Mississippi State University

Scholars Junction

1-1-2005

Parameter estimation and auto-calibration of the STREAM-C model

Sumit Sinha

Follow this and additional works at: https://scholarsjunction.msstate.edu/td

PARAMETER ESTIMATION AND AUTO-CALIBRATION OF THE STREAM-C

MODEL

By

Sumit Sinha

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Civil Engineering
in the Department of Civil Engineering

Mississippi State, Mississippi

May 2005

PARAMETER ESTIMATION AND AUTO-CALIBRATION OF THE STREAM-C
MODEL

By

Sumit Sinha

Approved:

_____   _____
Dr. James L. Martin   Dr. William H. McAnally
Professor of Civil Engineering   Associate Professor of Civil Engineering
(Committee Chairman)   (Committee Member)

_____   _____
Dr. Christopher D. Eamon   Dr. Thomas D. White
Assistant Professor of Civil Engineering Professor and Head of Civil Engineering
(Committee Member)   (Graduate Coordinator)

                    _____
                    Dr. Kirk H. Schulz
                    Dean of the College of Engineering

Name: Sumit Sinha

Date of Degree: May 7, 2005

Institution: Mississippi State University

Major Field: Civil Engineering

Major Professor: Dr. James L. Martin

Title of Study: PARAMETER ESTIMATION AND AUTO-CALIBRATION OF THE STREAM-C MODEL

Pages in Study: 92

Candidate for Degree of Master of Science

The STREAM-C model is based on the same algorithm as implemented by the Steady Riverine Environmental Assessment Model (STREAM), a mathematical model for the dissolved oxygen (DO) distribution in freshwater streams used by Mississippi Department of Environmental Quality (MDEQ). Typically the water quality models are calibrated manually. In some cases where some objective criterion can be identified to quantify a successful calibration, an auto calibration may be preferable to the manual calibration approach. The auto calibration may be particularly applicable to relatively simple analytical models such as the steady-state STREAM-C model. Various techniques of parameter estimation were identified for the model. The model was then subjected to various techniques of parameter estimation identified and/or developed. The parameter estimates obtained by different techniques were tabulated and compared. A final recommendation regarding a preferable parameter estimation technique leading to auto calibration of the STREAM-C model was made.

DEDICATION

I would like to dedicate this work to my parents, Shobha and Shiva Kumar Sinha, and my brother Shishir. Their love, and encouragement, made the completion of this project possible. I thank God for having them in my life.

ACKNOWLEDGEMENTS

The author would like to acknowledge those who assisted in the completion of this research. I would like to thank Dr. James L. Martin, my committee chairman, for his guidance, commitment, and support; not only in this project, but also throughout the pursuit of my Masters Degree. I would also like to acknowledge the members of my thesis committee, Dr. William H. McAnally and Dr. Christopher D. Eamon, for their direction and constructive criticism throughout this research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Computer models are often used to predict the behavior of environmental systems. These models vary in complexity from a simplified representation of the dominant processes using a small number of analytical equations, to detailed mathematical descriptions of the many interactive operative processes, the solution of which can only be obtained through complex numerical manipulations. When a model is developed for representing processes at certain specific locations it is imperative to calibrate the model using historical data or collected field data. Calibration is the process of adjusting the model parameters so that the discrepancy between the model prediction and the field data is minimized. Once the model has been calibrated it can then become a device for environmental analysis in which physical/chemical properties or other significant descriptors of environmental processes operative at a particular study area are inferred. Model-predicted environmental behavior is often used as a basis for environmental management. The process of calibrating a model can be a daunting task and the effort required to do so increases as the number of uncertain parameters in the model increases. The nonlinearity of the model is another factor which makes the process of calibration a time consuming and a tedious affair. A major problem in using the computer model which has been calibrated manually is that the parameter estimate

obtained in such a manner is that the parameter estimate obtained in such a manner is fraught with uncertainty. The uncertainty in the parameter also leads to the uncertainty in the model prediction. Much has been written on the subject over a long period of time (e.g. Beck (1987), Beven (1993), Jakeman and Hornberger (1993), Hill (1998)) to name just a few.

In some cases where some objective criterion can be identified to quantify a successful calibration, an auto calibration may be preferable to the manual calibration approach. The auto calibration approach may be particularly applicable to relatively simple analytical models such as the steady-state STREAM model routinely used by the Mississippi Department of Environmental Quality (MDEQ) for the purpose of waste load allocation and determination of Total Maximum Daily Loads (TMDLS).

Objectives

The objective of this work is as follows.

a) To test and develop techniques of parameter estimation which will lead to auto calibration of the STREAM (Steady Riverine Environmental Assessment Model) model. This is dissolved oxygen -sag model used by the M.D.E.Q.

b) To identify the suitable method or technique which will be apt for carrying out the process of parameter estimation of the STREAM model and which will finally calibrate the model.

Preview

To accomplish the above objectives following steps were taken.

a)   The present structure of STREAM is based on Java algorithms implemented in an Oracle data base. This structure was prohibitive in the usage of lot of parameter estimation technique hence the whole program was rewritten using the C programming language which resulted in STREAM-C, which was tested against the original code. Once the best parameter estimation technique for the STREAM-C is identified, it is to be incorporated in the MDEQ STREAM model.

b)   A detailed description of the STREAM-C model is presented in chapter 2, where in the hydraulic and water quality constituents of the model are presented.

c)   A detailed literature review was then undertaken which would lead us to identify the suitable methods and/or software which could then be implemented for the purpose of parameter estimation of the model was described in Chapter 3.

d)   Selected methods and/or software were then implemented on a number of test cases made and conclusions were drawn, which are presented in chapter 4.

e)   Conclusions against the initial objectives and results drawn by this work are discussed in Chapter 5.

CHAPTER II

STREAM-C MODEL AND PARAMETERS

The STREAM-C model is based on the Steady Riverine Environmental Assessment Model (STREAM), a mathematical model for the dissolved oxygen (DO) distribution in freshwater streams used by MDEQ; hence this discussion is valid for both STREAM-C and the STREAM model. STREAM is a steady state model that utilizes a modified Streeter Phelps DO sag equation. The model is one dimensional, assuming that the channel is well mixed in vertical and lateral directions. The stream processes simulated by the model are ultimate carbon biochemical oxygen demand (CBODu) decay, Nitrification and dissolved oxygen deficit change. Reaction rates for the above stream processes are input by the user and corrected for the temperature. The hydrological processes simulated by the model include stream velocity and flow from point sources and spatially distributed inputs. The model output includes water quality conditions in each computational element for DO deficit, CBODu and ammonia nitrogen (NH3-N) concentrations. The STREAM-C model incorporates an option to perform parameter estimation through Monte Carlo sampling subjected to an objective function which finally gives us the best parameter estimate for a reach see chapter 4.

Hydraulic and Water Quality Characteristics

A detailed description of the formulas used for defining the hydraulic and water quality characteristics is discussed below.

Segmentation

The first step in using the STREAM model is to determine and specify the model segmentation, or computational representation of the river or stream. To simulate a river or stream system along with a tributary would require specification of two water bodies. The most upstream point of each water body is called the headwater element. The initial flow and concentrations of water quality constituents must be specified for each headwater element.

The second step in the segmentation is dividing the water body into reaches and each water body will consist of one or more reaches. Reaches represent the segments of water body having similar characteristics. A water body may be divided into reaches such that the model parameters, including velocity, depth, slope, reaction rates, and sediment oxygen demand are relatively consistent within each reach. Flows and concentrations from upstream sources (such as a tributary) or point sources inflows (such as waste sources) can only be specified at the origin (most upstream point) of a reach. For example, in simulation of the two water bodies, a river and tributary, the water body representing the river would be subdivided into a minimum of two reaches, one upstream

and one downstream of the tributary inflow. The river (and/or tributary) could be further subdivided based on the following characteristics.

a)   Additional inflow from point or non-point sources.

b)   Differences in hydraulic characteristics (e.g. depths and velocities).

c)   Differences in reaction rates, which includes sediment oxygen demand (SOD), rates of productivity and respiration, and reaction rates for CBODu and ammonia. These rates may vary between reaches but for the modeling purpose are assumed to be constant within a reach.

The flows and concentrations at the most upstream point of a contiguous reach are computed using a mass balance, based upon values for external inflows and the end condition for upstream reach(es).

The final segmentation step is to further subdivide each reach into computational elements which are based on a computational length specified by the user as a part of the input file. The model then subdivides each reach into an integer number of computational elements of equal length. The length of computational elements is consistent throughout the model for all reaches.

In the model STREAM-C, flow and water quality constituents are calculated at the end of each computational element, using predictions from the previous condition as the boundary condition. Parameters and other rates specified at the origin of the reach apply to the entire reach. Inputs of tributaries, point sources and spatially distributed sources begin at the origin river mile in a reach. Thus, reach divisions must occur at the confluence of any of these inputs. If a branching system is modeled, the program requires

that the terminus river mile is zero for a reach that flows into another water body. Please see Appendix B for a typical input file.

Flow

The model calculates a steady state flow in each computational element based on flow input (headwater, point and non-point source). Headwater flows are assigned for each headwater element (the most upstream point in each modeled water body). Point sources of flow are added to a particular origin river mile of a reach. Distributed or nonpoint sources are distributed evenly among the computational elements in a reach.

Reach Velocity

The velocity in the STREAM-C model can be input in two different ways. The first being simply specifying the value of velocity (in fps) in the input file for each reach, according to the specified format for the model. The model developed can also calculate the reach velocity using an exponential equation for which the user must specify the coefficient and exponents, as in:

$$U = C_u \times Q^{N_u},$$ (2.1)

where:

U   =   reach velocity (fps),

Cu   =   Velocity coefficient, and

Nu   =   Velocity exponent.

Reach Depth

Depths for each reach can be also input by the user in two different ways. The first method is to specify the depth in each reach which is taken as constant over the entire reach. The STREAM-C model can also use an exponential equation similar to that for velocity for the computation of depth within a reach, as given by:

$$D = C_h \times Q^{N_h},$$

(2.2)

where:

D $\quad=\quad$ depth (ft),

C_h $\quad=\quad$ depth coefficient, and

N_h $\quad=\quad$ depth exponent

Reaeration

Reaeration is the process of oxygen exchange between the atmosphere and a water body in contact with the atmosphere. There are many empirical formulas available for estimating the reaeration rate. The STREAM-C model uses two of these empirical formulas, Tsivoglou and O'Connor-Dobbins formulas. The user also has an option of specifying the value of the reaeration rate based on field measurements.

The Tsivolglou formulation is given as follows:

$$K_a = C \times S \times U$$

(2.3)

where

Ka $\quad=\quad$ reaeration rate (day -1),

C   =   escape coefficient,

S   =   slope (ft/mile), and

U   =   velocity (fps)

The Tsivoglu formulation is appropriate for use in small water bodies. The O'Connor-Dobbins formulas is recommended for water bodies with depth greater than 5 ft and where the depth profiles are available (see Brown and Barnwell 1985). The governing formula is as follows:

$$K_a = \frac{12.9 \times \sqrt{U}}{\left(\sqrt{D}\right)^3}$$

(2.4)

where

Ka   =   reaeration rate (day-1),

U   =   velocity (fps), and

D = depth (ft)

It's worth mentioning that of all the available options for specifying velocity, depth and reaeration a single method or any combination can be used in the model.

Water Quality Constituents

The STREAM-C model is capable of simulating the following water quality constituents: DO DO-deficit, CBODu and NH3-N. Under the conditions when the predicted DO is greater than zero, the DO is equal to the saturated dissolved oxygen concentration minus the estimated dissolved oxygen deficit that can be presented by the following equation:

$$DO = DO_{sat} - D_i$$

(2.5)

where:

DOsat = DO saturation (mg/l), and

Di   = DO deficit (mg/l).

