CHAPTER -2

LITERATURE SURVEY

CHAPTER-2

LITERATURE SURVEY

2.1 MATHEMATICAL MODELING

Mathematical modeling is the method of translation of the given physical or other information and data into mathematical relation, into a mathematical model. The mechanism of translating physical or biological concepts of any system into a set of mathematical relationships, and the manipulation of the mathematical system thus derived is called a model and it is an abstract representation of real world.

Models are, by their very nature, abstractions of reality used to simulate, rather than mimic, natural systems. They are seldom, if ever, truly correct (van Waveren, Groot et al. 1999) and application of models for management is often considered as much an art as it is a science. This does not imply a lack of rigour, but rather a recognition of inherent uncertainties and the need for the modeller to make intelligent choices in the development, use and reporting of models. This is both the strength and weakness of the modelling process. On one hand a model can reduce highly complex processes to simple output but, on the other, the strength of a model is determined by the relevance, and often extent, of the input data. Modeling can provide a powerful tool for management, but can be fairly meaningless if there is an ill defined objective, poor conceptualization of the causative relationships and their uncertainty, or if insufficient attention is paid to essential technical aspects of the modelling process. Failure to address, or at least be acutely aware of, these issues restricts sensible interpretation of results.

Mathematical models fall into a number of generic types that assemble and use data in different ways. All models have a *domain*, which provides the boundaries within which they were designed. Operating outside the defined domain is ill-advised.

Mathematical modeling is a fast developing area of research and development field, which has a tremendous scope with respect to environmental planning and conservation.

Here simulation models are built or developed and experimented in nature. Deterministic Models can be developed for Water quality to study the impacts or movement of BOD, COD, and DO etc of the river.

A mathematical model is a qualitative representation of a system based on mathematical relationships. Models are built in order to:

- understand functioning and internal structure of a system
- program measuring campaigns or positioning sensors
- monitor the system and control interventions

Building a model involves a number of steps, but it is not a straight procedure, as illustrated in Figure 2.1. Where more than one model is possible for a judicious choice of a particular model based on any system physical simulation is very important.

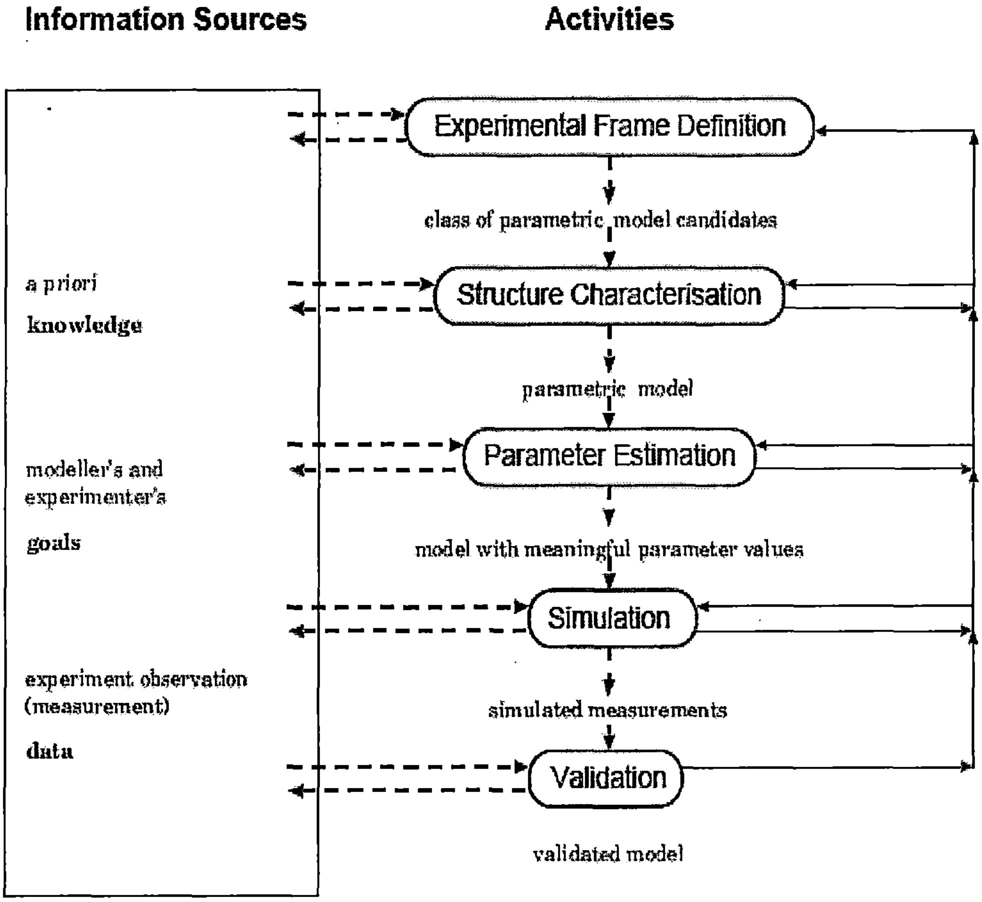

Figure 2.1: Model-based systems analysis

2.1.1 Water Quality Models

Water is required to sustain human life. Rapid growth of population and industrialization has resulted in scarcity and pollution. Absences of good management practices and regulations have exacerbated the problems. Assessment of water quality of river is therefore very important for a sustainable future.

General guidance relevant to application of models to support the implementation of the Water Framework Directive (WFD) includes:

- Define the objective that specifies the domain of the problem and the scenarios to be addressed;
- Determine if a mathematical model is needed to reach the objective;
- What reliability of model solution is required and does the expertise exist to apply the model;
- If a model is thought to be needed, provide a conceptual framework;
- Determine scope and boundary conditions, which guide data needs;
- Select a type of model;
- Verify that the conceptual model is effectively addressed by the computer program chosen;
- Check the suitability or robustness of the model to extreme values of input data;
- Check the sensitivity of the model to changes in input values;
- Calibrate the model against empirical data sets;
- Validate the calibrated model with independent data; and
- Check if objective has been achieved (did the model answer the question that is was supposed to).

2.2 The Streeter-Phelps River Pollution Model: Oxygen Sag Curve

Streeter-Phelps derived an equation (1925) that described the oxygen profile in a river, which undergoes a steady influx of pollutant at some point upstream. Initially, biodegradation of the pollutant causes a decline in dissolved oxygen C or, viewed slightly differently, an increase in the oxygen deficit $D=C^{*}-C$, where C^{*} is the equilibrium solubility of oxygen in water. As the pollutant concentration L decreases through biodegradation, the decline in oxygen concentration slows and ultimately passes through a minimum, the so-called critical point, as oxygen supply from the atmosphere replenishes the river. Further "reaeration" ultimately restores the oxygen concentration to fill saturation level.(See the figure).

