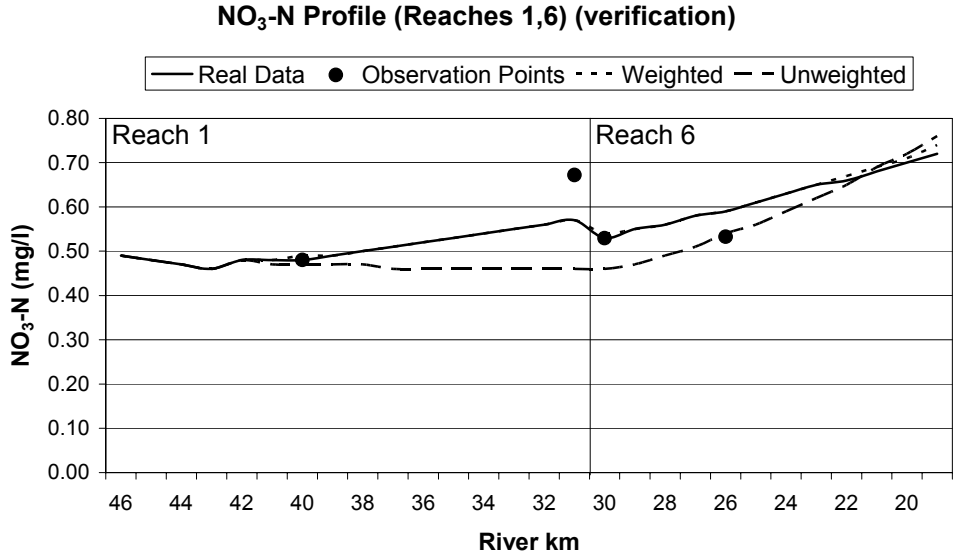

Figure C.11: NO_3-N profile along the main river for Complex Problem II for the verification conditions. (Biased observation data) **(Corresponds to Figure 4.45)**

C.6 Complex Problem III

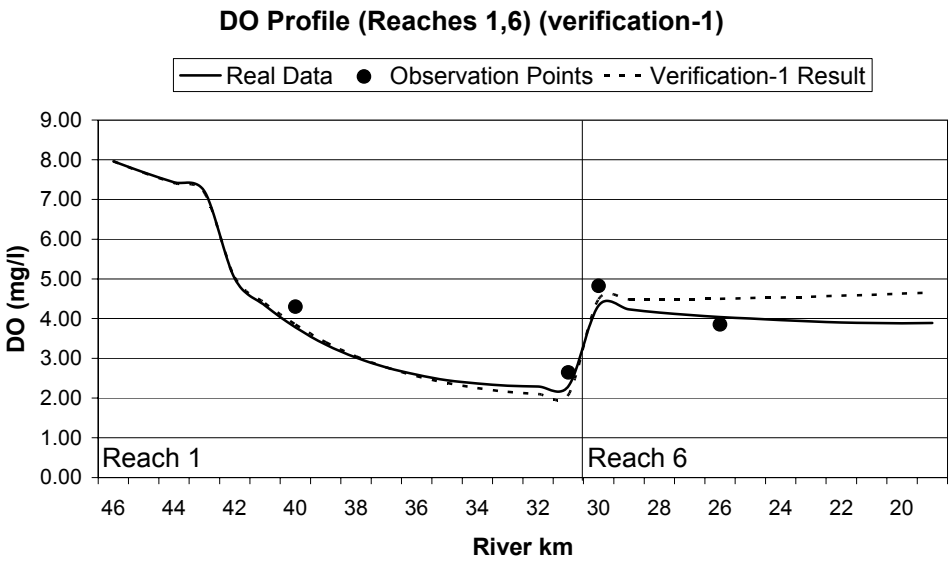

Figure C.12: DO profile along the main river for Complex Problem III for the first verification conditions. (Biased observation data) **(Corresponds to Figure 4.59)**