DO saturation for a particular reach is dependent on the average temperature of the reach which is specified as the part of the input file of the model. The model uses a modified version of the Elmore and Hayes equation for the saturation concentration:

$$DO_{sat} = 14.652 + (-0.41022 + (0.007991 - 0.000077774 \times T) \times T) \times T + 0.2$$

(2.6)

where:

DOsat = DO saturation (mg/l), and

T   = Temperature (0 C)

The model calculates the concentration of CBODu in each computational element based on the headwater concentration combined with point and distributed inputs of CBODu. The first order decay rate Kr is used to simulate the removal of CBODu from the water body through decomposition and settling. The equation is as follows:

$$L_i = L_{i-1} \times \exp\left(-K_r \times \left(\frac{\Delta x}{U}\right)\right)$$

(2.7)

where:

Li   =   CBODu at the end of the computational element (mg/l),

Li-1   =   initial CBODu concentration (from previous computation),

Kr   =   overall CBOD removal rate (day -1),

Δx =   computational element distance (mile), and

U   =   reach velocity (mile day-1)

The model calculates the concentration of NH3-N in each computation element based on the headwater concentrations combined with point and distributed inputs of NH3-N A first order rate (KN) is used to simulate nitrification, the oxidation of NH3-N to nitrate nitrogen (NO3-N). It is assumed that each gram of NH3-N oxidized to NO3-N uses 4.57 grams of oxygen. The relevant equation is as follows:

$$N_i = N_{i-1} \times \exp\left(-K_N\left(\frac{\Delta x}{U}\right)\right)$$

(2.8)

where:

Ni $\quad=\quad$ NH3-N concentration (mg/l) at the end of element i,

Ni-1 $\quad=\quad$ initial NH3-N concentration (from previous computation),

KN $\quad=\quad$ nitrification rate (day-1),

Δx $\quad=\quad$ computational element distance (mile), and

U $\quad=\quad$ reach velocity (mile day-1)

The most important water quality constituent (from a regulatory prespective) is the DO deficit. The DO deficit in each computational element is calculated using a modified Streeter-Phelps equation. It's worth noting that reaeration is the major source of DO in the flowing water, other sources include algal photosynthesis and the oxygen contained in incoming flow from point and distributed sources. DO sinks include CBODu decay, nitrification, algal respiration and sediment oxygen demand (SOD). In STREAM-C the DO deficit in water body is computed using the following equation:

$$
\begin{aligned}
D = &\; D_o \exp\left(-K_a\frac{\Delta x}{U}\right) + \left\{\frac{K_d}{K_a-K_r}\left[\exp\left(-K_r\frac{\Delta x}{U}\right) - \exp\left(-K_a\frac{\Delta x}{U}\right)\right]\right\}L_0 \\
&+ \left\{\frac{K_N}{K_a-K_N}\left[\exp\left(-K_N\frac{\Delta x}{U}\right) - \exp\left(-K_a\frac{\Delta x}{U}\right)\right]\right\}N_o\,4.57 \\
&- \left[1-\exp\left(-K_a\frac{\Delta x}{U}\right)\right]\frac{P}{K_a} \\
&+ \left[1-\exp\left(-K_a\frac{\Delta x}{U}\right)\right]\frac{R}{K_a} \\
&+ \left[1-\exp\left(-K_a\frac{\Delta x}{U}\right)\right]\frac{S}{K_a}
\end{aligned}
$$

(2.9)

where:

Di-1	=	Do (mg/L),
Ka	=	reaeration rate (day-1),
Δx	=	computational element distance (mile),
U	=	reach velocity (mile day-1),
K_d	=	CBOD decay rate (day -1),
K_r	=	overall CBOD removal rate (day-1),
Li-1	=	Lo (mg/L),
K_N	=	nitrification rate (day-1),
Ni-1	=	No (mg/L),
P	=	photosynthesis rate (mg/l/day),
R	=	respiration rate (mg/l/day),
S	=	SOD rate (mg/l/day),

As a general rule the rates of reactions in natural water bodies increases with increasing temperatures. As a result of this all the reaction rates that are input to the model are at standard temperature of 200C and are then adjusted for temperature by the following formula:

$$k(T) = k(20) \times \theta^{T-20}$$
(2.10)

where:

k	=	first order reaction rate (day-1),
T	=	temperature in degrees Celsius, and
θ	=	temperature coefficient

The value of temperature coefficient varies for each reaction rate. Temperature coefficients are basically assigned based on the literature values. In this model the user doesn't have the choice of changing the temperature coefficient and all the coefficients are hard coded. There values are as follows.

Table 2-1

Temperature Coefficient for Various Reaction Rate

REACTION RATE	TEMPERATURE COEFFICIENT
Reaeration Rate (Ka)	1.024
Overall CBOD Removal (Kr)	1.04
CBOD Decay (Kd)	1.04
NH3-N Decay (KN)	1.04
Photosynthesis, Respiration and SOD	1.08

Key Parameters of the Model

There are three equations that form the basis of the STREAM-C model, which forms the basis for predicting the DO deficit, NH3-N and CBODu (concentrations).The three equations are presented in 2.7, 2.8 and 2.9. There are three essential unknown parameters for the model. The first is K_d the carbonaceous deoxygenation rate. This parameter defines the rate of oxidation of carbonaceous organic material ($CBOD_u$) in a water body. The model's second important parameter is K_r, which is overall CBOD removal rate; this is composed of both oxidation and settling of CBOD. When K_r and K_d are assumed to be equal settling is assumed to be negligible. The last but equally

important parameter is K_N, the nitrification rate which defines the rate at which the NH3-N oxidizes to nitrate nitrogen (NO3-N). The parameter Kr and Kd are connected by the following relationship.

$$K_r = K_d + \frac{V_s}{H}$$

(2.11)

where

      Kr   =   Overall CBOD removal rate,

      Kd   =   CBOD decomposition rate,

      Vs   =   settling velocity for CBOD, and

      H   =   average depth

The three important parameters just mentioned are specified to the model via the input file. The values of these parameters are usually taken from the standard literature or by manual calibration. But, then these parameters might not give us the prediction which tallies with the field data or observed data. Alternatively which brings us back to the central issue of this research the parameter estimation may be accomplished using some kind of auto calibration technique, if some information about the parameter distribution is known. From all the input files that were obtained courtesy of MDEQ, and which were used in the various simulation and test cases, the average initial parameter values varied between 0.2 and 0.4 and coefficient of variation varied from 0.15 to 0.20 for all the three parameters.

CHAPTER III

TECHNIQUES OF PARAMETER ESTIMATION

Any mechanistic water quality model or for that purpose any model can be mathematically represented by the following mathematical equation:-

$$y = f(p;x)$$

where x represents the structural aspects of the model, as well as its input dataset, which are assumed to be fixed for the present discussion and not adjustable for the process of calibration. If we take our fundamental equation the dissolved oxygen deficit equation:

$$D = D_o \exp\left(-K_a \frac{x}{U}\right) + \left\{\frac{K_d}{K_a - K_r}\left[\exp\left(-K_r \frac{x}{U}\right) - \exp\left(-K_a \frac{x}{U}\right)\right]\right\}L_0$$

$$+ \left\{\frac{K_N}{K_a - K_N}\left[\exp\left(-K_N \frac{x}{U}\right) - \exp\left(-K_a \frac{x}{U}\right)\right]\right\}N_o 4.57$$

$$- \left[1 - \exp\left(-K_a \frac{x}{U}\right)\right]\frac{P}{K_a}$$

$$+ \left[1 - \exp\left(-K_a \frac{x}{U}\right)\right]\frac{R}{K_a}$$

$$+ \left[1 - \exp\left(-K_a \frac{x}{U}\right)\right]\frac{S}{K_a}$$

$$(3.1)$$

All the other variables apart from the three fundamental parameters i.e. Kr, Kd and KN will form structural aspect of the model and we don't adjust them or change then while trying to calibrate the model.

The problem of parameter estimation is encountered by people in all disciplines and is the subject of extensive ongoing research (e.g., Esposito and Floudas 2000, Nazin and Polyak 2003, Dilks, et al. 1992; etc).Over the last couple of decades with exponential advancement in computing power and with new ideas emerging in the fields of statistical science, different techniques of parameter estimation have emerged. The choice of technique for parameter estimation often varies with the:

a)   Numbers of parameters that you are trying to estimate, and

b)   Degree of non-linearity of the model.

It's not uncommon to have a water quality model in which it is required to estimate as high as ten or more parameters. There are basically two broad categories for conducting the parameter estimation process:

a)   Stochastic Techniques, and

b)   Deterministic Techniques.

Explanations of both of the techniques are presented here. The fundamental difference between the techniques is that the deterministic technique doesn't use any kind of probability measure to make relevant decisions. Stochastic technique uses some probability rules and measures to reach to valid results.

STOCHASTIC METHOD

The stochastic method for parameter estimation is by and large based on the Bayesian concepts (e.g. Qian and Stow 2002, Borsuk and Stow 2000, Hakanson 2000 etc). Bayesian methods are currently experiencing an increase in popularity in science and engineering as a means of probabilistic inference (Malakoff 1999). The common advantages of the Bayesian method can be listed as follows.

 a) Ability to include prior information.

 b) Ease of incorporation into a formal decision analytic context.

 c) Explicit handling of uncertainty.

The Bayesian approach has been applied to lot of ecological models where there is little information about the parameter values (e.g Reichert and Omlin 1997).The fundamental equation in Bayesian inference is Bayes' Theorem, discovered by English cleric Thomas Bayes and published posthumously. It was rediscovered and systematically exploited by later by Laplace. The basic Bayesian theorem as applied to the modeling framework is discussed below.

Suppose we have to estimate the values of p unknown model parameters: $\theta : \left(\theta_{1,...,}\theta_{p} \right)$ about which there may be some prior beliefs. These prior beliefs can be expressed as a probability density function $\pi(\theta)$, and may be interpreted as the probability placed on all possible parameter values before collecting any new data. Now let's say that we collect some field data described by D = (d1… dm). These observations will differ from the model predictions for any given set of model parameters θ and we

can write the probability of obtaining them as $L(D\,|\,\theta)d\theta$ (e g. Harmon and Challenor

2000). Now onwards we will drop the term $d\theta$ and only work with the probability

density function $L(D\,|\,\theta)$. This probability density function is referred as likelihood

function and is used to update the prior distribution of θ to account for the new data, D.

This updating is performed using the Bayes' theorem which can be expressed as follows

(e.g. Stow, Borsuk et al. 2002):

$$\pi\left(\frac{\theta}{D}\right) = \frac{\pi(\theta)L\left(\dfrac{D}{\theta}\right)}{\int \pi(\theta)L\left(\dfrac{D}{\theta}\right)d\theta}$$

(3.2)

Where $\pi\left(\dfrac{\theta}{D}\right)$ is called the posterior parameter distribution and expresses the

probability of the parameter values after observing the new data. Because the

denominator of the above equation is a normalizing constant, Bayes' theorem is often

expressed as (e.g. Stow, Borsuk et al. 2002):

$$\pi\left(\frac{\theta}{D}\right) \propto \pi(\theta)L\left(\frac{D}{\theta}\right)$$

(3.3)

The above equation states that the posterior distribution is affected or modified by

the likelihood function and the prior distribution of the parameters (e.g. Stow, Borsuk et

al. 2002). A major difference between the Bayesian methods and deterministic methods

of parameter estimation like least squares (ordinary and nonlinear), or maximum

likelihood is that the inference using Bayes theorem is typically made over the whole

support of $\pi\left(\dfrac{\theta}{D}\right)$, not just a single value of θ.

The posterior distribution calculated by the Bayesian technique can be exploited to obtain a number of different characteristics of the parameters, such as the marginal distributions or means and variances of the individual θ_i. But, any integration over the posterior distribution (for example to calculate it's mean) involves integrating a complicated function over multidimensional space and would be prohibitively expensive in computer time (e.g. Harmon and Challenor 2000). Fortunately there are techniques to deal with this problem which saves lots of our computational time. A few of them are as follows

a) Bayesian Monte Carlo technique (see Dilks, Canale et al. 1992),

b) Metropolis Hasting Algorithm (Metropolis et al.. 1953) as described by Clifford (1994), and

c) Shuffled Complex Evolution Metropolis algorithm (Vrugt, Gupta et al. 2002).

As mentioned before the Bayesian technique has it's limitations and in some problems the Bayesian approach has been shown to lead to a very different conclusion than that fom a classical approach see (Ludwig 1996, Al-Khatib et al. 2001).Monte Carlo sampling is another technique that is used to exploit the posterior distribution obtained by the Bayesian technique which has been widely used in the field of ecological modeling for uncertainty analysis(e.g. Van Horssen, et al. 2002; Carroll and Warwick, 2001;

Hakanson, 2000; Philips and Marks, 1996; Yool, 1999; Annan, 1997), and parameter estimation see(Dilks et al. 1992; Gertner, et al. 1999).

A detailed discussion on one of the algorithm which exploits the posterior distribution obtained by the Bayesian technique for finding other relevant parameter characteristics is presented below.

Markov Chain Monte Carlo Methods

This method is also known by the name of Metropolis-Hastings algorithm; the basic step which forms the algorithm is presented below. Choose a starting parameter vector say θ for the model for which you are implementing the Metropolis-Hastings algorithm, and then follow the steps listed to upgrade the parameter vector so that the model prediction converges to the field data. The number of iteration for which we need to carry out the listed steps would differ from one model to the other and will also be dependent on other things like quantity of the observed data present.

Repeat for a set number of iterations which will lead to a stationary Markov Chain:

i) Generate a new vector $\theta^x = \theta + \Delta\theta$ where $\Delta\theta$ is drawn from specified distribution Q.

ii)If $\left[\dfrac{\pi\left(\dfrac{\theta^x}{D}\right)}{\pi\left(\dfrac{\theta}{D}\right)}\right] > U$ where U is a random variable from a uniform distribution on the interval (0,1), then accept θ^x as the next step in the chain.

iii) Else use θ as the next step in the chain.

iv) Endif

To implement the above algorithm, posterior distribution needs to be computed. Posterior distribution for the parameter vector can be computed using Bayesian technique as explained above. To compute the ratio of posterior distribution in the Metropolis Hastings algorithm, likelihood function and prior distribution function for the parameter space is required. The prior distribution function for the parameter vector is taken from the literature value and the method for calculating the likely hood function for the water quality models is presented.