In this model, Streeter and Phelps did not consider pollutant adsorption on river sediment, an important removal mechanism that will be addressed in the following example. Here only biodegradation and reaeration rates needed to be considered, both of which were assumed to be first order in concentration. The model equations are then as follows:

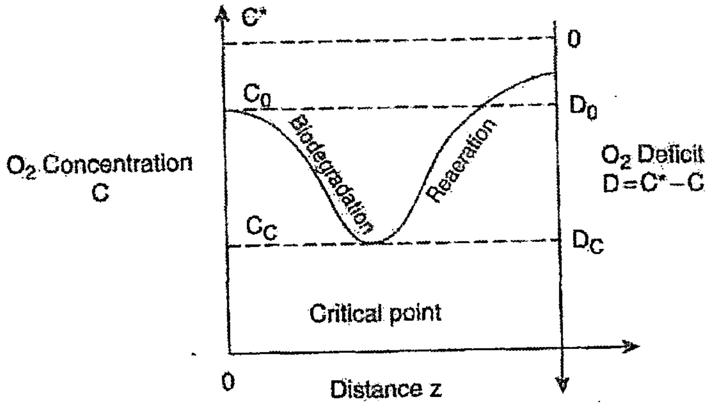

Figure: 2. 1, Streeter-Phelps model. Dissolved oxygen profiles in a river with a steady influx of a pollutant

Oxygen mass balance:

Rate of oxygen in – Rate of oxygen out = 0

$[QC\, l_z + k_L\, a\, (C^*\text{-}C)_{avg} A_C \Delta z\,]\text{-}\ [QC\, l_{z+\Delta z} + k_r\, LA_C \Delta z] = 0$_______(2.1)

which becomes, in the limit $\Delta z \longrightarrow 0$ then the above expression becomes

$$v\frac{dC}{dz}+k_rL-k_La(C^*-C)=0 \qquad \_\_\_\_\_\_\_\_\_\_\_\_(2.2)$$

where $v=Q/A_C$=Superficial river velocity, $k_L\ a$ = volumetric mass transfer coefficient m^2/m^3 river, k_r = reaction rate coefficient and L = pollutant concentration.

Alternatively we can write in terms of the oxygen deficit $D=C^*\text{-}C$ then the above equation becomes

$$v\frac{dC}{dz}-k_rL+k_LaD=0 \qquad \text{-----------------------}(2.3)$$

Pollutant mass balance:

Rate of pollutant in – Rate of pollutant out =0

$QL\, l_z\, \text{-}\ [QL\, l_{z+\Delta z} + k_r\, LA_C \Delta z] = 0$

Which yields the ODE:

$$v\frac{dL}{dz}+k_rL=0 \qquad \_\_\_\_\_\_\_\_\_\_\_\_(2.4)$$

Using separable of variables it is obtained

$$L = L_0 \exp(-\frac{k_r}{v} z) \qquad \_\_\_\_\_\_\_\_\_\_(2.5)$$

this gives the pollutant concentration profile in river. Now from (2.3) and (2.5) we get the following ODE in the oxygen deficit:

$$v\frac{dD}{dz} + k_L aD = k_r L_0 \exp(-\frac{k}{v} z) \qquad \_\_\_\_\_\_\_\_\_\_(2.6)$$

This equation is of the form

y' +f(x) y = g(x)

which has the solution of the form

y=exp -f(x)dx [g(x) exp f(x)dx+K]

Upon evaluation of the integrals we obtain:

$$D = \exp(-\frac{k_L a}{v} z)[\frac{k_r L_0}{k_L a - k_r} \exp(\frac{k_L a - k_r}{v} z) + K] \qquad \_\_\_\_\_\_\_\_\_\_(2.7)$$

Using the boundary condition D=D₀ at z=0 to evaluate the integration constant K finally yields:

$$D = (D_0 - \frac{k_r L_0}{k_L a - k_r}) \exp(-\frac{k_L a}{v} z)[\frac{k_r L_0}{k_L a - k_r} \exp(\frac{k_L a - k_r}{v} z) + K]$$

$$\_\_\_\_\_\_\_\_\_\_(2.8)$$

This expression (2.8) is the oxygen deficit profile in the river shown in the figure 2.1.

2.3. River water quality modeling Concept

Rivers are vital and vulnerable freshwater systems that are critical for the sustenance of all life. However, the declining quality of the water in these systems threatens their sustainability and is therefore a cause for concern. This subsection gives an insight about the general concepts of water quality modelling. First, the River Continuum Concept (RCC) is presented, and then the general water quality modeling methods including types, terminologies and comparison of different modelling approaches are presented.

2.3.1. River Continuum Concept:

Rivers are longitudinal functional units whose flow is unidirectional, as explained by the 'river continuum' concept (Davies & Day, 1998). This concept considers rivers as characterised by physical and chemical conditions that are progressively and continuously modified downstream from the headwaters to the sea (Davies & Day, 1998). Whilst the attributes and constituents of water vary naturally within and between rivers, almost all surface water bodies tend to exhibit some degree of natural intra and inter annual variation (King *et al*, 2003).

Understanding the theoretical framework of the river as an ecological continuum provides a fundamental basis for understanding ecosystem dynamics and water quality modeling strategies in rivers. The River Continuum Concept (RCC) was first introduced by Vannote *et al.* (1980) (see Figure 2.2). It provides an insight in the way biological communities may change from the headwater stream to larger rivers in the absence of human influence. Since then, the concept is widely used in river water quality assessment and modelling (e.g. Shanahan *et al.*, 2001; Carpenter, 2001). The physical basis of the RCC is the size of the river or stream (stream order) and location along the stream. According to the RCC concept, the biotic and abiotic structure and function of the running water is characterized by longitudinal, vertical and lateral gradients.

According to the concept of river continuum, longitudinally, the river system can be subdivided into three zones: the headwater stream (1st – 3rd order stream), mid-reach (4th – 6th order stream) and downstream regions (> 6th order stream). The headwater streams depend on the surrounding forest for energy and nutrients sources, where leaf-shredding macro invertebrates constitute a large portion of the macroinvertebrate population. As this part of the river is shaded by riparian vegetation, there is almost no algal growth and the main source of organic carbon is from allochthonous primary production, and heteroterophs dominate the microbial community. Consequently, the ratio of gross primary productivity (P) to community respiration (R), P/R, is less than one. The important compartments in this river section are the bulk water, the riverbed and the hyporheic compartment (the transition zone between surface water and groundwater).

In the mid-reach, the influence of riparian vegetation becomes less important. The amount of grazers that scrape algae from rock surfaces and that are nearly absent in head streams becomes larger.

The absence of the shading effect enhances the stream shift from heterotrophic to authotrophic, and the growth of attached algae or periphyton. Thus, P/R is larger than 1, and the main source of nutrient is from authochthonous (internal source) and allochthonous (external source) detritus. The important compartments are bulk water, riverbed and hyporheic compartment.

The downstream regions receive fine particulate organic matters that are washed downstream from the upstream sections of the river. The collectors that filter or gather fine particles from the stream become dominant. As the effect of riparian vegetation is insignificant, the main source of nutrients is authochthonous detritus. The primary production is often limited by depth and turbidity, and hence P/R is less than 1. In this river section therefore, the important compartments are bulk water and flood plain.