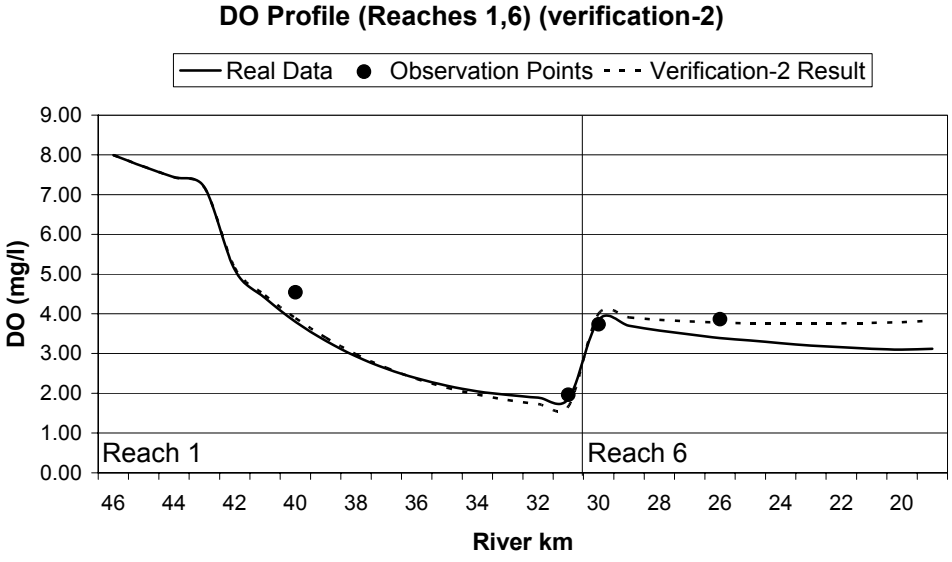

Figure C.13: DO profile along the main river for Complex Problem III for the second verification conditions. (Biased observation data) **(Corresponds to Figure 4.59)**

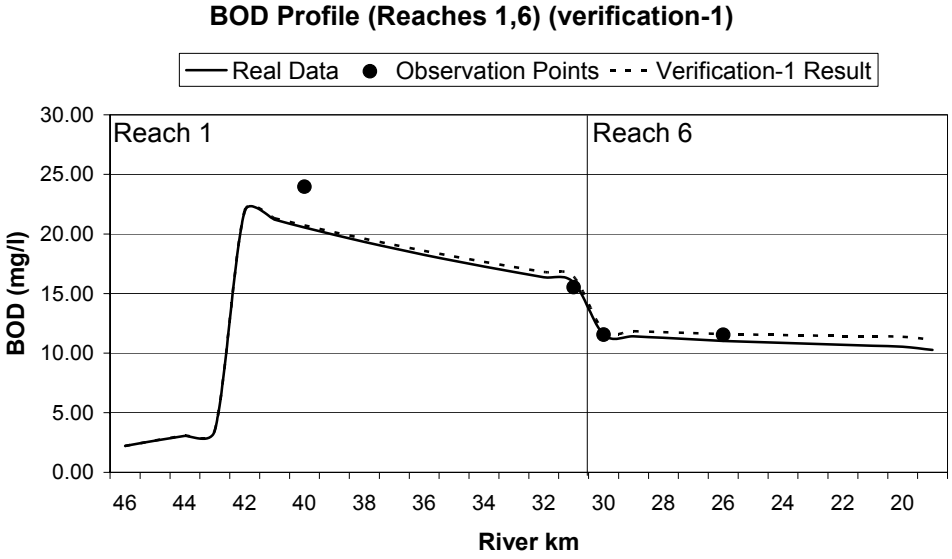

Figure C.14: BOD profile along the main river for Complex Problem III for the first verification conditions. (Biased observation data) **(Corresponds to Figure 4.61)**