Calculation of Likelihood

The likelihood function for the water quality model can be easily calculated as long as we can desribe the probability density function of the observed data (e.g. Dilks, Canale et al. 1992). The formula for the calculation of likelihood is based on the assumption that the errors in the data are normally and independently distributed with a mean zero. The error term is defined as (e.g. Dilks, Canale et al. 1992):

$$e_i = x_i - u_i$$

Where e, x and u are the error, field data and model predicted data at the point i. The likelihood function is given by the following formula (e.g. Dilks, Canale et al. 1992):

$$L\left(\frac{\theta}{D}\right) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{e_i}{\sigma}\right)^2\right]$$

(3.4)

Where n is the number of observed data points and σ the standard deviation for the data error. The equation for the computation of likelihood as can be seen is not only the function of the standard deviation of the data error but is also a function of the number of observed data points. As the standard deviation decreases, the likelihood value decreases sharply as the relative error becomes non zero. This is consistent with the benefit of the Bayesian inference, in that data with greater certainty have a stronger impact on the posterior distribution than highly uncertain data. (e.g. Dilks, Canale et al. 1992)

DETERMINISTIC METHOD

As opposed to stochastic technique, the deterministic technique doesn't consider any kind of data to be random It also doesn't rely on any kind of assumed probability distribution to make decisions. The common deterministic technique is by and large based on upgrading or modifying the parameter vector space to an extent where it gives optima of the desired objective function which is related to the model output. One such common tool which is gaining popularity in field of parameter estimation is PEST (e.g. http://www.sspa.com/pest/). PEST is a host of utility software which has been successfully applied for the purpose of parameter estimation in groundwater flow modeling (e.g. Meillier, Clark et al 2001) and is gradually finding its use as a parameter estimating tool in other kind of models (e.g. Skahill 2004).

The different kinds of deterministic techniques essentially differ in the kind of algorithm that they implement to reach to the optimum value of the objective function.

The problem of parameter estimation which different deterministic techniques try to solve can be restated as minimization of a cost function or an objective function that measures the goodness of the fit of the model with respect to a given experimental data set, subject to the dynamics of the system. More often than not this cost function or the objective function is sum of square of residual. that is the error between the observed and the predicted data Therefore the problem of the parameter estimation which our various deterministic techniques try to solve can be mathematically presented as follows (e.g. http://www.sspa.com/PEST/pestman.pdf):

$$\phi = \sum_{i=1}^{m} \left(w_i r_i \right)^2$$

(3.5)

Where ϕ is our cost or the objective function, ri (the ith , residual) expresses the difference between the model outcome and the actual field or laboratory measurement for the i'th observation, wi is the weight assigned to the ith observation which is assigned in order to increase the contribution of the observation which the user might think should make a greater impact on the objective function (e.g. http://www.sspa.com/PEST/pestman.pdf).

There are various algorithms which are used by different deterministic techniques to optimize the objective function. Owning to the success of the application of PEST to various ground water flow models (e.g. Meillier, Clark et al 2001) and other kind of models (e.g. Skahill 2004), it was decided to use PEST for the purpose of parameter estimation and compare the results obtained by the software developed using a stochastic technique.

The algorithm implemented by PEST to reach to the minima of the objective function is a variant of gradient based Gauss-Marquardt-Levenberg (GML) technique (e.g. Levenberg 1944, Marqardt 1963).

All kinds of deterministic technique suffer from one serious drawback and that is running into the point of local minima, which might not be the global minima of the objective function the algorithm is trying to optimize. This problem is especially faced in a nonlinear model. In a deterministic technique we start with some initial value of parameter vector and then keep updating it until it reaches one of the terminating criteria which might be number of model runs or relative reduction in the objective function etc. In such a scenario it's possible that starting from some initial value of the parameter vector we might be able to reach to another parameter vector which stops the process of minimization of the objective function but might not be the set of parameter which yields the lowest value of the objective function. Hence using both stochastic and deterministic techniques can increase our confidence about the parameter estimated. Like all deterministic techniques the algorithm followed by PEST (the GML technique) also requires that the initial value with which we start the process of parameter estimation should be somewhat in the vicinity of the final parameter estimated. Other wise the process of parameter estimation is bound to run into the problem of local minima which has been discussed.

PEST Methodology

As explained in the preceding section PEST (e.g. www.sspa.com/PEST/pestman.pdf). software is based on the variant of a (GML) algorithm detailed explanation of the algorithm is presented here.

Let the relationships between the parameters and model generated observations for a particular model is represented by the function M which maps n-dimensional parameter space into m- dimensional observation space. We make an assumption that this function be continuously differentiable with respect to all the model parameters for which the estimates are sought. Suppose that for the set of parameters comprising of vector bo the corresponding set of model calculated observations (generated using M) is Co i.e. (e.g. http://www.sspa.com/PEST/pestman.pdf).

$$Co=M(bo). \tag{3.6}$$

Now to generate a set of observations c corresponding to a parameter vector b that differs only slightly from bo, Taylor's theorem tells us that the following relationship is approximately correct, the approximation improving with proximity of b to bo:

$$C=Co+J(b-b0) \tag{3.7}$$

where J is the Jacobian matrix of M, i.e. the matrix comprised of m rows (one for each observation), the n elements of each row being the derivatives of one particular observation with respect to each of the jth parameter. We are actually trying with the help of PEST derive a set of model parameters for which model generated observations are as close as possible to our set of field observations in the least square sense, i.e. we wish to

determine a parameter set for which the objective function ϕ defined by is minimum (e.g. http://www.sspa.com/PEST/pestman.pdf)..

$$\phi = (c - c_o - J(b - b_o))^t Q(c - c_o - J(b - b_o)) \tag{3.8}$$

where c in the above equation represents field observation vector, the vector c-co represents the discrepancy between the model-calculated observations co and their experimental counterparts c and b-b0 is the parameter upgrade vector which we are trying to figure out which will minimize our objective function (e.g. http://www.sspa.com/PEST/pestman.pdf).The formula for parameter upgrade vector which is used is as follows

$$u = (J^t Q J)^{-1} J^t Q(c - c_o) \tag{3.9}$$

Now the vector b obtained by adding the parameter upgrade vector u by the above equation to the current parameter values bo is not guaranteed to be that for which the objective function is at it's minimum. Hence the new set of parameters contained in b must be then used as a starting point in determining a further parameter upgrade vector, and so until, hopefully, we arrive at the global ϕ minimum. The last equation forms the basis of non linear weighted least squares parameter estimation. It can be rewritten as (e.g. http://www.sspa.com/PEST/pestman.pdf).:

$$u = (J^t Q J)^{-1} J^t Q r \tag{3.10}$$

where u is the parameter upgrade vector and r is the vector of residuals for the current parameter set. Let the gradient of the objective function ϕ in parameter space be denoted by the vector g. the i'th element of g is thus defined as

$$g_i = \frac{\partial \phi}{\partial b_i}$$ i.e. by the partial derivative of the objective function with respect to the ith

parameter. The parameter upgrade vector cannot be an angle of greater than 90 degrees to

the negative gradient vector. If the angle between u and –g is greater than 90 degrees, u

would have a component along the positive direction of the gradient vector and

movement along u would thus cause the objective function to rise, which is opposite of

what we want (e.g. Doherty J 2002, PEST manual).

Most parameter estimation problem benefit from adjusting u such that it is a little

closer to the direction of –g in the initial stages of the estimation process (see Doherty J

2002, PEST manual). Mathematically this can be achieved by including in the last

equation so called "Marquardt Parameter" named after Marquardt (1963), though the use

of this parameter was pioneered by Levenberg (1944). The equation gets modified to:

$$u = \left(J^t Q J + \alpha I\right)^{-1} J^t Q r \tag{3.11}$$

where α is the Marquardt parameter and I is the n x n identity matrix. Hence we

start with a initial parameter vector and then evaluate the objective function and upgrade

the parameter vector using the above formula and re evaluate the objective function and

keep on carrying the iteration until we don't see any more decrease in the objective

function or some other terminating criteria set up in the input file of PEST is met. The

current parameter set for the last iteration of the procedure forms the vector of best

parameter estimate.

Conclusion

As discussed in the preceding section there are two kinds of broad techniques for carrying out the process of parameter estimation but it's worth reiterating that both the techniques come along with their advantages and disadvantages. If the common disadvantage of the stochastic method is that it can take lot of computational time, the deterministic technique can run into the problem of local minima or the deterministic technique might not be applicable if the model function is not differentiable with respect of the parameters as shown in the case of PEST in the preceding section. For the present scope of work it was decided to apply both stochastic and deterministic techniques and then make final recommendation based on the results obtained from these two techniques.

CHAPTER IV

PARAMETER ESTIMATION OF STREAM-C

As discussed in the previous section, there are a number of techniques available for carrying out the process of parameter estimation for a given model. As also stated, the choice of the method is governed by diverse factors and a particular technique which might be a good choice for a given model might not serve the purpose in some other case. For the purpose of this study for parameter estimation for the model STREAM-C, it was decided that the model would be subjected to both deterministic and stochastic techniques. Both the techniques were applied to a number of test cases and valid conclusions and observations were made, which are presented here.

Stochastic Technique for STREAM-C

The program developed to implement the stochastic technique of parameter estimation for the STREAM-C model was based on the simple Monte-Carlo sampling technique coupled with an objective function. For the development of this program we merged both stochastic and deterministic techniques. The details for the algorithm are presented below. DO deficit equation on which the algorithm is based is presented below.

$$D = D_o \exp\left(-K_a \frac{\Delta x}{U}\right) + \left\{\frac{K_d}{K_a - K_r}\left[\exp\left(-K_r \frac{\Delta x}{U}\right) - \exp\left(-K_a \frac{\Delta x}{U}\right)\right]\right\} L_0$$

$$+ \left\{\frac{K_N}{K_a - K_N}\left[\exp\left(-K_N \frac{\Delta x}{U}\right) - \exp\left(-K_a \frac{\Delta x}{U}\right)\right]\right\} N_o \, 4.57$$

$$- \left[1 - \exp\left(-K_a \frac{\Delta x}{U}\right)\right]\frac{P}{K_a}$$

$$+ \left[1 - \exp\left(-K_a \frac{\Delta x}{U}\right)\right]\frac{R}{K_a}$$

$$+ \left[1 - \exp\left(-K_a \frac{\Delta x}{U}\right)\right]\frac{S}{K_a}$$

$$(4.1)$$

The parameters to be estimated, which have been stated before also, are:

Kr   =   Overall CBOD removal rate,

Kd   =   CBOD decomposition rate, and

KN   =   Nitrification rate.

The model allows these parameters to be varied by river reach, resulting in 3xn unknown parameters where n is the number of reaches. The algorithm implemented for the stochastic technique for the development of the software is as follows

a) The parameters were assumed to be normally distributed with a fixed mean and coefficient of variation, which were based on literature values.

b) Random Monte-Carlo sampling was conducted using the following formula

$$\phi^{-1}(\xi) = -t + \left(\frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right)$$

Where ξ is a random number generated between 0 and 1 and value of other constant are as follows C0=2.515517 C1 = 0.802853 C2 = 0.010328 d1 = 1.432788 d2 = 0.189269 d3=0.001308 t = (-ln (ξ) 2)1/2. The Gaussian random number thus obtained can be converted to a sample of a Gaussian variable of desired   mean and coefficient of variation by applying the following formula: $SampleVal. = y \times C.O.V \times Mean + Mean$ where y = $\phi^{-1}(\xi)$, if ξ > 0.5 take value of ξ = 1-ξ, compute the $\phi^{-1}(\xi)$ by the given formula and then reverse the sign and use it as y for linear transformation to a Gaussian variable of desired mean and coefficient of variation as given in the formula for the sample val .

c) Starting with specified value of mean and coefficient of variation for the normally distributed parameter samples were drawn from the Monte Carlo sampling technique as mentioned in step b. These values were plugged into the fundamental DO deficit equation (4.1) and model predictions made.

d)   The model predicted data was used for the evaluation of the objective function; the objective function used for this program was the sum of square of residual, which are the differences between the predicted and observed values. For the evaluation of the objective function, we used the field data which were obtained by the courtesy of MDEQ in two test cases and fabricated data in the rest of the test cases.

e)   The model was run for a user specified number of iterations and the parameters corresponding to the minimum objective function stored as the best model parameter set.