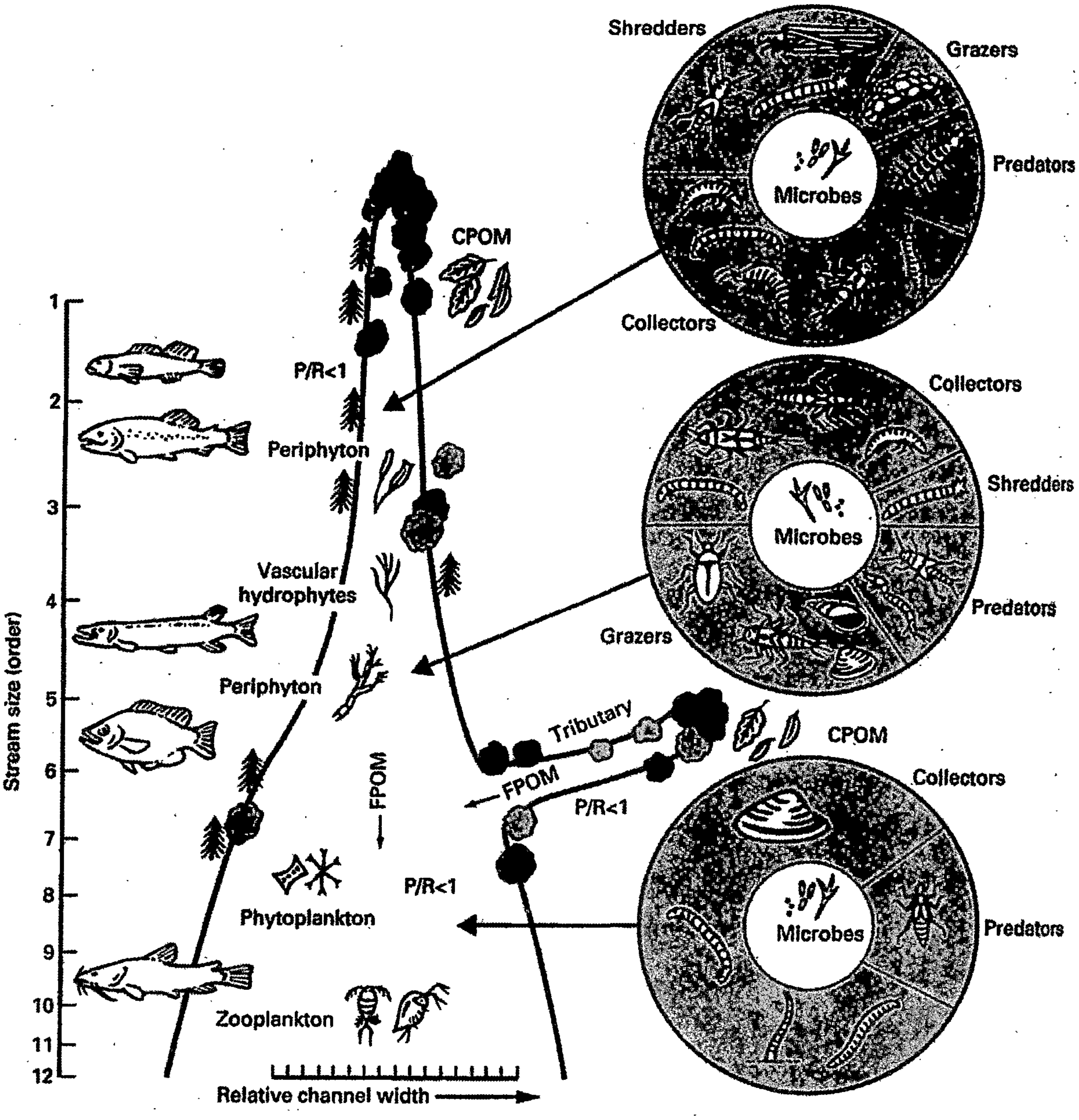

Figure 2.2: River Continuum Concept (RCC): relationship between the stream size and the progressive shift of structure and function of lotic communities (after Vannote *et al.*, 1980)

Impact of human activity and influence, the structure and function of river ecosystems become unpredictable with the RCC (Stout, 1994; Carpenter, 2001). The presence of disturbances such as nutrient enrichment, organic pollution and alteration of riparian vegetation through grazing, clear-cutting or impoundment may cause the overall continuum response to be shifted. The shift in structure and function of the river continuum indicates strong human influence and pollution of the river system, and thus the river needs restoration. In order to restore the polluted river, identification of the water quality problems and the sources of pollution are required, and this must be supported by water quality rules and regulations.

RCC clears the characteristics of the river ecosystem that should be considered in the river water quality modelling. The modelling approaches must address the ecological characteristics of the river by appropriate consideration of compartments (bulk water and benthic sediment) and processes description in representative spatial and temporal scales (Shanahan *et al.*, 2001). For example, in the headwater and middle stream with coarse substrates, it is necessary to consider both suspended and benthic compartments and their processes. This means the model should include the activities of microbial biomass both in suspension and attached to the benthic sediment. The activities of attached bacteria and algae predominate the activities of suspended microbial biomass, especially in small rivers where the wetted surface area to volume ration is large. In large rivers or in the downstream region of a river on the other hand, the suspended bacteria and algae in the water column dominate the conversion rates.

2.3.2. Water quality model types and terminologies:

Types of water quality models and related terminologies are well documented in Carstensen *et al.* (1997). Basically, there are two types of water quality models called stochastic and deterministic. If the model contains elements of randomness it is called stochastic. Including randomness in a model can be considered in order to account for the uncertainty associated with the model input variables and parameter values and model structure. Thus, a stochastic model will generate a range of values (rather than a single one) as model output in the form of a frequency distribution of e.g. pollutant concentration.

If the model contains no elements of randomness or does not comprise uncertainty, the model output is a single value. This type of model is termed deterministic. A deterministic model can be further described as mechanistic (white-box), grey-box and black-box model. Mechanistic (white-box) models are based on physical, biological and chemical laws, whereas the black-box models are those models that are not based on any physical or biological laws; instead they are based on a data driven transfer function, e.g. a neural network (Lek *et al.*, 1999). If a model contains elements of both the white-box (mechanistic submodel) and the black-box model, the model is called grey-box.

Mechanistic models may vary in their model components, level of representation and temporal representation. On the basis of the level of representation, the mechanistic models vary from very simple (lumped models) to very complex (distributed models). In the lumped models, several processes may be combined and expressed as one, whereas in the distributed model, the model attempts to represent every significant process.

Other important concepts that must be considered in water quality modelling are related to the model components. There are three basic components: variables, constants and parameters. The model inputs (components that influence the system), outputs (the one that one wants to predict) and states (the components that are required to be known to calculate the output) are all variables, and therefore called input variables, output variables and state variables. The state variables can be all variables that must be calculated based on the other variables. Model components that do not change their values throughout all possible applications are termed constants, e.g. the gravity constant and unit conversion constants (day to seconds; mg to kg, etc). Model components that change their value according to the application are called parameters. The value of the parameter can be related to time, location or input variables, e.g. bacterial maximum growth rate, temperature correction factors, etc.

Concerning the temporal representation of the model, the distinction should be made between steady-state and dynamic (unsteady-state) models. In steady-state models, all inputs and state variables are constant in time. In dynamic models however, inputs variables and state variables may vary with time, and thus result in a time variable output.