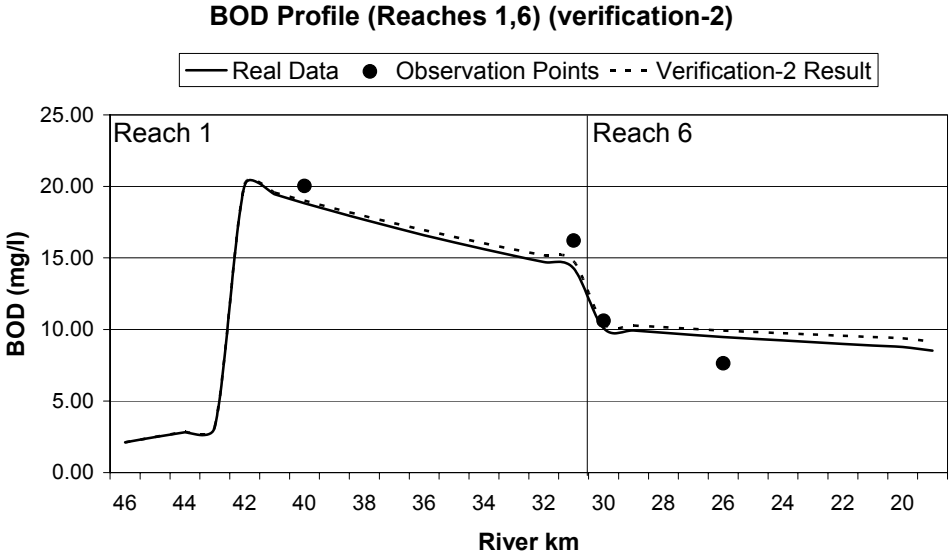

Figure C.15: BOD profile along the main river for Complex Problem III for the second verification conditions. (Biased observation data) **(Corresponds to Figure 4.61)**

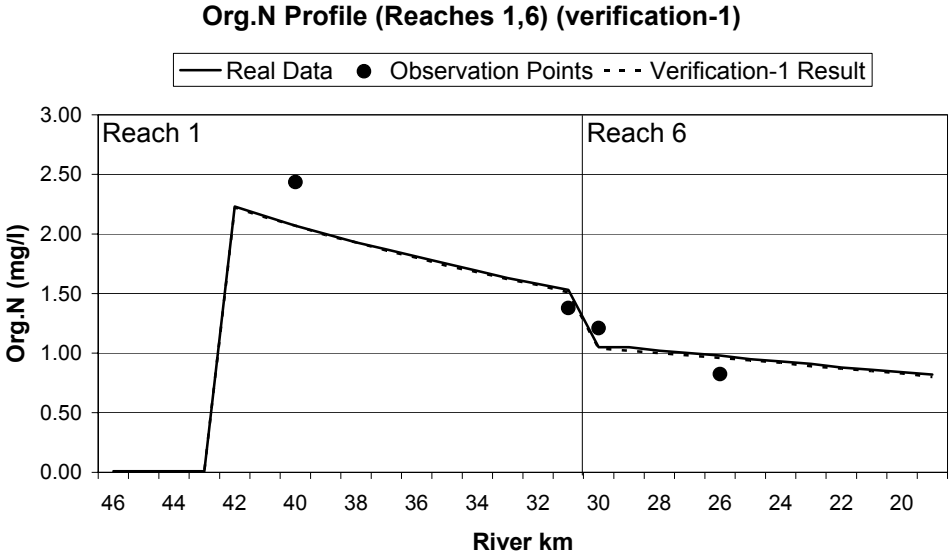

Figure C.16: Organic N profile along the main river for Complex Problem III for the first verification conditions. (Biased observation data)
(Corresponds to Figure 4.63)