Deterministic Technique for STREAM-C

The popular parameter estimation tool PEST, based on a deterministic technique (e.g. Doherty J 2002, PEST manual), was selected as the software for carrying the process of parameter estimation largely because of it's successful implementation to hydrologic and ground water flow model (e.g. Meillier, Clark et al 2001) and also because of it's model independent framework. The model independent frame work can be explained as the ability of PEST to be used for the process of parameter estimation for just about any existing computer model, whether or not a user has the access to the model source code. PEST is able to "take control" of a model, running it as many times as it needs to while adjusting its parameters until the discrepancies between selected model outputs and a complementary set of field measurements is reduced to a minimum in the weighted least squares sense. The universal applicability of PEST lies in it's ability to

perform the task of parameter estimation for any model that reads it's input data from one or a number of ASCII (i.e. text) input files and writes the outcomes of it's calculation to one or more ASCII output files. Thus a model does not have to be recast as a subroutine and recompiled before it can be used within a parameter estimation process. PEST adapts to the model, the model does not need to adapt to PEST. For PEST to take control of an existing model in this fashion in order to optimize its parameters, certain conditions must be met (e.g. PEST manual www.sspa.com/PEST/pestman.pdf). These are as follows:

i) While a model may read many input files, some of which may be binary and some of which may be ASCII, the file containing, those parameters which PEST is required to adjust must be ASCII (i.e. text) files.

ii) While a model may write many output files, some of which may be binary and some of which may be ASCII, the file or files containing those model outcomes which contains field measurement must be ASCII (i.e. text) files.

iii) The model must be capable of being run using a system command and requiring no user intervention.

iv) PEST (as has been mentioned in chapter 3) uses a nonlinear estimation technique known as the Gauss-Marquardt Levenberg method. The basic strength of this method (e.g. Levenberg 1944 and Marquardt 1963) lies in the fact that it can generally estimate parameters using fewer model runs than any other estimation method, a definite bonus for large models whose run times may be considerable.

v) PEST requires that upper and lower bounds be supplied for the adjustable parameters (i.e. parameters which are neither fixed nor tied); this formulation is vital to PEST, for it informs PEST the range of permissible values that a parameter can take.

vi) The model for which we plan to use PEST must be continuous with respect to the parameters.

vii) PEST requires three types of input files. These are:

a) Template file, one for each model input file on which parameters are identified.

b) Instruction files, one for each model output file on which model generated observations are identified.

c) An input control file, supplying PEST with names of all template and instruction files, the names of corresponding model input and output files, the problem size, control variables, initial parameter values, measurement values and weights etc.

For the STREAM-C model in particular there was one template file, one instruction file and one input control file. As PEST executes, it writes a detailed record of the parameter estimation process to the file case.rec where case is the name of the test cases and .rec is the extension given by the PEST. As PEST prepares its parameter estimation record file, which is case.rec initially, it records all the input data set and the current value of the parameter and the value of the objective function. Then it starts the estimation process in earnest, beginning with the first optimization iteration. After

calculating the Jacobian matrix as explained in the chapter 3, PEST attempts objective function improvement using one or more Marquardt lambdas. As it does this it records the corresponding objective function value both in absolute terms and as a fraction of the objective function value at the commencement of the optimization iteration. After completing the parameter estimation process, PEST prints the outcomes of this process to the run record file. PEST also calculates 95% confidence limits for the adjustable parameters. The presentation of 95% confidence limits provides a useful means of comparing the certainty with which different parameter values are estimated. Note that at the end of a PEST optimization run a listing of the optimized parameter values can also be found in the PEST parameters value file in case.par.

Test Cases and Results Obtained

Test cases developed for both the deterministic and stochastic technique were based on input files obtained by the courtesy of MDEQ. Corresponding to some of the input files the observed or the field data were available In rest of the cases the initial parameter values were mutated and model predictions obtained using those changed parameter value. The model predictions obtained in such a manner were then used as the observed data and the parameter values were set back to the values from which they were changed. All test cases were subjected to both deterministic and probabilistic technique. It's worth mentioning that any software developed or existing which we decide to use for solving a particular problem must be used initially for the test cases for which we already know the answer before we use it on the real world data and draw our conclusion, and

that is what we intend do in the starting test cases before we come to the real world data. We subject the test case to both deterministic and stochastic technique and try to study the answer that we get from the existing and the developed software and try to see how far it is from the answers that we already know.

Sowashee Test Case Type 1 (Fabricated Observed Data)

The Sowashee creek is a twelve mile long creek which is broken down into four reaches for the modeling purpose. Presented here is the tabular description of the reach on which the process of parameter estimation was carried out. The coefficient of variation for Kr and Kd was kept as 0.2 and for Kn it was kept at 0.15 based on literature values.

Table 4-1

Description of the Sowashee Reach Type 1.

River Name	Origin Mile	End Mile	Kr	Kd	KN
Sowashee	11.5	4.3	0.300	0.300	0.300

The reach was also subjected to various loading conditions which are as presented in the table.

Table 4-2

Description of loads on the Sowashee Reach Type 1.

	Flow (Cfs)	DO (Mg/L)	CBOD(lbs/day)	TKN (lbs/day)
Upstream	1.370	5.291	61.51	7.68
Distributed Input	0.100	7.00	1.07	0.05
Waste Source	1.550	6.00	125.10	16.70

The input data file for this particular test case is set up in such a manner that the model gives the relevant prediction at every 0.1 mile in any reach that it processes. Hence, when we mutate the parameters to obtain data which will serve as the pseudo-field data, we have 73 possible predicted points to choose from in this particular reach. To begin with we selected 15 points randomly which were taken as the observed data. The extent to which parameters were mutated is shown in the following table.

Table 4 - 3

Initial and Mutated Parameters for Sowashee Reach Type 1

Parameters	Kr	Kd	KN
Initial	0.300	0.300	0.300
Mutated	0.330	0.330	0.325

Stochastic Results for Sowashee Reach Type 1

The mutated parameters were plugged into the model to obtain data which could serve as the pseudo-field data, parameters were then adjusted back to their initial value

and both stochastic and deterministic techniques applied to the model, the results of which are presented as follows.

Table 4 - 4

Stochastic Result for Showashee Reach Type 1

No. of Itrations.	Min. ObjFun.	Kr	Kd	KN
2000	0.000019	0.320411	0.320411	0.341625
4000	0.000022	0.322589	0.322589	0.337192
6000	0	0.328066	0.328066	0.328189
8000	0	0.330222	0.330222	0.324674
10000	0	0.330222	0.330222	0.324674

We can see that for a single reach the Monte-Carlo simulation technique gives fairly good approximations. The variation of various parameter values with increasing number of Monte Carlo iterations is graphed. It's worth reiterating that every time the model was run for specified number of iterations, from those random values of parameters generated by Monte Carlo sampling only those were taken which gave the minima of the objective function. We can see that with increase in the number of iterations the parameter values corresponding to the minimum objective function reaches a constant value. The line corresponding to the OptKr&Kd and OptKn in the plot presented shows the mutated value of the parameters which were used for producing the pseudo field data. It can be seen from the plot that after certain number of iterations parameter values converges to the mutated or optimum values as expected.
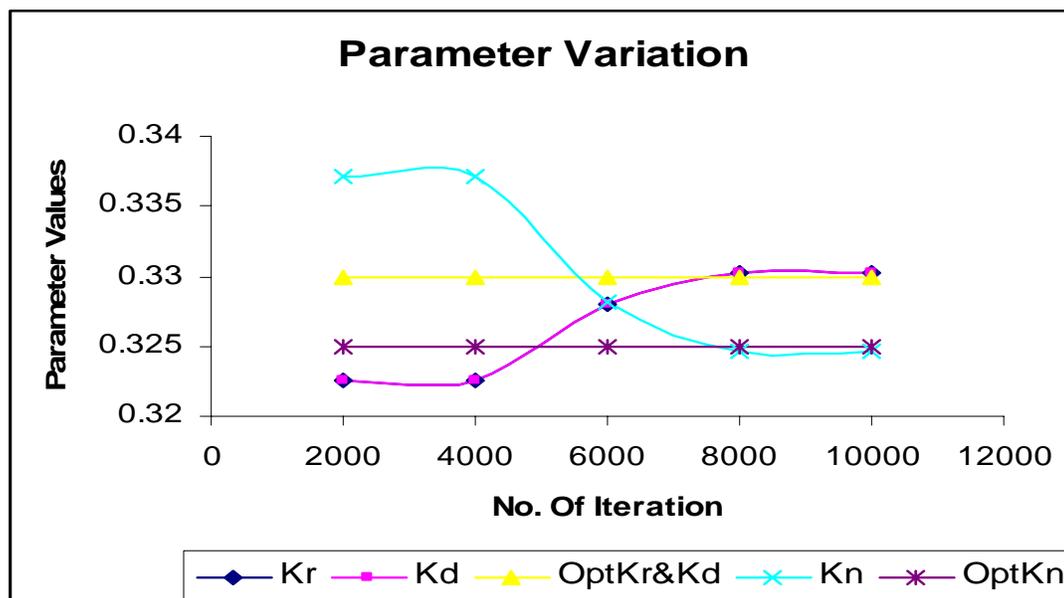
Figure 4.1 Parameter Fluctuation for Sowashee Reach Type 1

Deterministic Results for Sowashee Reach Type 1

The test case was subjected to the deterministic technique of parameter estimation by using the software PEST the result of which is presented as follows.The best thing about using PEST is the advantage of computational time , which though of not a great significance in this model can be a crucial decision making factor for the selection of a parameter estimation tool in other more complex models. In this case the optimized parameter estimate by PEST was much faster than the Monte Carlo simulations. The results obtained through PEST are as follows.

Table 4-5

Deterministic Results for Showashee Reach Type 1

Parameters	Estimated Values.	95% Confidence Lower Limit.	95%Confidence Upper Limit.
Kr	0.333	0.331	0.335
Kd	0.333	0.331	0.335
KN	0.319	0.316	0.323

The confidence limit provided in the table specifies the boundary of 95% confidence interval for the estimated parameter values. This implies that we are having a probability of 95% that the parameter corresponding to the minima of the objective function will be in this interval. It should be noted that parameter confidence limits provide only an idea about the parameters uncertainty since they are based on the same linearity assumption which is necessary to carry out the process minimization of sum of square of residual. In case of nonlinear functions the function is made linear with the help of Jacobian matrix as explained in the chapter three. Therefore, it's worth noting that if the confidence limits are large they will, in all probability; extend further into parameter space than the linearity assumption itself.

Sowashee Test Case Type 2 (Fabricated Observed Data)

This was essentially the same test case as the preceding one with only difference being in the number of points taken or chosen to be the observation points. In this case we took only 8 from the 73 possible points as the observed data and results are presented.

The coefficient of variation for Kr and Kd was kept as 0.2 and for Kn it was kept at 0.15

based on literature values.

Stochastic Results for Sowashee Reach Type 2

Table 4-6

Stochastic Results for Sowashee Reach Type 2

Number of Iterations.	Minimum Objective Function	Kr	Kd	KN
2000	0.000011	0.322589	0.322589	0.337192
4000	0.000011	0.322589	0.322589	0.337192
6000	0	0.328066	0.328066	0.328189
8000	0	0.330222	0.330222	0.324674
10000	0	0.330222	0.330222	0.324674

The parameter variation with respect to the number of iterations is graphically plotted.
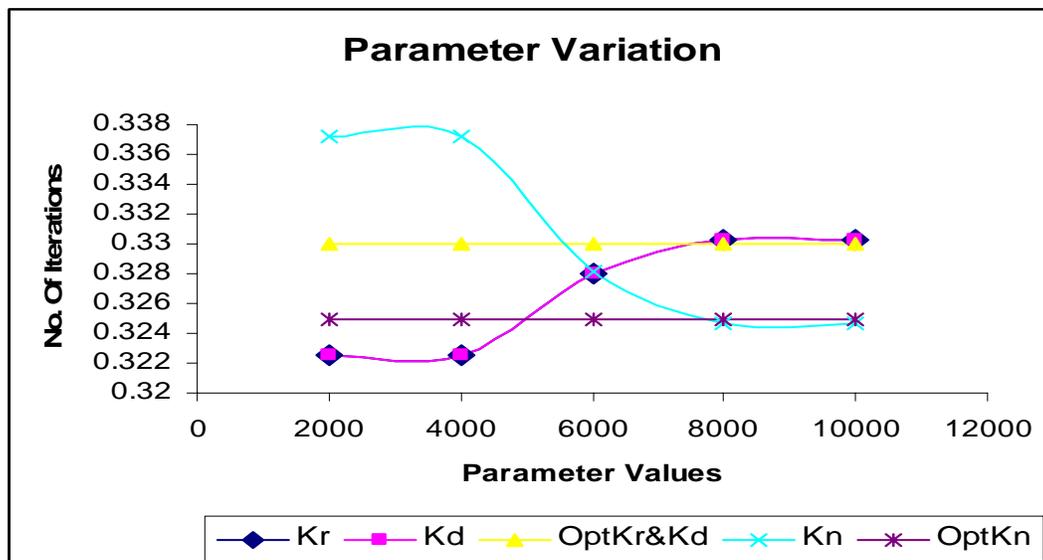
Figure 4.2 Parameter Fluctuation for Sowashee Reach Type 2

We can very well see that the fluctuation in the parameter ceases after 6000 iterations. The results are not very much different from the preceding test case but in case where the observation points are very less, we might end up with lot of different combination of parameters giving the minima of the objective function under consideration. The parameter estimates obtained in such a scenario would not be limited to just a value rather a posterior parameter distribution will be recommended.