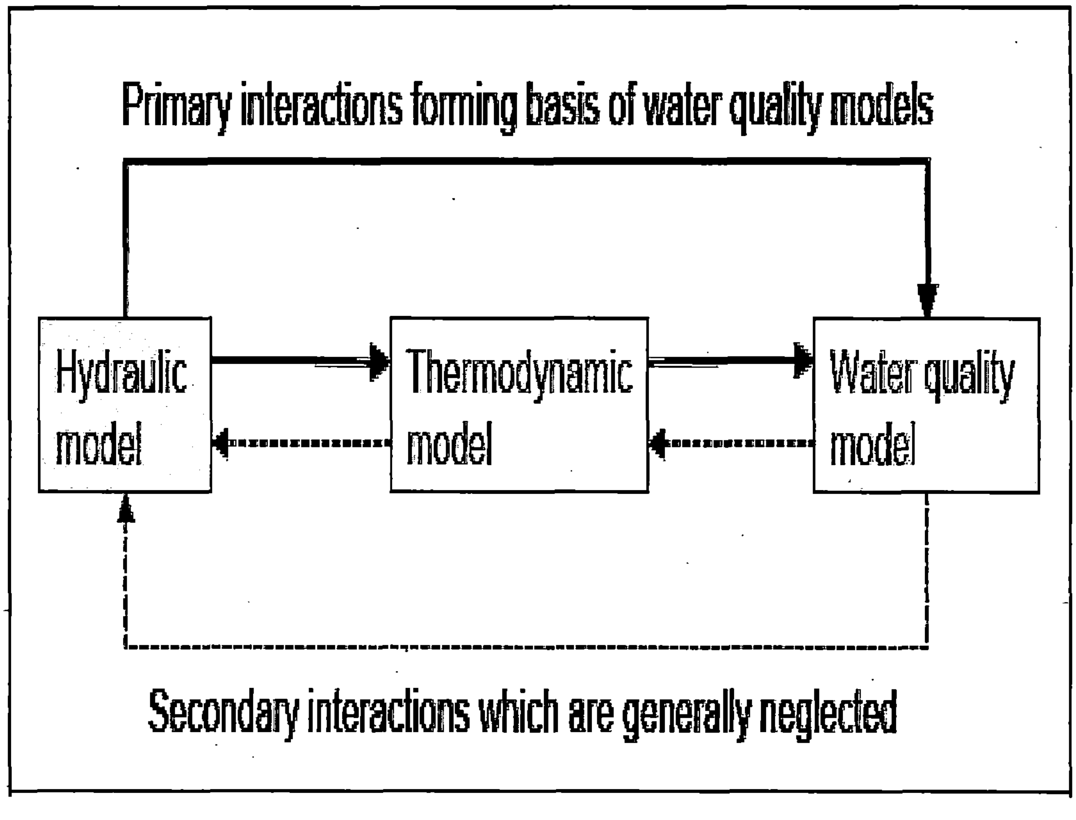

Figure 2.3 The basic sub-models which make up a water quality model (Adapted from Thomann and Meuller 1987)

2.4. Dynamic mechanistic river water quality modeling

The conservation of mass the main component of mechanistic river water quality models, i.e. within a finite volume of water, mass is neither created nor destroyed. In quantitative terms, the principle is expressed as mass-balance equations that account for all transfers of matter across the system's boundaries and all transformations occurring within the system.

Due to various factors the pollutant loading into the river and the water quality variables (e.g. the concentration of dissolved oxygen) may vary such as storm events and sewer overflows. A dynamic mechanistic river water quality model therefore takes into account such temporal variability of pollutant loading and water quality variables in rivers.

In the following subsection dynamic mechanistic river water quality modelling approaches that are applied in both basic water quality and organic contaminant fate and effect modeling are discussed under these subsections: (1) complex hydraulics routing, (2) complex pollutant transport, (3) conceptual hydraulic routing, (4) conceptual pollutant routing,

2.4.1. Complex hydraulic routing

Various models available to simulate dynamic water movement (flood propagation) in rivers are based on the usage of the St. Venant equations which comprise the mass and momentum balances on a one-dimensional channel. When the wind shear and eddy losses are omitted, the model is as follows:

Continuity equation including lateral inflow (mass balance):

$$\frac{\partial Q}{\partial x}+\frac{\partial A_{cross}}{\partial t}=q$$

Momentum equation (momentum balance):

$$\underbrace{\frac{1}{A_{cross}}\frac{\partial Q}{\partial t}}_{\substack{\textit{Local}\\ \textit{Acceleration}}}+\underbrace{\frac{1}{A_{cross}}\frac{\partial}{\partial x}\left(\frac{Q^2}{A_{cross}}\right)}_{\substack{\textit{convective}\\ \textit{acceleration}}}+\underbrace{g\frac{\partial h}{\partial x}}_{\substack{\textit{pressure}\\ \textit{force}}}-g(\underbrace{S_o}_{\substack{\textit{gravity}\\ \textit{force}}}-\underbrace{S_f}_{\substack{\textit{friction}\\ \textit{force}}})=0 \qquad \text{-----------------------(2.9)}$$

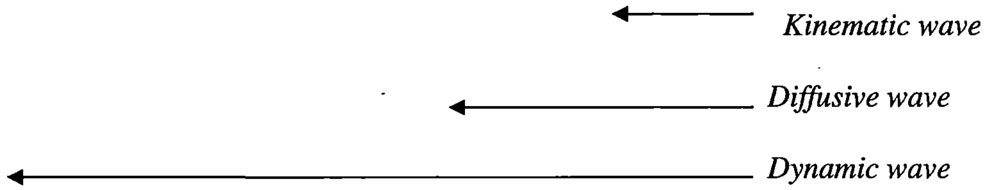

Figure 2.4: Simplification of momentum equation by dropping terms (after Chow, 1988)

Where

Q	flow rate [m^3 s^{-1}]
A_{cross}	cross-sectional area [m^2]
h	absolute elevation of water level from the datum [m]
g	gravitational acceleration constant [m^2 s^{-1}]
q	lateral inflow per unit length [m^2 s^{-1}]
S_o	river channel side slope [-]
S_f	friction slope [-]
x	longitudinal distance of the river [m]

The momentum balance in Figure 2.4 can be simplified by dropping terms (Chow, 1981).

the pressure and acceleration terms are when dropped or only the friction term and gravity force are considered, the equation describes the kinematical wave only, which is limited to the monotonically decreasing of the riverbed. When one ignores the variation of flow, the equation is simplified to the diffusive wave approximation, which allows describing backwater effects of weirs or other hydraulic controls like tidal effects. It can be applied when the river is not monotonically decreasing. If no term is ignored, the dynamic wave equations are able to describe the full dynamic wave.

By considering constant water width and rearranging equation 2.9 and equation indicated in Figure 2.2, one can get the St. Venant equations as follows:

$$\frac{\partial h}{\partial t} = -\frac{1}{b}\frac{\partial Q}{\partial x} + \frac{q}{b} \qquad \text{---2.10}$$

$$\frac{\partial q}{\partial t} = -\frac{\partial}{\partial x}\left(\frac{Q^2}{A_{cross}}\right) - gA_{cross}\frac{\partial h}{\partial x} + gA_{cross}(S_O - S_f) \text{--------------2.11}$$

where b is the river water width [m].