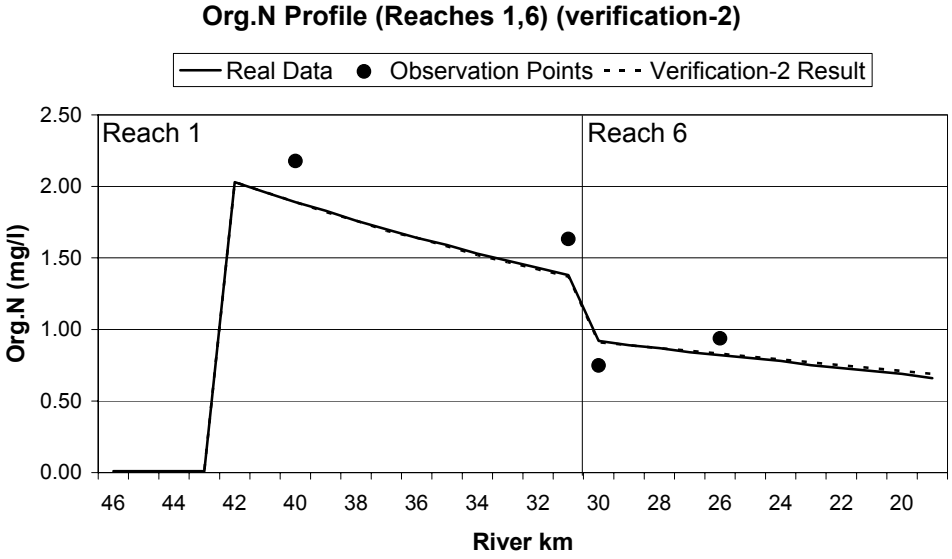

Figure C.17: Organic N profile along the main river for Complex Problem III for the second verification conditions. (Biased observation data)
(Corresponds to Figure 4.63)

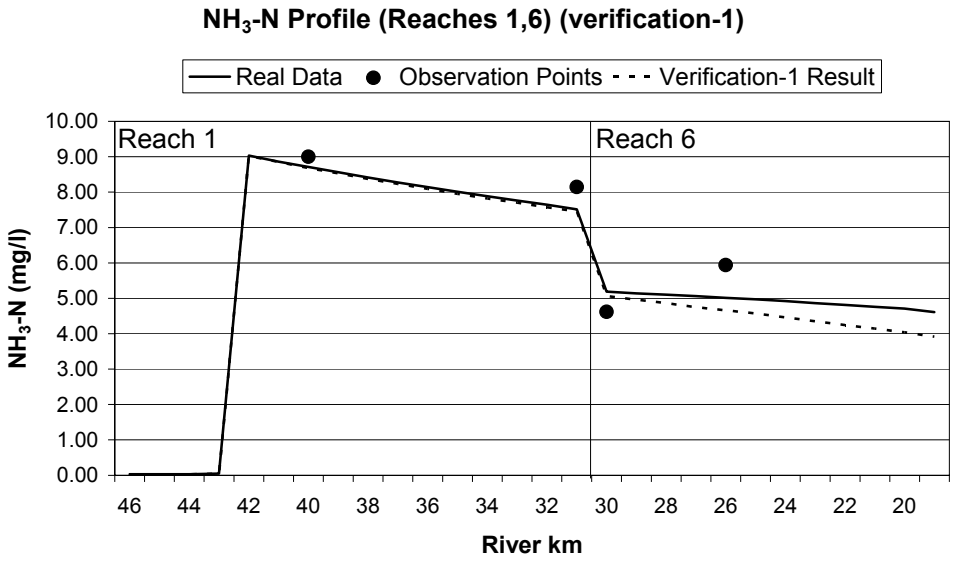

Figure C.18: NH_3-N profile along the main river for Complex Problem III for the first verification conditions. (Biased observation data) **(Corresponds to Figure 4.65)**