Deterministic Result for Sowashee Reach Type 2

The result from PEST doesn't show much change, which implies that minimization of the sum of square of residual i.e. the difference between the predicted and observed data can work just fine with less than optimum number of observed data. Presented here is the result from PEST, though the number of observation points should not be less than the parameters to be estimated.

Table 4-7

Deterministic Result for Sowashee Reach Type 2

Parameters	Estimated Values	95% Confidence Lower Limit	95% Confidence Upper Limit
Kr	0.333	0.331	0.335
Kd	0.333	0.331	0.335
KN	0.319	0.316	0.323

Sowashee Test Case Type 3 (Fabricated Observed Data)

In this test case there were two different reaches from Sowashee river application were taken and 3 parameters for each, i.e. 6 parameters were tried to be estimated. The detail description of the reaches for which the parameters are being estimated is as follows.

Table 4-8

Initial Parameter Values for Sowashee Test Case Type 3

River Name	Origin Mile	End Mile	Kr	Kd	KN
Sowashee	11.5	4.3	0.300	0.300	0.300
Sowashee	2.7	0	0.400	0.400	0.300

The initial parameters were mutated as shown in the following table and 6 predicted points from the first reach and 4 predicted points from the second reach were taken and then used as the field or the observed data.

Table 4-9

Initial and Mutated Parameter Values for Sowashee Test Case Type 3

Parameters	Kr	Kd	KN
Initial Mile Pt=11.5	0.300	0.300	0.300
Mutated	0.330	0.330	0.325
Initial Mile Pt = 2.7	0.400	0.400	0.300
Mutated	0.420	0.420	0.320

The coefficients of variation for the above parameters were kept at 0.15 for Kr and Kd and 0.2 for KN. The test case was subjected to both deterministic and stochastic technique and the results are compiled.

Stochastic Results for Sowashee Reach Type 3

The reach was subjected to different number of iterations and parameter sets corresponding to minimum objective function was recorded. The following table shows the results.

Table 4-10

Stochastic Results for Sowashee Reach Type 3

Number of Iterations	Minimum Objective function	Origin Mile point	Kr	Kd	KN
10000	0.000122	11.5	0.319729	0.319729	0.34428
		2.7	0.422932	0.422932	0.313219
20000	0.000122	11.5	0.319729	0.319729	0.34428
		2.7	0.422932	0.422932	0.313219
30000	0.000122	11.5	0.319729	0.319729	0.34428
		2.7	0.422932	0.422932	0.313219
40000	0.000091	11.5	0.337902	0.337902	0.313733
		2.7	0.432262	0.432262	0.30123
50000	0.000011	11.5	0.337902	0.337902	0.313733
		2.7	0.432262	0.432262	0.30123

We can see that the parameter values corresponding to the minimum objective function reach a constant value after forty thousand iterations. It's worth noting that the accurate and close parameter estimates are obtained only after lot of iterations, the reason being increase in the number of parameters to be estimated. The number of parameter to be estimated in this case when we are taking two reaches simultaneously is six and as the results show that we converge to the true optimum value of parameters only after forty thousand iterations. It is concluded that the process of parameter estimation should be carried out in a reach wise manner because as we increase the number of reaches the number of parameters to be estimated increases by three times and the process of parameter estimation will be a very time consuming exercise in such a scenario. But, Its worth mentioning that the best parameter estimate can still be estimated using this technique but the time required for the process will be enormous. Never the less the fluctuation in the parameter values corresponding to the minimum objective function reaches a constant value after 60000 iterations the corresponding graph is presented, for origin mile point 11.5.
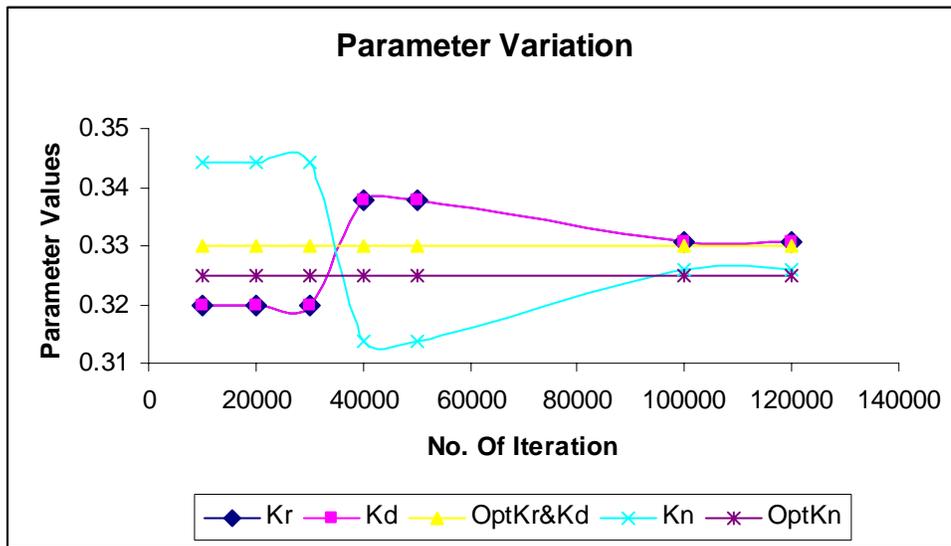
Figure 4 - 3 Parameter Fluctuation for Sowashee Reach Type 3 Origin Mile 11.5

There is not much fluctuation in the parameter values for the second reach. As the objective function is made up of the observed data points in both the reaches there seems to be some kind of correlation between the parameters in both the reaches.
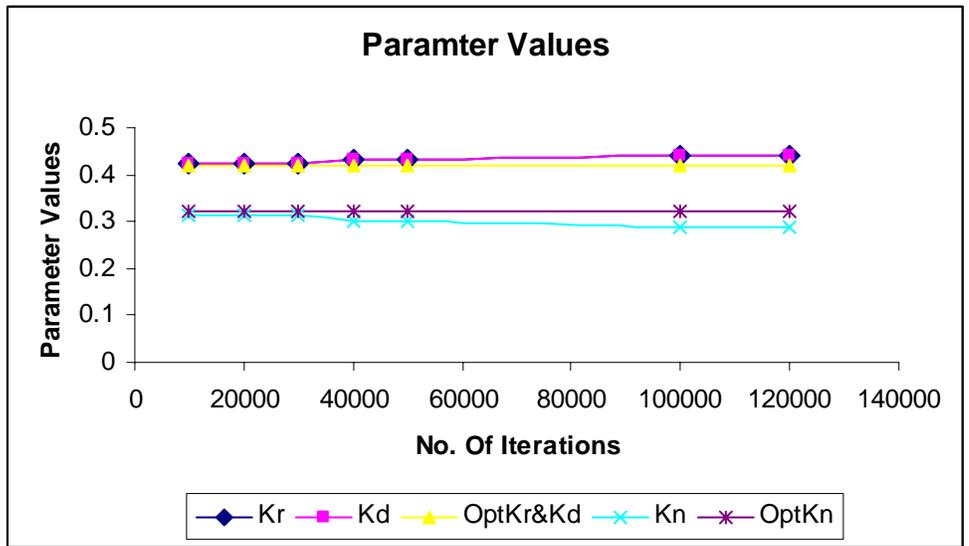
Figure 4.4 Parameter Fluctuation for Sowashee Reach Type 3 Origin Mile 2.7

Deterministic Results for Showashee Reach Type 3

The results obtained from the PEST are compiled here and interpreted subsequently.

Table 4-11

Deterministic Results for Sowashee Reach Type 3

Estimated Parameter values	Origin Mile = 11.5	Origin Mile = 2.7
Kr	0.340	0.600
Kd	0.340	0.600
KN	0.328	0.600

It can be seen that the parameter estimates obtained for the origin mile 11.5 works out close to the expected value but the estimation for the subsequent reach (i.e for the origin mile 2.7) is totally out of the range. This leads us to conclude that PSET works in most efficient manner if it is used for parameter estimation in a reach wise manner for the STREAM-C model, the reason being the continuity of the D.O. deficit with respect to the parameters is broken as we go from one reach to the other. One of the prerequisite for the use of PEST is that the model output should be continuous with respect to the parameters all along..

Sowashee Test Case Type 4 (Fabricated Observed Data)

The factor that was varied in this test case was the extent of mutation to obtain the observed data. The initial and mutated parameters are presented here, the coefficient of variation for Kr and Kd was 0.15 and for KN it was again kept at 0.20.

Table 4-12

Initial and Mutated Parameter Values for Showashee Reach Type 4

Parameters	Kr	Kd	KN
Initial	0.300	0.300	0.300
Mutated	0.450	0.450	0.150

Stochastic Results for Sowashee Reach Type 4

For this test case 15 from 73 possible predicted points with mutated parameters were taken to serve as the observed data. The Monte Carlo simulation results are presented here.

Table 4-13

Stochastic Result for Showashee Reach Type 4

No. of Itrations.	Min. ObjFun.	Kr	Kd	KN
2000	0.072649	0.425232	0.425232	0.179265
4000	0.010181	0.21445	0.21445	0.55334
6000	0.010181	0.21445	0.21445	0.55334
8000	0.010181	0.21445	0.21445	0.55334
10000	0.007027	0.4509	0.4509	0.150789

The parameters estimated from Monte Carlo simulations were again exactly the same as the optimum value of the parameters. The point that is worth noting here is that the optimum value of the parameters were obtained only after 10000 iterations and which might again increase by many folds if the optimum value of parameters were far away from the initial value of the parameter which is also assumed as the mean for the sampling algorithm. It is desired and recommended to have a good idea about the optimum value of the parameter so that we can start with a value some what in the vicinity of the optimum value and can cut down on the number of iterations required to reach to the optimum value of the parameters. In a worst case scenario, if we don't have

any idea about the optimum value of the parameters we can use the stochastic technique with fairly large coefficient of variation and do lot of iterations to reach to the optimum value of the parameters.

Deterministic Results for Showashee Reach Type 4

The results obtained by PEST were as accurate as what we got from stochastic technique. Its worth reiterating that a significant advantage of using PEST is the amount of time consumed which is much less than the stochastic technique.

Table 4-14

Deterministic Result for Showashee Reach Type 4

Parameters	Estimated Values	95% Confidence Lower Limit	95% Confidence Upper Limit
Kr	0.450	0.450	0.450
Kd	0.450	0.450	0.450
KN	0.150	0.150	0.150

The results obtained by PEST in this test case were exactly same as the mutated parameters. We are inclined to conclude that PEST may be a good tool for conducting parameter estimation for the STEAM-C model as long as we have got more observed data points than the number of parameters to be estimated, which is an essential perquisite for the usage of PEST and the estimation process is carried out in a reach wise manner for the model. The use of PEST for the STREAM-C model is recommended to be

carried out in a reach wise manner and the result produced, when applied to more than a single reach is not accurate, because the D.O. deficit equation (4.1) is not continuous that is, it is not differentiable with respect to parameters between reaches because of different source/sink of D.O. that usually is there when another reach starts. The continuity of the model equation with respect to the parameters is must for the usage of PEST as mentioned section 4.3-f.

Big Black Test Case Type 1 (Fabricated Observed Data)

The Big Black River that we try to model is 14.6 mile long and is composed of just a single reach. The single reach on which the process of parameter estimation is carried out is as follows.

Table 4-15

Initial Parameters for Big Black Reach Type 1

River Name	Origin Mile	End Mile	Kr	Kd	KN
Big Black R	66.6	52.0	0.300	0.300	0.300

The input file to model the Big Black River is set up in such a manner that the model predictions are obtained at every 0.73 miles. The river is also subjected to different kind of loads which are presented here.

Table 4-16

Description of Loads for Big Black Reach Type 1

	Flow (Cfs)	DO (Mg/L)	CBOD(lbs/day)	TKN (lbs/day)
Headwater	85.000	5.200	913.90	45.81
Waste Source	1.550	5.200	938.25	166.80
Distributed Input	0.100	5.200	1.08	0.05

The fashion in which the model is set up results in model predictions at 20 points along the river. These prediction points were obtained with the help of mutated parameters and 10 out of these 20 points were taken as the observed data. The initial and mutated parameters are presented as follows:

Table 4-17

Initial and Mutated Parameters for Big Black Reach Type 1

Parameters	Kr	Kd	KN
Initial	0.300	0.300	0.300
Mutated	0.335	0.335	0.320

The test case was subjected to both stochastic and deterministic parameter estimation techniques and results compiled. For this modeled river the deterministic technique seems to work better, although stochastic technique also produces good results, when the initial parameter is in close vicinity of the parameter that is finally estimated.