2.4.2. Complex pollutant transport (mass balance) modeling

Pollutant routing in a river is often described by the advection-dispersion (mass balance) equation, which is based on the principle of conservation of mass and Fick's law. The mass balance for non-conservative (non-reactive) pollutants in three directions (longitudinal x, vertical y and lateral z) is written as:

$$\frac{\partial C}{\partial t} = -u_x\frac{\partial C}{\partial x} - u_y\frac{\partial C}{\partial y} - u_z\frac{\partial C}{\partial z} + E_x\frac{\partial^2 C}{\partial x^2} + E_y\frac{\partial^2 C}{\partial y^2} + E_z\frac{\partial^2 C}{\partial z^2} - R \text{--------2.12}$$

where

C	concentration [g m^{-3}]
T	time [s]
$u_{x,y,z}$	average velocity in the ith direction [m s^{-1}]
x, y, z	distances in x, y and z directions [m]
$E_{x,y,z}$	the dispersion coefficients in the x, y and z directions [m^2 s^{-1}]
R	reaction transformation rate [g m^{-3} s^{-1}]

Equation 2.12 is also called the basic water quality equation. It assumes both the flow velocity and dispersion coefficient to be constant over all the three directions. It cannot be solved alone in itself but requires the application of a hydraulic model as an input.

The full-form of equation 2.12 is rarely applied in river water quality studies, as it requires a lot of data for the three-dimensional velocity fields. It is hence often applied in a simplified form.

Assuming the absence of temporal velocity gradient in both vertical and lateral direction, pollution routing for one-dimensional river water quality can be expressed as follows:

$$\frac{\partial(A_{cross}C)}{\partial t}=-\frac{\partial(QC)}{\partial x}+EA_{cross}\frac{\partial^2 C}{\partial x^2}-A_{cross}R \text{------------------------}2.13$$

To solve equation 2.13 numerically, it is coupled to the numerical solution of open channel flow such as provided for the St. Venant equations (equations 2.10 and 2.11).

The St. Venant equations require numerical methods (typically finite difference and finite element method) to solve them. These methods require small time steps to overcome the numerical problem of instability. As the application of the full St. Venant equations already requires long computation times, further extension of this model towards integrated water quality modelling will result in even more computation time.

Consequently, a conceptual mechanistic surrogate model was proposed for the sake of faster simulation and easy implementation of water quality models (e.g. Meirlaen *et al.*, 2001).

Mathematical Modeling related to Flow Analysis:

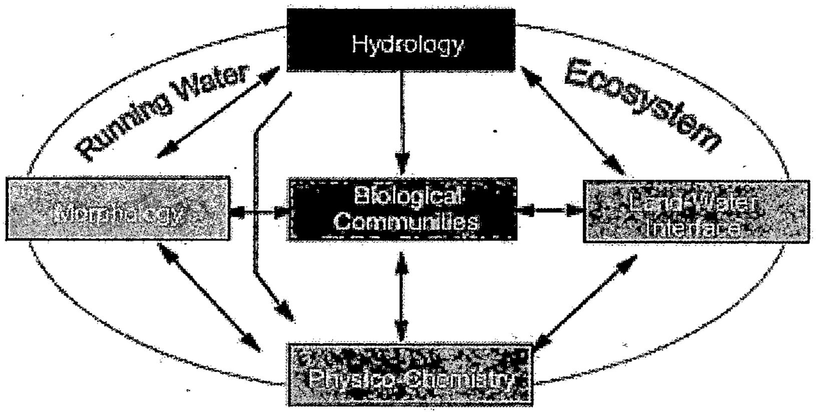

Fig 2.5 Structure of running water ecosystems

2.4.3. Hydraulic routing (Conceptual Model)

The conceptual model consisting of a cascade of Continuously Stirred Tank Reactor in Series (CSTRS) can be applied for dynamic hydraulic modelling in rivers. The schematic representation of this modelling approach is indicated in Figure 2.6. It is based on the mass balance of water that can be expressed around a control volume, an incremental element (slice) of stream volume, as follows:

$$\frac{dv}{dt} = Q_{in} - Q_{out} \text{ --------------2.14}$$

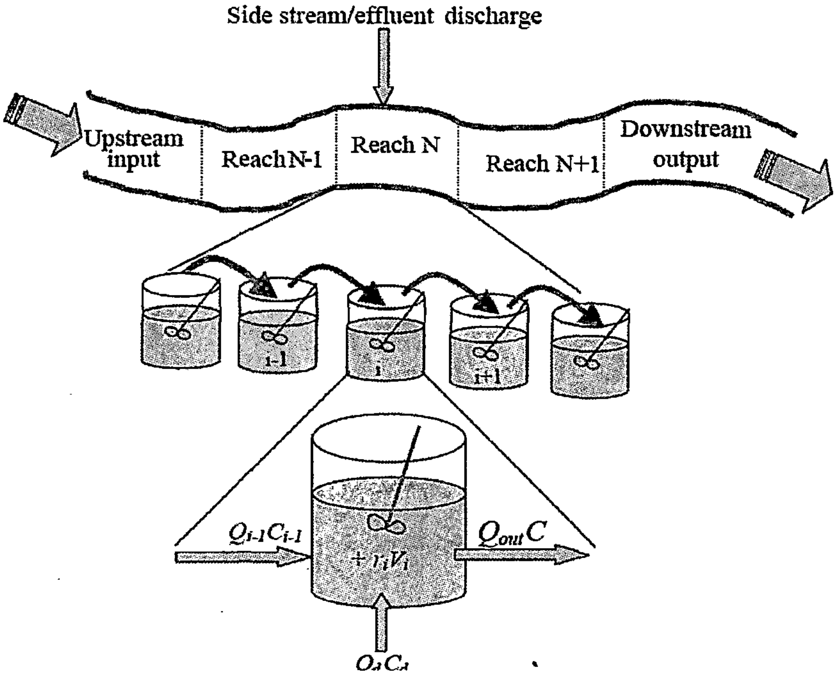

Figure 2.6: River discretisation, a cascade of CSTRS model and mass balance

where V control volume [m^3]

Q_{in}	inflow rate [$m^3\ d^{-1}$]
Q_{out}	outflow rate [$m^3\ d^{-1}$]

Around a control volume, box or river tank, a side stream or effluent discharge can be connected. In that case, equation 2.14 can be extended as follows:

$$\frac{dv}{dt} = Q_{in} + Q_d - Q_{out} \text{ ------2.15}$$

where Q_d is the flow rate of the side stream or effluent discharge [$m^3\ d^{-1}$].

Equations 2.14 and 2.15 are based on simple mass balance where the change in volume over time is the difference between the overall inflow rates and outflow rates. It respects the continuity equation, equation 2.9. The outflow rate Q_{out} can be calculated in different ways, e.g. Shaw (1996), but the simplest way is as indicated in equation 2.16. The parameters of the power function a and b can be calculated on the basis of stage-flow data on available relationships, as follows:

$$Q_{out} = \alpha h^{\beta} \quad \text{---------2.16}$$

Equation 2.16 is conceptually equivalent to the momentum conservation.

The underlying concept of the conceptual hydraulic model is that the river must be discretised (segmented) into a series of tanks (control volume). The only limitation of this formulation, however, is that it does not simulate backwater-effects due to dams, weirs or tidal effects. The conceptual hydraulic model can, however, be applied for dynamic flow propagation in nontidal influenced rivers.

2.4.4. Pollutant routing (Conceptual Model)

A cascade of CSTRS is among the simplest systems that can be used to model a natural water body (Chapra, 1997). This conceptual approach is also called moving segment model, box model or cells in series model, and has been successfully applied in river water quality modelling (Beck and Reda, 1994; Park and Lee, 1996; Lewis *et al.*, 1997; Park and Lee, 2002).