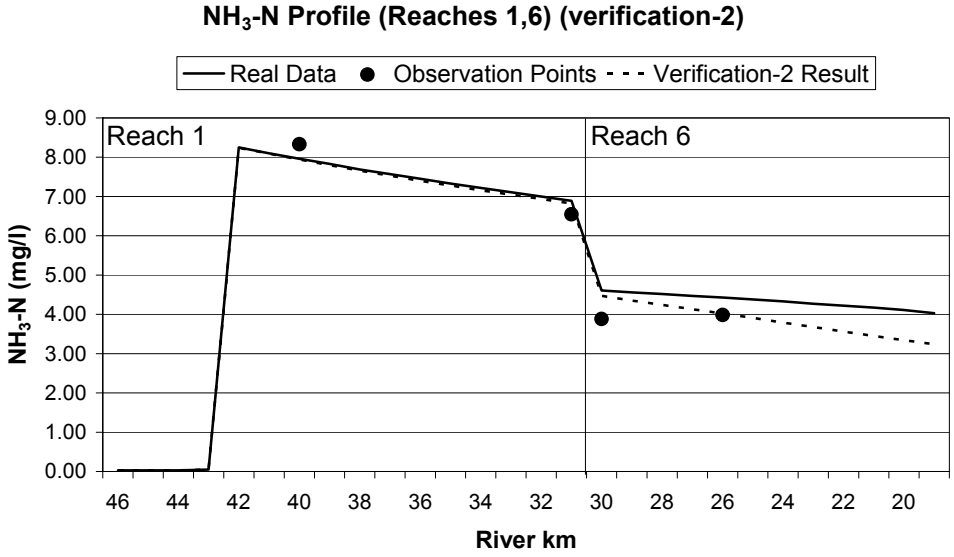

Figure C.19: NH_3-N profile along the main river for Complex Problem III for the second verification conditions. (Biased observation data) **(Corresponds to Figure 4.65)**

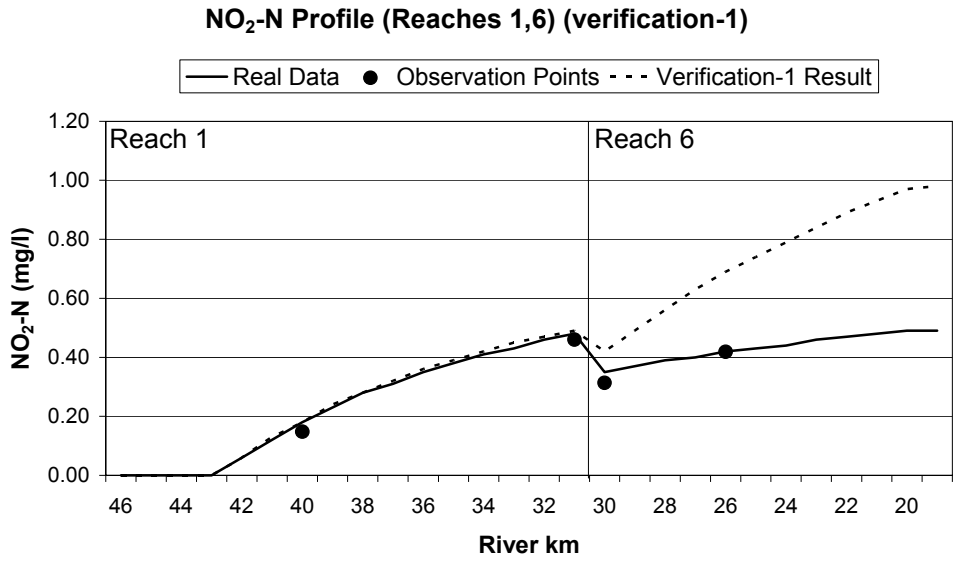

Figure C.20: NO_2-N profile along the main river for Complex Problem III for the first verification conditions. (Biased observation data) **(Corresponds to Figure 4.67)**