To develop a better understanding, the results of both stochastic and deterministic technique are compiled here:

Stochastic Results for Big Black Reach Type 1

The results from the Monte Carlo iterations are plotted and compiled. The interpretation of the result is given subsequently

Table 4-18

Stochastic Result for Big Black Reach Type 1

Number of Iterations	Minimum Objective Function.	Kr	Kd	KN
2000	2000	0.190126	0.190126	0.112379
4000	4000	0.2078	0.2078	0.119302
6000	6000	0.2078	0.2078	0.119302
8000	8000	0.335144	0.335144	0.315963
10000	10000	0.335035	0.335035	0.314373

Monte Carlo iteration gives a fairly accurate results; the parameter variation with number of iterations is graphed. The plot corresponding to OptKr&Kd and OptKn is the value of the parameters which were used for the generation of the data points which served as the pseudo field data.
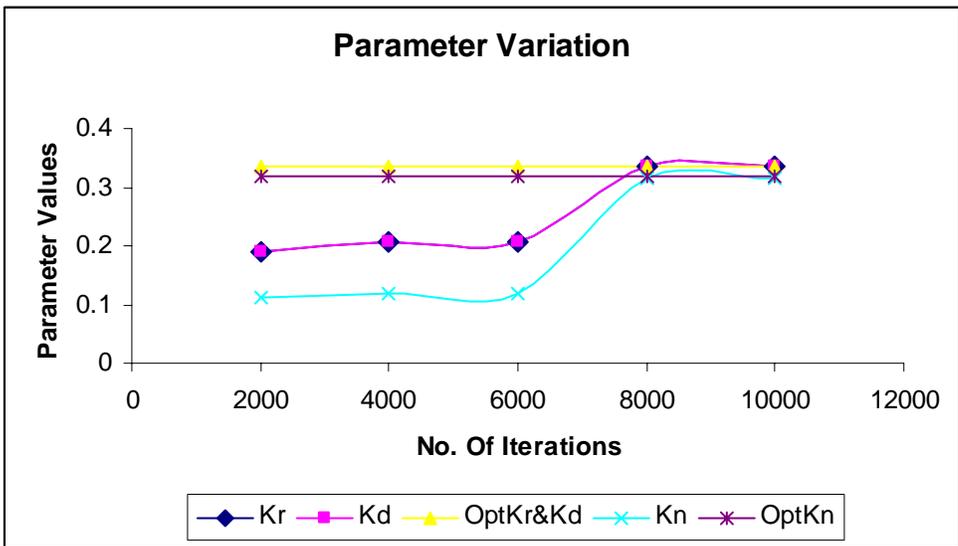
Figure 4.5 Parameter Fluctuation for Big Black Reach Type 1

After six thousand iterations the fluctuation in the parameter values begins to reduce and it start converging to the optimum value.

Deterministic Results for Big Black Reach Type 1

The test case was also subjected to parameter estimation through PEST, the results are as follows.

Table 4-19

Deterministic Result for Big Black Reach Type 1

Parameters	Estimated Values	95% Confidence Lower Limit	95% Confidence Upper Limit
Kr	0.3357	0.3357	0.3357
Kd	0.3357	0.3357	0.3357
KN	0.3219	0.3219	0.3219

As far as accuracy is concerned the values estimated by PEST are as good as what we get from the stochastic technique. The other thing that is worth mentioning is that when ever we apply PEST it's necessary that the numbers of observation points are always more than the number of parameter to be estimated.

Big Black Test Case Type 2 (Fabricated Observed Data)

This is the same test case as the preceding case; The only difference between test cases being that 5 of the 20 predicted points were taken as the pseudo field data as opposed to 10 in the last test case. The results of parameter estimation are presented as follows.

Stochastic Results for Big Black Reach Type 2

The result of obtained by Monte Carlo iteration is compiled and subsequently plotted.

Table 4-20

Stochastic Results for Big Black Reach Type 2

Number of iterations	Minimum Obj. function	Kr	Kd	KN
2000	0.000291	0.24162	0.24162	0.474397
4000	0.000206	0.328356	0.328356	0.331083
6000	0.000006	0.335028	0.335028	0.320825
8000	0.000006	0.335028	0.335028	0.320825
10000	0.000006	0.335028	0.335028	0.320825

The Monte Carlo simulation still gives accurate parameter estimates and the optimum values of the parameters are reached after six thousand iterations in this case.
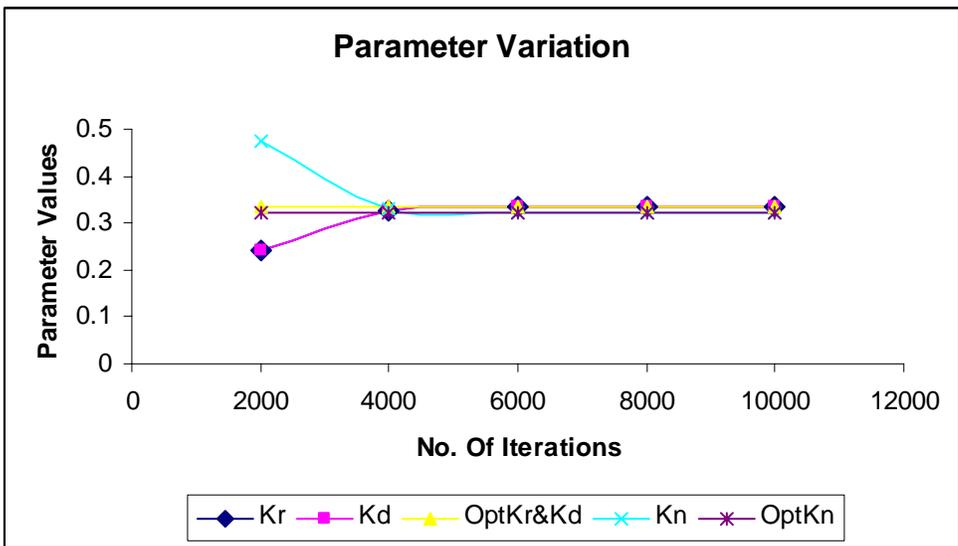
Figure 4.6 Parameter Fluctuation for Big Black Reach Type 2

Deterministic Results for Big Black Reach Type 2

When we estimated the parameter for the same test case using the software PEST the results were again close to the correct (mutated) values. Even with fewer observation points the accuracy was not compromised, which strengthens our belief in the Gauss Marquardt Levenberg algorithm which is working behind PEST. The results obtained by PEST are as follows.

Table 4-21

Deterministic Result for Big Black Reach Type 2

Parameters	Estimated Values.	95% Confi. L. Limt.	95%Confi. U. Limt.
Kr	0.3349	0.3349	0.3349
Kd	0.3349	0.3349	0.3349
KN	0.3228	0.3228	0.3228

Looking at the result we can easily conclude that the accuracy of PEST is as good as what we get from stochastic technique. The added advantage of the technique is that it takes much less time to determine the optimum value of the parameters.

Little Tang – Test Case Type 1 (Fabricated Observed Data)

This test case was based on a study conducted by MDEQ for a watershed comprising of two rivers, one of which was Town Creek which flows into Little Tang. The river Little Tang is 15.7 miles long and Town Creek is 3.8 miles long. The system is subjected to number of point and distributed loads. The river Little Tang has been broken down into 18 reaches for the purpose of modeling. The Town Creek had been broken down into 4 reaches based on different rates and other hydraulic characteristics for the purpose of modeling. For this test case a single reach was taken first and the process of parameter estimation was carried out, the reach on which the technique of parameter estimation was implemented is tabulated as follows.

Table 4-22

Initial Parameters for Little Tang Reach

River Name	Origin Mile	End Mile	Kr	Kd	KN
Little Tang	9.30	8.10	0.400	0.400	0.300

The above reach was also subjected to a number of different kinds of load which are summarized and tabulated as follows.

Table 4-23

Description of Load for Little Tang Reach

	Flow (Cfs)	DO (Mg/L)	CBOD(lbs/day)	TKN (lbs/day)
Upstream	1.850	7.372	10.41	0.50
Dist. Input	1.470	6.00	213.93	15.85
Waste Source	0.340	5.000	3.66	0.18

The model input file with all the loads and other parameters was set up in such a manner that model gave us predictions at every 0.1 mile; hence we had 13 predicted points in this reach. The initial parameters were mutated and 4 from the 13 predicted points were used as the pseudo field data and both stochastic and deterministic techniques were applied for parameter estimation and results are presented. Values of initial and mutated parameters are presented in the following table.

Table 4-24

Initial and Mutated Parameters for Little Tang Reach

Parameters	Kr	Kd	KN
Initial	0.400	0.400	0.300
Mutated	0.335	0.335	0.320

Stochastic Results for Little Tang-Town Creek Reach Type 1

Monte Carlo sampling was done for the specified number of iterations and the parameters corresponding to the minimum objective function was recorded the results of which are presented as follows.

Table 4-25

Stochastic Result for Little Tang Reach Type 1

Number of Iterations.	Minimum Obj Function	Kr	Kd	KN
2000	0	0.320323	0.320323	0.364057
4000	0	0.320323	0.320323	0.364057
6000	0	0.333866	0.333866	0.322877
8000	0	0.333866	0.333866	0.322877
10000	0	0.33754	0.33754	0.320612

If we graph the parameter fluctuation with number of Monte Carlo iterations it is interesting to see that the parameter values converge to the optimum value after six thousand iterations.
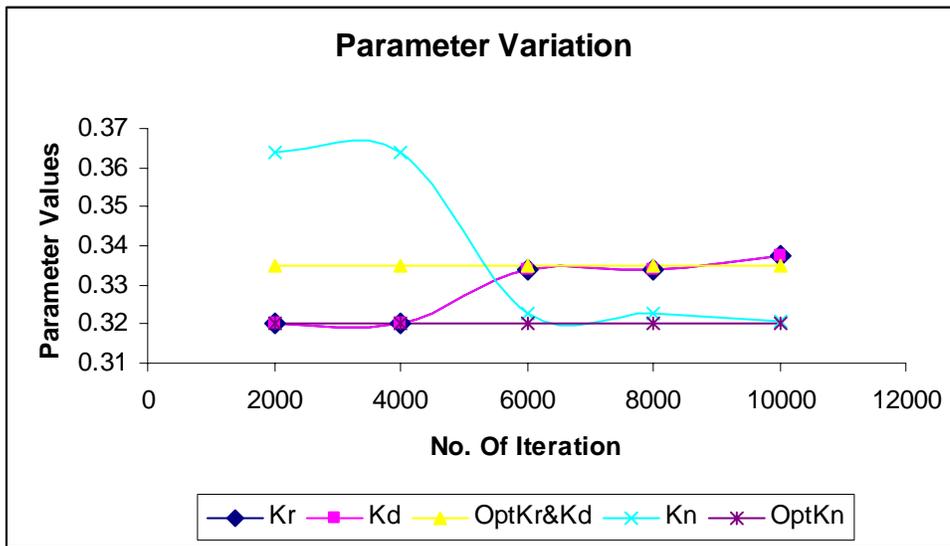
Figure 4 - 7 Parameter Fluctuation for Little Tang Reach

Deterministic Results for Little Tang-Town Creek Reach Type 1

The same test case was subjected to the process of parameter estimation using

PEST and the results are as follows.

Table 4-26

Deterministic Result for Little Tang Reach Type 1

Parameters	Estimated Values	95% Confi. L. Limt.	95%Confi. U. Limt.
Kr	0.3149	0.3049	0.3549
Kd	0.3149	0.3049	0.3549
KN	0.3228	0.3228	0.3228

Although the estimated parameters do not appear to be as close to the true (mutated) value as what we were getting in the previous test cases, it's worth mentioning that algorithm behind the PEST strives for minima of sum of square of residual and it's quite possible that there are more than one set of parameters which gives us the exactly same model predictions due to nonlinearity of the model. Presented here is the table which supports the above argument.

Table 4-27

DO Prediction with True and Estimated Parameters

Observation Mile Point- Little Tang	Observation with Mutated Paramts	Observation with Estimated Parmts	Residual
9.2	6.455 (mg/L)	6.455 (mg/L)	0.00
9.0	5.920 (mg/L)	5.920 (mg/L)	0.00
8.4	4.844 (mg/L)	4.844 (mg/L)	0.00
8.2	4.615 (mg/L)	4.615 (mg/L)	0.00

Hence while the estimated parameter might not be the same as the mutated parameters it produces the same results as far as model predictions are concerned. It's also possible that the in the particular test case the outcome that is the model prediction is not very much sensitive to the parameters and there fore we get exact same prediction with two different set of parameters. It's quite possible that we might get same prediction if we start with a slightly different set of parameters again

Indian Creek – Test Case (Real World Observed Data)

This test case is based on a modeling study conducted by MDEQ for which a limited field study was conducted to provide data for the model calibration, resulting in one observed data station per reach. The portion of the Indian Creek modeled by MDEQ is 3.7 miles long; it starts from mile point 10.7 and ends at mile point 7.0. It's broken down in 6 reaches for the modeling purpose. six reaches would imply that we need to estimate 6 times 3 i.e. 18 parameters. The process of parameter estimation using PEST requires more number of observation points than the number of parameters to be estimates which is not the case in this example. Therefore, it was decided to keep the parameters constant across the individual reaches more so as the length of the river being modeled was not very long and was just 3.7 miles. Such an assumption reduced the numbers of parameters to be estimated to just three. The sequence in which the input file for the Indian Creek is set up is presented.