On the basis of equation 2.14, a pollutant mass balance for unsteady flow conditions in every CSTRS can be expressed as follows:

$$\frac{d(VC)}{dt} = Q_{in}C_{in} - Q_{out}C + rV \quad \text{------------------------------------}2.17$$

Where

C_{in} concentration in the inflow [g m^{-3}]
C concentration in the outflow [g m^{-3}]
r the overall reaction rate [g m^{-3} d^{-1}]

If the side stream and wastewater effluent discharge is connected
equation 2.9 can be extended as follows:

$$\frac{d(VC)}{dt} = Q_{in}C_{in} + Q_dC_d - Q_{out}C + rV \quad \text{-----------------------------}2.18$$

where C_d is the concentration in the side stream or effluent discharge [g m^{-3}].

Equations 2.17 and 2.18 are based on a simple mass balance and relatively easy to apply. They can be used for both steady-state and dynamic flow conditions, where for the dynamic flow conditions, the derivative term must be solved numerically.

2.5 Dynamic river water quality modeling (Simplified form)

A simplified dynamic river water quality model is discussed in this chapter. It was derived from the already existing complex river water quality model, which was principally developed for data rich conditions. The simplified model is meant to be applied in data limited situations, as is the case in developing countries, as well as for integrated river water quality modelling. The model was applied on the Crocodile River case study (South Africa) in order to investigate the seasonal dynamics of nitrate and ammonia nitrogen concentrations. Its application was evaluated using monitoring data collected during the years 1987 to 1990. The relationship between river flow rates and inorganic nitrogen concentrations was analyzed. The sensitivity of the model output to changes of model input parameters is discussed.

2.5.1. Introduction

In developing countries the challenge of using mathematical models as a support tool to evaluate water quality remediation options is well documented (Ongley and Booty, 1999). Modelling is expensive, requires substantial investment in reliable data, development of scientific capacity and a relatively sophisticated management culture. Nevertheless, new developments in water quality management policies and strategies require a mathematical model to predict the in-stream fate of pollutants as well as to estimate the likely effects that the resultant water quality may have on recognized water uses. Furthermore, the complex relationships between waste load inputs, and the resulting water quality responses in receiving water bodies are best described using mathematical models.

Several types of river water quality models for conventional pollutants (e.g. organic and inorganic nutrients) are available. These models simulate the major reactions of nutrient cycles and their effect on the dissolved oxygen balance. The complexity and number of state variables of these models increase from the simplest Streeter-Phelps (Oxygen sag curve) (Streeter and Phelps, 1925) to extended models such as *QUAL1* (Masch and Associates, 1970), *QUAL2* (Water Resource Engineering, 1973), and *QUAL2E* (Brown and Barnwell, 1987). Currently, *QUAL2E* is the most widely available stream water quality model that has been adapted for use on a personal computer. However, the *QUAL2E* model types were indicated to have some limitations. One of the major limitations is the lack of provision for conversion of algal death to Biological Oxygen Demand (BOD) (autochtonous source of organic matter), denitrification, and Dissolved Oxygen (DO) change caused by fixed plants (Ambrose *et al.*, 1996; Park and Lee, 2002). Furthermore, these models take into account neither suspended nor attached microbial biomass as state variables despite the fact that microbial biomass are the main component in the biotransformation processes (Reichert *et al.*, 2001). They are also using Biological Oxygen Demand (BOD) as a measure of carbonaceous organic matter, which cannot be fractionated properly to different phases (dissolved and particulate) of organic carbon, and therefore not suitable to calculate mass balance.

An integrated water quality modelling approach requires a river water quality model that can be connected easily to, and is compatible with, a typical Activated Sludge wastewater treatment plant Models (ASM) (Reichert *et al.*, 2001). In contrast to QUAL2E type models, ASMs are based on Chemical Oxygen Demand (COD) as a measure of carbonaceous organic matter, and also consider microbial biomass as state variable. Accordingly, the IWA Task Group on River Water Quality Modelling recently proposed the River Water Quality Model number 1 (RWQM1) (Reichert *et al.*, 2001). This model considers microbial biomass as state variables, and it is also based on COD. It is thus compatible with the existing IWA Activated Sludge Models: *ASM1* (Henze *et al.*, 1987), *ASM2* (Henze *et al.*, 1995) and *ASM3* (Gujer *et al.*, 1999). RWQM1, however, is considered to be too comprehensive and complex to be applied directly in many situations, as is the case in developing countries where the availability of data is very limiting. In data poor situations, one needs to focus on a simple river water quality model that describes components of the C, O, N and P cycles reasonably and which is still compatible with Activated Sludge Models. In this study, a simple dynamic river water quality model was developed based on a simple conceptual hydraulic model and the simplification of the available complex RWQM1 model (Reichert *et al.*, 2001).

2.6 Basic river water quality models

2.6.1 Introduction

Rivers are waterways of strategic importance across the world, providing main water resources for domestic, industrial, and agricultural purposes. This is particularly true for India, a country that relies heavily on surface water resources to sustain its growing population and economy. So, water is life and the quality of water is an essential measure of the quality of life or rather the existence of life. Consequently water quality management is (or should be) one of the most important activities of mankind, so as to protect and save human life and the life of other living things, which latter is a precondition of human life as well. The management of water quality, or the protection of the aquatic ecosystem in a broader sense, means the control of pollution. Water pollution originates from point and non-point (diffuse) sources and it is always due to human action (the author strongly believes that no such thing as "natural pollution" exists, as sometimes advocated by other people). The control of water pollution, the protection of aquatic systems, is thus the control of human activities that result in pollution. One should also understand that the protection of the aquatic environment, and within this the control of pollution, is a profession and not an easy one. Professions like designing a house, a bridge, a road or just the making of a pair of shoes. This also means that no bridge designers (or hydraulic engineers) and no shoemakers and not even water chemists and aquatic ecologists can alone attempt the solving of water pollution control problems (although sometimes they think they can). A crucial element in the series of complex activities of planning and implementing water pollution control actions is the quantitative determination and description of the cause-and effect relationships between human activities and the state (the response) of the aquatic system, its quantity and quality. These activities together can be termed the modeling of aquatic systems (hydrological, hydraulic and water quality modelling). These activities are aimed at calculating the joint effect (the impact) of natural and anthropogenic processes on the state of water systems.

The subject of this chapter is to introduce the basics of water quality modelling to the user. Although the qualitative and quantitative modelling of water systems (rivers, lakes and reservoirs) should be done simultaneously we will have to separate them for the purpose of this programme, always assuming that the quantitative state (the hydrological and hydraulic parameters) of the water system is known and sufficiently well described. With this one can focus on the quantitative, mathematical, description of processes that affect water quality.

Even within water quality modelling one can go to deal with the most essential basics of river modeling which would deal with more details of it (including basic statistics of flow and quality data), lake and reservoir modelling and last but not least with the modelling of non-point source pollution, a crucial problem of ever growing importance of our era.

2.6.2 Theory of river water quality models

In logical order the teaching of this topic should have started with the description of both the quantitative and qualitative state of the water body. The more so since even the basic flow modelling techniques would fill a separate curriculum in itself. Consequently in the following sections of this programme all, hydraulic and hydrological river parameters (e.g rate of flow, flow velocity, stream depth and width, etc) will be considered as given input data. Thus we will start with the introduction of the basic mass transport and transformation processes, relying on continuity and conservation of mass considerations.