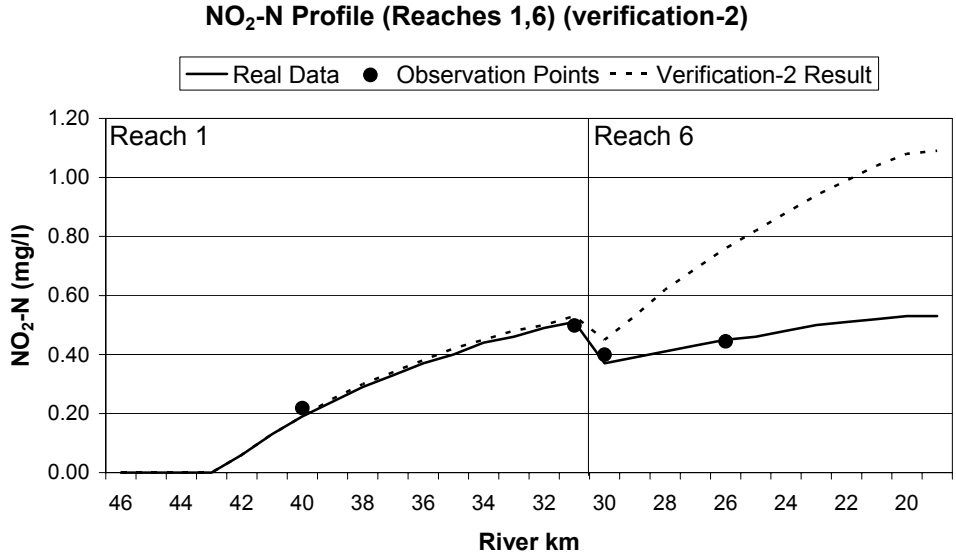

Figure C.21: NO_2-N profile along the main river for Complex Problem III for the second verification conditions. (Biased observation data) **(Corresponds to Figure 4.67)**

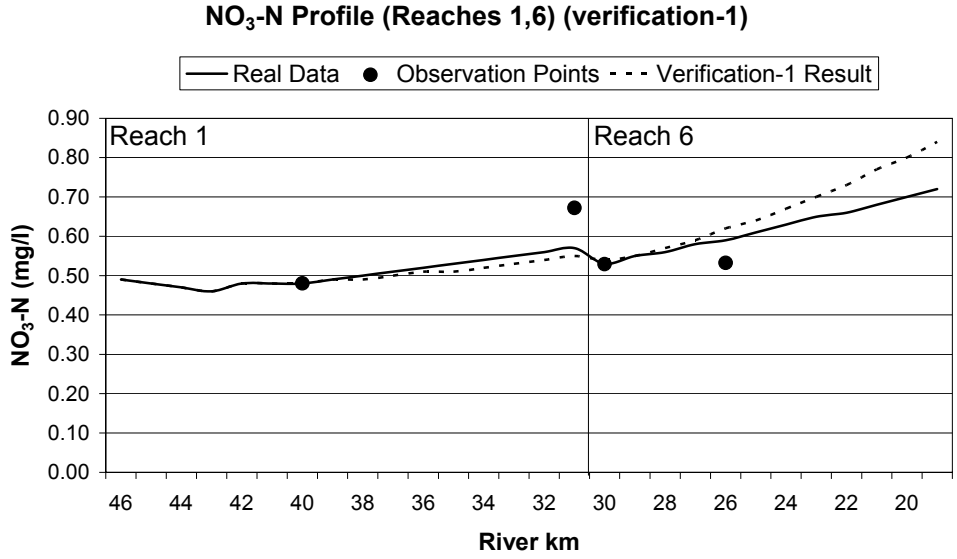

Figure C.22: NO_3-N profile along the main river for Complex Problem III for the first verification conditions. (Biased observation data) **(Corresponds to Figure 4.69)**