Table 4-28

Processing Sequence for Indian Creek

River Name	Initial Mile	Terminus Mile
Indian	10.7	10.2
Indian	10.2	9.4
Indian	9.4	8.8
Indian	8.8	8.4
Indian	8.4	8.2
Indian	8.2	7.0

The observed data that was obtained by MDEQ is also tabulated below.

Table 4-29

Observed Data by MDEQ for Indian Creek

River Name	Mile Point	Dissolved Oxy. (mg/L)
Indian	10.7	6.9
Indian	10.2	6.36
Indian	9.4	5.43
Indian	8.8	2.98
Indian	8.4	3.23
Indian	8.2	4.31
Indian	7.0	6.21

Stochastic Results for Indian Creek

The above test case was subjected to the stochastic technique and the results are compiled below.

Table 4-30

Stochastic Results for Indian Creek

Number of Iteration	Min. Objective Function	Kr	Kd	KN
10000	150.52	0.134	0.134	0.145
20000	150.48	0.134	0.134	0.145
30000	150.28	0.135	0.135	0.145
40000	150.45	0.135	0.135	0.145
50000	150.02	0.134	0.134	0.144

We can see from the results compiled that the minimum objective function doesn't decrease much with an increase in the number of iterations For this test case we have no idea about the initial value of the parameters hence we are only guided by the literature values. We also don't have any idea as to how close be our initial parameters to the optimum value unlike the previous test where the mutated values were the true set of parameters for the test case. The increase in number of iteration does not result in significant reduction in the value of the objective function and hence we are inclined to believe that we might require much more number of iterations which might take lot of time for the parameter estimation process. The other problem with stochastic technique which this program implements is that if we change the initial values of the parameters and let's say coefficient of variation also, we will have to run those thousands of iterations again which might be time consuming exercise again. But as mentioned before in such a scenario it is recommended to use large coefficient of variation which will surely increase our chance of finding the optimum value of the parameters.

Deterministic Results for Indian Creek

As we were not sure about the initial value of parameters to be estimated, this particular test case was run several times with different initial values and other changes in the file that is set up to run the PEST and the results that were obtained are as follows.

Table 4-31
Deterministic Result for Indian Creek

Parameters	Initial Values	Estimated Values
Kr	1.00	13.170
Kd	1.00	13.170
KN	0.20	0.20

The graph of the observed D.O. against D.O. readings obtained from initial parameter and estimated parameter is also presented.
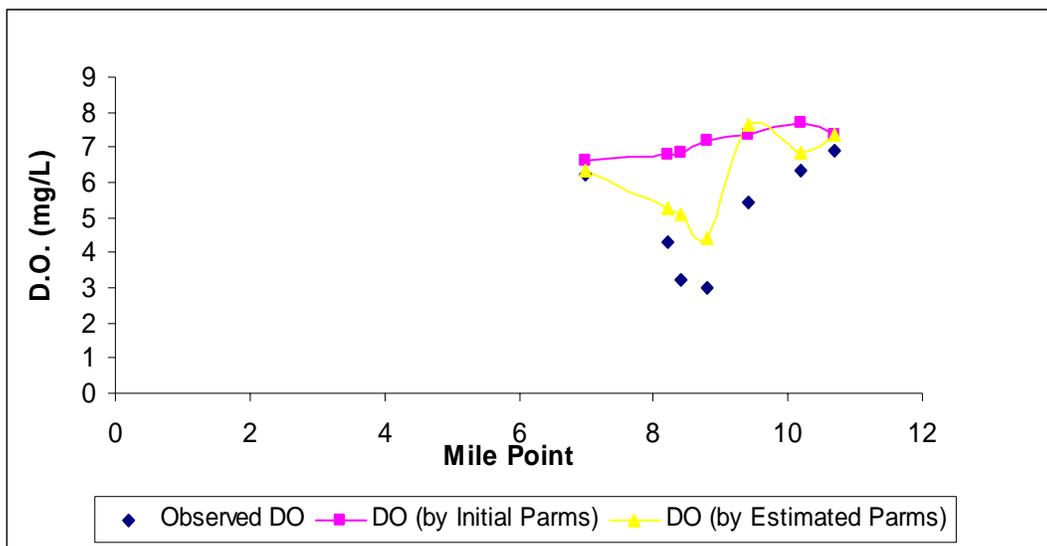

Figure 4.8 D.O. Prediction with initial and estimated parameters against observed

It's worth mentioning that the value of the objective function as computed by the initial parameters went down from 42.73 to 11.91 when computed with help of optimized

parameters. It should also be noted that due to lack in information about the initial value of the parameters the objective function computed with stochastic technique was of the order of 150 and above. Although we get better results with PEST the unrealistically high value of Kr and Kd suggest that it is not a biological reaction which is taking place in the reach under the study. Such a high value of the parameters suggest that there is some kind of chemical reaction taking place Nevertheless result of the research done strongly suggests that the PEST is very well suited to carry out the process of parameter estimation for the STREAM-C model.

Tallahala – Test Case (Real World Observed Data)

This test case was also based on a modeling study completed by MDEQ and for which limited field data were available. The portion of Tallahala River which was modeled is 15.8 mile long and has got a tributary, Tallahoma, flowing into it. For the modeling purpose Tallahala is broken down into 12 reaches and Tallahoma is processed as a single reach. By running different kind of test cases it has been found that the technique of parameter estimation both stochastic and deterministic works best when we implement them for a single reach and conduct reach wise parameter estimation. In essence whenever we try to carry out the process of parameter estimation for a river which has been broken down in multiple reaches, we are trying to estimate the value of three times the number of reach parameters, and for to PEST to work it's required that we always have greater number of observed data then the parameters to be estimated. Presented here is the table depicting the sequence of reach for the Tallahala input file:

Table 4-32

Processing Sequence for Tallahala Creek

River Name	Incoming. Trib.	Origin Mile	Terminus Mile
Tallahala	Nil-	53.5	53.1
Tallahala	-Nil-	53.1	52.0
Tallahala	-Nil-	52.0	50.8
Tallahala	-Nil-	50.8	50.3
Tallahala	-Nil-	50.3	49.5
Tallahala	-Nil-	49.5	48.4
Tallahala	-Nil-	48.4	44.4
Tallahoma	-Nil-	0.4	0.0
Tallahala	Tallahoma	44.4	44.0
Tallahala	-Nil-	44.0	41.8
Tallahala	-Nil-	41.8	37.4
Tallahala	-Nil-	37.4	27.7

The above table presents the number of reaches for which the parameter needs to be estimated, taking 3 parameters for each reach it comes out to be 36 parameters to be estimated. Both the techniques that we use for parameter estimation depends on the number of observed data. The observed data which was obtained by the courtesy of MDEQ is compiled here.

Table 4-33

Observed Data by MDEQ for Tallahala Creek

River Name	River Mile	Minimum D.O. (mg/L)	Average D.O(mg/L)	Max.D.O(mg/L)
Tallahala	53.1	7.0	7.8	9.6
Tallahala	52.0	6.8	7.6	9.5
Tallahala	49.9	5.1	6.9	9.3
Tallahala	48.4	5.7	7.3	9.6
Tallahala	44.4	7.5	8.4	9.9
Tallahala	37.4	8.4	9.6	12.0
Tallahoma	44.3	8.1	8.5	9.1

As it can be easily seen that we have just seven observed D.O. data for the Tallahala River where as the number of parameters to be estimated is three times the number of reaches (i.e. 36).Therefore PEST can't work in such a scenario and the results from stochastic technique will also be anything but exact. The only option that we are left with is the parameter estimation through manual calibration but, here again the observed data are not sufficient to constrain the calibration. Hence it's highly recommended and desired that we have more field data to carry out the process of parameter estimation in a significant manner

.

CHAPTER V

CONCLUSION

The STREAM-C model was developed on the lines of the STREAM model used by the Department of Environmental quality Mississippi State. STREAM-C was subjected to different techniques of parameter estimation, with the help of several artificial and real world test cases. This chapter summarizes the results and conclusions that were drawn by conducting the above research work.

Conclusions by Stochastic Technique

The stochastic technique implemented used Monte-Carlo sampling coupled with an objective function as explained in chapter 3. Several artificial and real world test cases were subjected to this technique. The following observations were made as a result of those test cases.

a)   Stochastic technique was more efficient in time when the parameter estimation was carried out in a reach wise manner.

b)   Initial values of the parameters which were also taken as the mean for the sampling technique should be in the vicinity of optimum parameter value to get the best parameter estimates with less number of iterations.

c) Reach wise parameter estimation with reasonable initial parameter values converged to the true parameter values after an average of 6000 iterations for the artificial test cases.

In absence of any idea about the initial parameter values and lack of information on the coefficient of variation the technique was not able to produce parameters corresponding to the minimum objective function and it was far from accurate see the test case of Indian Creek in chapter 4. But, it's worth mentioning that if we start with a very big coefficient of variation if we have no idea about the initial parameters and coefficient of variation then we might just get plausible or accurate parameter estimate but the number of iterations required by the program will be enormous and the technique will be far from efficient in time.

Conclusions by Deterministic Technique

The process of parameter estimation through deterministic technique was carried out using PEST. All the test cases developed both artificial and real were subjected to parameter estimation using PEST and following observations were made.

a) As in the case of stochastic technique more accurate parameter estimates were obtained when the estimation was carried out in a reach wise manner because inside the reach the DO deficit, the governing model outcome, is continuous and differentiable with respect to the parameters which is not

the case between the reaches. The continuity of the model with respect to the parameters is an essential prerequisite for the usage of PEST.

b)   Accurate parameter estimates for a river reach using PEST can only be obtained if we have greater number of observation data than the number of parameters to be estimated.

c)   The initial parameter values should again be in the vicinity of the optimum parameter values to get the best parameter estimates, other wise we might end up getting parameter corresponding to the local minima of the objective function then the global minima.

d)   In absence of any information about the initial parameter values the problem of local minima can be tackled by running PEST with different initial values of the parameters as seen in the case of Indian Creek in chapter 4.

It was finally concluded that for the STREAM-C model, PEST might be a preferable choice for parameter estimation. This conclusion is based on the premise that parameter estimation is being carried out in a reach wise manner. The accuracy of the results obtained by PEST are as good as the stochastic technique but the time required for the parameter estimation process is much less. On an average the amount of time required to run the model for 250,000 iterations while implementing the stochastic technique takes about twenty five to thirty minutes but the time taken by PEST is less than five minutes. PEST also gives us the 95% confidence limit for the estimated parameters which though not very accurate gives us some idea about the uncertainty of the parameter estimated.

APPENDIX A

INPUT FILES

APPENDIX A

INPUT FILES

This appendix provides the format of basic input file for the STREAM-C model and the other files required to run the program which implements the stochastic parameter estimation technique. Also the format of the output file produced by the program is explained. STREAM-C is a dos based program done in C and all the file required and made by the program are ASCII files. The various files required and produced by the STREAM-C model are as follows.

- Input File. The file that needs to be processed.

- Output File. The file in which the model results will be stored.

- Variance File. The file having coefficient of variation and number of iterations for which the program needs to run.

- Observation File. The file having the observed data.

- Model Simulation File, The file that will store the model results of all the simulations.

- Parameter Simulation File. The file that will store the base and all the simulated parameters.

- Best parameter File. The file that stores the parameter corresponding to the minimum objective function.

The description and the format of all the above listed files are presented in this section. The structure of the input file for the model is presented.

A-I) STREAM-C Input File

The input file for the STREAM-C model is a column formatted ASCII file and the entries made can be described as follows. There are six tables which together makes the complete input file. Detail description of the table is provided.

i)   Head Water or Tributary table.

Only one entry may be made in this table for each river. This entry identifies the beginning of that river reach. The structure of the column format is as follows.

Column

1 – 12   River Name.

13 – 80   Stream Flow (cfs).

Dissolved Oxygen (Mg/l)

Carbonaceous BOD (lbs/day) and TKN (lbs/day).

ii)   Waste Source and Specified Input table.

The waste source and specific input table are exactly alike and are processed in exactly the same manner the structure of which is provided here.

Column

1—12   River name.

13—62   Origin Mile.

Waste Discharge (mg/L),

Dissolved Oxygen (mg/L),

CBOD and TKN (lbs/day), Source type can be 0 for point load and 1 for distributed load.

iii)   Reach Parameter Table.

This table contains the data necessary to specify or calculate the velocity and the reaeration coefficient for the reach. There are six alternative forms of this type of card. The form to be used will depend on the type of reaeration formula specified for the reach. The types of the reaeartion formula that can be used are as follows

- O'Connor-Dobbins

- Tsivoglou

- Specifically Stated.

Column 63 of the reach parameter card is used to select the type of reareation formula. The character "T" placed in column 63 will cause Tsivoglous formulation to be used while character "K" placed in column 63 will specify that the user has included the value of the reaeration coefficient on the card. A blank or any other character other a "T" or a "K" placed in the column 63 will cause O'Conner-Dobbins to be used. If the character "U" is placed in the column 64 of the card the user should have included the velocity and possibly the depth as the part of the data. A blank or any other character placed in the column 64 specifies that the velocity (fps) and possibly the depth (ft) will be calculated by the exponential equations. In the later case the values of the coefficient will be on the card. One reach parameter card will be included for each reach.

iv)   Reach Rate Table.