Skipping again some of the details of deriving the basic equation (Jolankai 1979, Jolankai,1992) let us consider an elementary water body, a cube of dx, dy and dz dimensions. The quality of water within this elementary water body depends on the mass of a polluting substance present there. Water quality models then should describe the change of the mass of a polluting substance within this water body. The change of the mass of this substance is calculated as the difference between mass-flows (mass fluxes) entering and leaving this water body, considering also the effects of internal sources and sinks of the substance, if any. The mechanism of mass transfer into and out of this water body includes the following processes:

Mass transported by the flow, by the v_X, v_Y, and v_Z, components of the flow velocity vector. This process is termed the advective mass transfer. The transfer of mass, that is the mass flux (in mass per time, $M\ T^{-1}$, dimension) can be calculated in the direction x as $C*v_X,*dy*dz$, where C is the concentration of the substance in the water (in mass per volume dimension, M L), see also Equation 2.19. The other means of mass transfer is termed the dispersion or dispersive transport. Here one has to explain this term because there is usually considerable confusion with the terms diffusion and dispersion; -in short: dispersion is a term used for the combined effect of molecular diffusion and turbulent diffusion, and both of these latter processes is caused by pulsating motion, that is by the "Brownian" thermally induced motion of the molecule (molecular diffusion), and by the pulsation of the flow velocity around its mean value, caused by turbulence (called the turbulent diffusion).

The dispersive mass transfer (E_X, E_Y, E_Z) has the dimension of mass per time per area (M T^{-1} L^{-2}) and it is usually expressed by the law of Fick which states that the transport of the substance in a space direction is proportional to the gradient of the concentration of this substance in that direction the proportionality factor being the coefficient of dispersion, as shown in equation 2.19.These equations describe the dispersive and advective transport of a polluting substance from the x direction into an elementary water body. The first term is actually the law of Fick which states that the diffusive (dispersive) transport of the substance in a space direction is proportional to the gradient of the concentration of this substance in that direction the proportionality factor being the coefficient of dispersion. The user finds more information on dispersion in the "general" part of this basic theory chapter and on the programme part on "dispersion river models". The second term is the advective transport term, which states that the specific (per unit area) transfer of mass to a spatial direction is the product of the concentration of a substance and the velocity of flow in that spatial direction. These are the terms used in writing the overall mass balance (that is Eq. 2.20) of an elementary water body.

$$E_x = D_x \frac{dC}{dx} \quad ; [ML^{-2}T^{-1}]$$

$$ADV_x = C\, v_x \quad ; [ML^{-2}T^{-1}]$$

-------------------------------- 2.19

where

E_X – is the dispersive mass flux in the spatial direction x ($M\ L^{-2}\ T^{-1}$), with the assumption that the law of Fick holds for the joint effect of molecular diffusion and turbulent diffusion, that is for dispersion .

ADV_x is the advective mass flux in the spatial direction x (iM $L^{-2}\ T^{-1}$).

C - is the concentration, the mass of the quality constituent in a unit volume of water (mass per volume, $M\ L^{-3}$);

D_x - is the coefficient of dispersion in the direction of spatial coordinate x (in surface area per time, L^2T^{-1} units);

V_X - is the component of the flow velocity in spatial directions x. (length per time, $L\ T^{-1}$);

2.6.3 The mass-balance equation of an elementary water body

This equation was derived by writing a mass balance of in- and outflowing advective and dispersive mass fluxes of an elementary water body and expressing the change of the mass of the substance with time. The terms for one spatial direction include the inflowing mass flux and the outflowing mass flux, which latter is the difference between inflowing flux and the change of the flux within the water body. For more details see the “General description of basic theory”, the “mass transport terms for deriving the basic model” and the “General description of dispersion river models”.

$$\begin{aligned}\frac{\partial C}{\partial t}dxdydy = {} & \left[(v_x C)+E_x\right]dydz+\left[(v_y C)+E_y\right]dxdz+\left[(V_z C)+E_z\right]dxdy \\ & -\left\{(v_x C)+E_x+\frac{\partial}{\partial x}\left[(v_x C)+E_x\right]dx\right\}dydz- \\ & -\left\{(v_y C)E_y+\frac{\partial}{\partial y}\left[(v_y C)+E_y\right]dy\right\}dxdy- \\ & -\left\{(v_z C)+E_z+\frac{\partial}{\partial z}\left[(v_z C)+E_z\right]dz\right\}dydx\end{aligned}$$

--2.20

Where

C - is the concentration, the mass of the quality constituent in a unit volume of water (mass per volume, $M\,L^{-3}$);

E_X, E_Y, E_Z - are the dispersive mass fluxes in the spatial directions x, y, and z (in $M\,L^{-2}\,T^{-1}$ dimension), with the assumption that the law of Fick holds for the joint effect of molecular diffusion and turbulent diffusion, that is for dispersion;

V_X, V_Y, V_Z - are the components of the flow velocity in spatial directions x, y , and z, (length per time, $L\,T^{-1}$);

d_x, d_y, d_z - are the side lengths of an elementary cube, an elementary water body.

2.6.4 The basic water quality model equation

This equation forms the basis of all water quality models. It was derived from Equations 2.19 and 2.20, by combining them, carrying out the operations, rearranging the result and dividing the equation by the elementary water volume dx*dy*dz and also by considering internal sources and sinks of the substance, as well as external sources. The basic equation describes the variation of the concentration of a quality constituent C with the time and space. Apart from the advective and dispersive transport terms that were discussed in relation to Equations 2.19 and 2.20 in this basic equation there is a general term, the internal source/sink term, or internal reaction term, that should be also discussed in somewhat more detail. They are also called the transformation processes with the meaning that the substance in concern is being transformed by various physical, chemical, biochemical and biological processes resulting in the change of the quantity of the substance in an elementary water body. This change is either a "loss" or sink term caused by processes such as settling, chemical-biochemical decomposition, uptake by living organisms or a "gain", a source term, such as scouring from the stream bed, product of chemical-biochemical reactions, biological growth, that is the "build-up " of the substance in concern on the expense of other substances present in the system. The actual form of these transformation processes will be presented in relation to concrete model equations such as the BOD-DO models, the models of the oxygen household.

$$\frac{\partial C}{\partial t}+v_x\frac{\partial C}{\partial x}+v_y\frac{\partial C}{\partial y}+v_z\frac{\partial C}{\partial z}=$$

$$=\frac{\partial}{\partial x}\left(D_x\frac{\partial C}{\partial x}\right)+\frac{\partial}{\partial y}\left(D_y\frac{\partial C}{\partial y}\right)+\frac{\partial}{\partial z}\left(D_z\frac{\partial C}{\partial z}\right)+S(x,y,z,t)\pm S_{\mathrm{internal}}$$

--2.21

Where

C - is the concentration, the mass of the quality constituent in a unit volume of water (mass per volume, M L^{-3});

D_x, Dy, Dz- are the coefficient of dispersion in the direction of spatial coordinates x,y and z (in surface area per time, L^2T^{-1} units);

t- is the time(T):

S(x,y,z,t)- denotes external sources and sinks of the substance in concern that may vary both time and space (mass per volume per time, $ML^{-3}T^{-1}$);

$S_{nternal}$ – denotes the internal sources and sinks of substance ($ML^{-3}T^{-1}$);

2.6.5 Derivation of simple water quality model from the basic model equation

The basic three dimensional water quality model is seldom used in its original complex way (Eq. 2.21), mostly because three dimensional problems occur rarely. For example river problems can be frequently reduced to one-dimensional (linear) or two dimensional (longitudinal-transversal) problems, as it will be demonstrated in the programme. Another reason of using simplified models is that transversal or vertical velocity measurement data are seldom available.