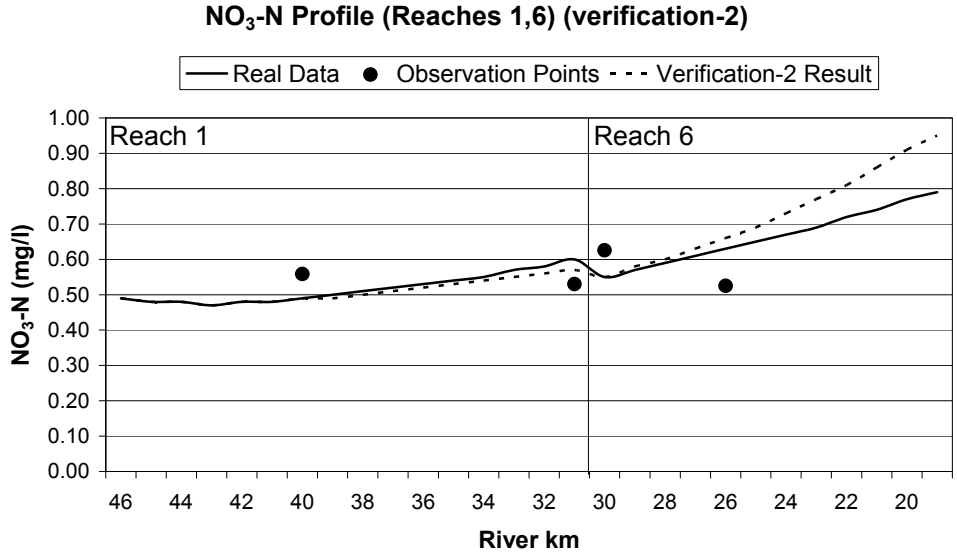

Figure C.23: NO_3-N profile along the main river for Complex Problem III for the second verification conditions. (Biased observation data) **(Corresponds to Figure 4.69)**

APPENDIX D

ASSIGNED WEIGHT VALUES IN THE COMPLEX PROBLEMS AND THE CASE STUDY

D.1 Assigned Weights in the Complex Problems

Table D.1: The values of the weights assigned to the error functions of the utilized water quality constituents in Complex Problem I.

Water Quality Constituent	The Assigned Weight	
	Perfect Data Case	Biased Data Case
DO	0.00000000986	0.00000000749
BOD	0.00000000186	0.00000000132

Table D.2: The values of the weights assigned to the error functions of the utilized water quality constituents in Complex Problem II.

Water Quality Constituent	The Assigned Weight	
	Perfect Data Case	Biased Data Case
DO	0.00000001154	0.00000001127
BOD	0.00000000178	0.00000000126
Org.N	0.00142826537	0.00079617834
NH_3-N	0.00000004478	0.00000003848
NO_2-N	0.00000007730	0.00000007706
NO_3-N	0.00000053611	0.00000058758

Table D.3: The values of the weights assigned to the error functions of the utilized water quality constituents in Complex Problem III.

Water Quality Constituent	The Assigned Weight	
	Perfect Data Case	Biased Data Case
DO	0.00000001154	0.00000001127
BOD	0.00000000178	0.00000000111
Org.N	0.00142826537	0.00079617834
NH_3-N	0.00000004444	0.00000003848
NO_2-N	0.00000007714	0.00000007706
NO_3-N	0.00000048964	0.00000050687

D.2 Assigned Weights in the Case Study

Table D.4: The values of the weights assigned to the error functions of the utilized water quality constituents in Case Study.

Water Quality Constituent	The Assigned Weight	
	Dry Period	Wet Period
DO	0.00000003034	0.00000989805
BOD	0.00000000694	0.00000954745
Org.N	0.00005153045	0.00318035811
NH_3-N	0.00004771448	0.00607496507
NO_2-N	0.00021761838	1.07680876953
NO_3-N	0.00004026900	0.03268614761
Dis.P	0.46436034363	59.6196267811