This table contains the temperature in the reach and the reaction rates. In addition this table may contain the photosynthetic oxygen production, oxygen utilized by aquatic plants trough there respiration and sediment oxygen demand. The structure of this table is as follows

Column

1—12   River Name

13—80   Origin of River,

Temperature in degrees centigrade,

Kr (/day),

Kd (/day),

KN (/day),

P (Photosynthetic oxygen demand), R (respiration) and Sediment oxygen demand.

v)   Sequence Table

This table specifies the order in which the reaches are to be processed. A tributary which is included in the solution must be processed prior to the specification as a previously processed input. Stream and tributary must be specified in the order from upstream to the mouth. The mouth of all the tributaries must be river mile zero of all the tributaries. The format of the table is as follows.

Column

   1—12   River Name,

   13—24   Name of previs processed tributary,

   37—80   Origin Mile and Terminus Mile.

Example is provided

0.73

BIG BLACK R 66.600 85.0 5.200 913.90 45.815

****** PRECEDED BY DELTA INCREMENT AND HEADWATER TABLE ******

BIG BLACK R 66.600 1.550 5.200 938.250 166.800 0   CERES 45-20-6

****** PRECEDED BY WASTE SOURCE TABLE ******

BIG BLACK R 66.600 0.100 5.200 1.075 0.054 1   SPATIAL

****** PRECEDED BY SPECIFIC INPUTS TABLE ******

BIG BLACK R 66.600 1.000 0.000 0.070 0.500

****** PRECEDED BY REACH PARAMETERS TABLE ******

BIG BLACK R 66.600 28.000 0.300 0.300 0.300 0.000 0.000 0.000

****** PRECEDED BY REACH RATE TABLE ******

BIG BLACK R   66.600 52.000

****** PRECEDED BY SEQUENCE TABLE ******

A-2) Output File

The structure of the model out put file is as follows

River Name = ***** Origin Mile = ** Terminus Mile = **

MilePt Flow   Do   CBOD   Tkn   Vel

The output is produced reachwise at every delta x mile which is specified at the beginning of the model input file. The sequence in which the reaches are prpcessed and written to the outout file comes from the sequence table of the input file.

Example is provided

River Name = Town Creek Origin Mile = 3.80 Terminus Mile = 2.70

MilePt	Flow	Do	CBOD	Tkn	Vel
3.800	0.039	5.512	1.994	0.098	0.100
3.700	0.068	5.929	1.948	0.096	0.100
3.600	0.097	5.947	1.921	0.095	0.100
3.500	0.126	6.054	1.898	0.094	0.100

A-3) Variance File

The structure of the variance file which is supplied by the user to the model is as follows

No. of Iterations. ObjFun No. Caliration Variable

Parmeter Distribution function C.O.V

The first row of the file consist of three entries number of iteration for which model is suppose to run as the first entry, objective function number decides the kind of objective function that the program evaluates. A value of 1 for this entry will evaluate sum of square of residual. The third entry on the first line is the calibration variable using which the program decides the calibration to be carried out on what kind of observed data. It can be either CBODu, TKN or DO. We have carried out the calibration on DO in all the test cases. The second line contains name of the parameter as the first entry, the probability distribution function attached to it as the second entry and the coefficient of variation as the third entry. Corresponding to each parameter there will be an entry like this in the variance file hence the number of lines in the variance file will be the number of parameters to be estimated plus one an example is presented.

8000 1 DO

Kr   Normal   0.15

Kd   Normal   0.15

Kn   Normal   0.20

A-4) Observation File

The observation file supplied to the model has the follwing structure.

River Name Origin Mile Obs DO Obs CBOD Obs TKN

The observation file corresponding to a particular test case will have as many lines as the number of observations. All the line will have the same entries. The first entry being the

name of the river followed by the origin mile which in turn is followed by the observed

DO, CBOD and TKN.an example is presented

INDIAN   10.7 6.9 7.7 8.56

INDIAN   10.2 6.36 7.7 8.56

INDIAN   9.4 5.43 7.7 8.56

A-5) Model Simulation File

The model simulation file has the model out put in the format explained in A-2 for all the

simulations performed for the number of iterations supplied through the variance fie.

Example is provided

SIMULATION 1

INDIAN   7.000 6.645 2.128 0.000

INDIAN   7.100 6.587 2.269 0.000

A -6) Parameter Simulation File

This file like the model simulation file is generated by the program number and it has the

objective function corresponding to all the simulation and the parameter values

corresponding to that simulation.

A-7) Best Parameter File

The best parameter file generated by the program is the file that has the parameter corresponding to the minimum objective function among all the total number of simulation for which the model was run with randomly sampled parameter values.

Example is provided

PARAMETERS CORRESPONDING TO MINIMUM OBJECTIVE FUNCTION

RIVER	ORIGINMILE	Kr	Kd	Kn
SOWASHEE CRK	12.00	0.300	0.300	0.300
SOWASHEE CRK	11.50	0.325	0.325	0.334
SOWASHEE CRK	4.30	0.400	0.400	0.300
SOWASHEE CRK	2.70	0.434	0.434	0.334

APPENDIX B

PEST FILES

APPENDIX B

This appendix contains the description of the files needed by PEST to carry out the process of parameter estimation on the model STREAM-C. There are three ASCII files that PEST needs to set up to carry the process of parameter estimation. They are.

- Case.tpl

- Case.ins

- Case.pst

B-1) Template File

The case.tpl is the template file and it tell PEST as to which are the parameters being estimated. The format of this file is exactly alike the input file which was described in the appendix A-1. The only difference being that the parameter to be estimated is enclosed within a special symbol declared at the beginning of the file. The user will have to set up a template file corresponding to each input file having a parameter to be estimated. In case of STREAM-C model we had to set up just one template file. An example is presented.

ptf #

Model Input File

0.73

BIG BLACK R 66.600 85.0 5.200 913.90 45.815

****** PRECEDED BY DELTA INCREMENT AND HEADWATER TABLE ******

BIG BLACK R 66.600 1.550 5.200 938.250 166.800 0   CERES 45-20-6

****** PRECEDED BY WASTE SOURCE TABLE ******

BIG BLACK R 66.600 0.100 5.200 1.075 0.054 1   SPATIAL

****** PRECEDED BY SPECIFIC INPUTS TABLE ******

BIG BLACK R 66.600 1.000 0.000 0.070 0.500

****** PRECEDED BY REACH PARAMETERS TABLE ******

BIG BLACK R 66.600 28.000 #RchKr   # #RchKd   # #RchKn   # 0.000 0.000 0.000

****** PRECEDED BY REACH RATE TABLE ******

BIG BLACK R   66.600 52.000

****** PRECEDED BY SEQUENCE TABLE ******

The first entry in the file ptf denotes pest template file and the symbol #, which can be any thing that the user sets up, helps PEST in locating the place where the estimated parameter will go in the input file as can be seen in the example presented also.

B-2) Instruction File

The case.ins is the instruction file which is set according to some rules which helps PEST to locate those data in the output file corresponding to which there are field data . The rules to set up the instruction file can be obtained from the PEST manual (www.sspa.com/PEST). User is required to set up different instruction file if the data corresponding to which there are field data is in more than one file. In the case of STREAM-C we had to set up just one instruction file.

B-3) Control File

The case.pst is the control file which holds every thing together in this file we have information like how many template files are there, how many instruction files are there, how many parameters to be estimated, how many observed data points present, the executable file that the PEST needs to call to get started with the process of parameter estimation etc.

For the STREAM-C model all the three files whose structure and content has been described was changed every time a new test case was run. The detail description of the steps needed to be followed to set up files for using PEST is present in the PEST manual see (Doherty J 2002, "PEST User's Manual", Watermark Numerical Computing Brisbane, Australia").

REFERENCES

1) Al-Khatib, S., Califf, R.M., Hasselblad, V., Alexander, J.H., McCrory, D.C., Sugarman, J.,"Placebo-controls in short-term clinical trials of hypertension", Science 292, 2013-/2015, 2001.

2) Annan, J.D.,"On repeated parameter sampling in Monte Carlo simulations", Ecological Modelling 97, 111-/115, 1997.

3) Beck, M B ,"Water Quality Modeling: A Review of the Analysis of Uncertainty", Water Resources Research, 23(8), pp 1393-1442, 1987.

4) Beven, K J and Binley, A M ,"The Future of Distributed Models: Model Calibration and Predictive Uncertainty", Hydrological Processes, 6, 279-298, 1992.
Borsuk, M.E., Stow, C.A.,"Bayesian parameter estimation in a mixed-order model of BOD decay", Water Research34, 1830-1836, 2000.

5) Brown, L C ,"Uncertainty Analysis in Water Quality Modeling Using QUAL2E in Systems Analysis in Water Quality Management (M B Beck, ed), Pergamon", Oxford, pp 309-319, 1987.

6) Carroll, R.W.H., Warwick, J.J., "Uncertainty analysis of the Carson River mercury transport model", Ecological Modelling 137, 211-/224, 2001

7) Dilks, D.W., Canale, R.P. and Meier, P.G., "Development of Bayesian Monte-Carlo techniques for water-quality model uncertainty". Ecological Modeling, 62:149-162, 1992.

8) Doherty, J." PEST user's manual", Watermark Numerical computing, Brisbane, Australia, 2002.

9) Doherty J. and Jhonston."Methodlolgies for calibration and predictive analysis for a water shed model." Journal of the American Water Resource Association 39(2), 251-265, 2003.

10) Elmore, H. L., and Hayes, T. W, "Solubility of Atmospheric Oxygen in Water," Journal, Sanitation Engineering Division, American Society of Civil Engineers, Vol 86, No. SA4, pp 41, 1960.

11) Esposito, W.R. and Floudas, "Global optimization for the parameter estimation of differential-algebraic systems". Indust. Eng. Chem. Res. 39: 1291-1310, 2000.

12) Gertner, G.Z., Fang, S., Skovsgaard, J.P.," A Bayesian approach for estimating the parameters of a forest process model based on long-term growth data", Ecological Modelling, "119, 249-/265, 1999.

13) Hakanson, L., "The role of characteristic coefficients of variation in uncertainty and sensitivity analyses, with examples related to the structuring of Lake Eutrophication models", Ecological Modeling 131, 1-/20, 2000.

14) Harmon, R. and Challenor, P.,"A Markov chain Monte Carlo method for estimation and assimilation into models", Ecological Modeling 101, 41-59, 1997.

15) Hastings, W.K.,"Monte Carlo sampling methods using Markov chains and their applications", Biometrika 57, 97-/109, 1970.

16) Hill, M.C., Methods and guidelines for effective model calibration. U.S Geological Survey Water- Resources Investigations Report 98-4005, 90p, 1998.

17) Jakeman, A.J. and Hornberger, G.M. "How much complexity is warranted in a rainfall-runoff model?" Water Resources Research, 29(8), 2637-2649, 1993.

18) Jasper , A. V,Gupta, H. V.,Bouten W., and Sorooshian S.,"A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters." Water Resources Research Vol. 39. No.8 1201, 2003.

19) Levenberg, K.,"A method for solution of certain nonlinear problems in least squares.",Quart. App. Math., 2 pp. 164-168, 1944.

20) Ludwig, D.,"Uncertainty and the assessment of extinction probabilities.",. Ecological Applications 6, 1067-/1076, 1996.

21) Malakoff, D.M.,"Bayes offers 'new' way to make sense of numbers", Science 286, 1460_/1464, 1999.

22) Marquardt Donald, W,"An algorithm for Least-Squares Estimation of nonlinear parameters." journal of the society of Industrial and applied mathematics, vol 11, No 2 pp 431-441, June ,1963.

23) Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller,A.H., Teller, E.,"Equations of state calculations by fast computing machine",Journal of Chemistry and Physics 21, 1087-1091, 1953.

24) O'Connor, D., Dobbins, W., "Mechanism of reaeration in natural streams", Proc. Amer. Soc. Cvil., 123,641-684, 1956.

25) Phillips, D.L., Marks, D.G.," Spatial uncertainty analysis: propagation of interpolation errors in spatially distributed models", Ecological Modelling 91, 213-/229, 1996.

26) Qian, S.S., Stow, A.C. and Borsuk, M.E.,"On Monte Carlo Methods for a Bayesian Inference". Ecological Modeling 159:269-277, 2003.

27) "Rates, Constants and Kinetics Formulations in Surface Water Quality Modeling, 2nd Edition, EPA/600/3-85/040, June 1985"

28) Reichert, P., Omlin, M.,"On the usefulness of over parameterized ecological models",. Ecological Modeling 95, 1997.

29) Van Horssen, P.W., Pebesma, E.J., Schot, P.P., "Uncertainties in spatially aggregated predictions from a logistic regression model", Ecological Modelling 154, 93-/101, 2002.

30) Yool, A.," Comments on the paper: on repeated parameter sampling in Monte Carlo simulations", Ecological Modelling115, 95-/98, 1999.

31) http://www.sspa.com/pest/PESTMAN.pdf