The internal source-sink terms that were only denoted in Eq. 2.21 should be specified for each problem explicitly and they vary with the components considered. Here it will be briefly demonstrated how can one derive the most simple (river) models version of Eq. 2.21, which can be used in the practice. In order to arrive to the possible most simple water quality model we have to make first series of assumptions and approximations: a, Neglect, for the time being, all terms accounting for dispersion. With this we assume that the system is fully mixed, which means that any external material input (load) to the river will be instantaneously and fully mixed with the water. This is a very rough approximation and its consequences will be discussed in a subsequent sections dealing with dispersion and mixing problems. However, this approximation holds with long linear systems, e.g in the case of smaller rivers with continuous steady input loads (waste water discharges).

b, Considering a river and a sewage discharge of steady state conditions (with flow not varying in time) the initial concentration Co downstream of an effluent outfall can be described by the general dilution equation (see Equation 2.22).

2.6.5.1 The general dilution equation

Considering a river and an effluent discharge of steady state conditions (with flows and concentrations not varying in time) and assuming instantaneous full cross-sectional mixing of the sewage water with the river water the initial concentration Co downstream of an effluent outfall can be calculated by the dilution equation (Eq. 2.22), which stems from the balance equation of in- and outflowing fluxes written for the section of the discharge point (e.g. background river mass flux plus pollutant discharge mass flux equals the combined mass flow downstream of the point of discharge). This equation is used very frequently in simple analytical water quality models for calculating the initial concentration of pollutants.

$$C_0 = \frac{C_s q_s + C_b Q_b}{q_s + Q_b} \qquad \text{------------------------------------2.22}$$

Where

C_b – background concentration of the polluting substance in concern in the river, (ML^{-3});

Cs - concentration of the pollutant in the waste water, (ML^{-3});

Q- discharge (rate of flow) of the river upstream of the effluent outfall, ($L^3\ T^{-1}$);

q- the effluent discharge, ($L^3\ T^{-1}$);

2.6.5.2 The most simple water quality model (2)

Averaging flow and concentration over the cross section Equation 2.21 simplifies into Equation 2.23 where v is the average flow velocity along the stream. Introducing the "time of travel" $t = x/v$ and assuming first order reaction kinetics for a single decay or decomposition process, as the only internal process (sink) one obtains the possible most simple river water quality model in the form of Equation 2.24. This equation (the principle of first order reaction kinetics) states that the decay/decomposition of a pollutant is proportional to the concentration of the pollutant and the factor of proportionality is K, the decay rate coefficient (T^{-1}). Solving Eq. 2.24 for the initial conditions defined above ($C = C_0$ at $x=x_0$, that is $t = t_0$ the simple exponential decay equation (Equation 2.25) is obtained, which is at the same time the most simple water quality model used in the practice. Equation 2.25 will be subsequently referred to also as the "Decay Equation". This equation can be used for a number of water quality 'modelling purposes (such as the "decay" of BOD, COD, etc, see also at the description of ' BOD-DO models), and forms an essential part in developing coupled reaction models (see under this heading for more details).

$$v\frac{dC}{dx} = \pm S_{internal}$$

---------------------------------------2.23

$$\frac{dC}{dt} = -KC$$

---------------------------------------2.24

$$C = C_0 e^{-Kt}$$

---------------------------------------2.25

Where

C – is the concentration, the mass of quality constituent in a unit volume of water (mass per volume, ML^{-3});

C_0- is the initial concentration of the pollutant downstream of a point source of pollution (ref.eq.1.4);

v- is the mean flow velocity of a river reach investigated (LT^{-1});

$S_{internal}$ – denotes the internal sources and sinks of substance, ($M\ L^{-3}\ T^{-1}$);

K- is the reaction rate coefficient for first order kinetics(T^{-1})

t- is the time of travel interpreted as $t=x/v$

x- the distance downstream (L)

2.6.5.3 Derivation of coupled reaction models

Chemical, biological or biochemical processes to which water quality constituents are subjected seldom occur alone but in a coupled way. If we consider such a coupled process situation, still in a generalizable way, assuming that the product of a decomposition/decay process of a water quality component (C_1) is another water quality constituent (C_2) which latter is subjected to further decay/decomposition then we can derive a simple set of coupled reaction models in the form of Equations 2.26 and 2.27, where K_1 and K_2 are the respective reaction rate coefficients of the not yet named water quality processes. With this we have actually derived the still most frequently used basic river model, the oxygen sag curve model (Streeter and Phelps, 1925). Assuming that the parameter C_1 is the biologically decomposable organic matter content of the water (expressed in Biochemical Oxygen Demand, BOD which is the amount of oxygen utilized by microorganisms from a unit volume of water for the decomposition of organic matter during a selected period of time) and assuming that the other parameter C_2 is the oxygen deficit compared to saturation level Eq 2.26 and 2.27 are the basic equations of the traditional oxygen sag curve model which states that the oxygen consumed by microorganisms adds to the oxygen deficit, while the process of aeration (or reaeration; the uptake of oxygen across the water surface due to turbulence and molecular diffusion) reduces this deficit.

Here the reaction rate coefficients gain specific meaning, that is

K_1- is the rate coefficient of biochemical decomposition of organic matter (T^{-1})

K_2 - is the reaeration rate coefficient (T^{-1})

t- is the time, that is the time of travel in the river interpreted as t=x/v, where

x is the distance downstream of the point of effluent discharge

The set of differential equations (Eq 2.26 and 2.27) can be solved for initial conditions $C_1 = C_{1,O}$ and $C_2=C_{2,0}$ at x=0; ($t=t_0$) (to be calculated with the dilution equation (Eq 2.20) in a similar way as shown there), obtaining Equations 2.28 and 2.29. Equation 2.29 is termed in the relevant literature the "Oxygen-Sag Equation" and will be referred to as such later on.

At this point we have arrived to a model which is actually used in the practice (along with its more or less modified, expanded, versions) as it will demonstrate in the subsequent parts of the programme.

$$\frac{dC_1}{dt} = -K_1C_1 \qquad \text{-----------------------------2.26}$$

$$\frac{dC_2}{dt} = K_1C_1 - K_2C_2 \qquad \text{-------------------------------2.27}$$

$$C_1 = C_{1,0}\, e^{-K_1t} \qquad \text{----------------------------2.28}$$

$$C_2 = \frac{K_1 C_{1,0}}{K_2 - K_1}\left(e^{-K_1 t} - e^{-K_2 t}\right) + C_{2,0} e^{-K_2 t} \qquad \text{-----2.29}$$

Where

$C_{1,}$, C_2 - are concentrations of interacting water quality constituents (the product of the "decomposition" process of C1 is C,, which latter is also a decaying decomposing constituent (ML^{-3})

C_{10}, C_{20} - are initial concentrations of the above two water quality constituents (see also Eq. 2.22) (ML^{-3})

K_1, K_2 - are the reaction rate coefficients of the above processes, (T^{-1})

t- is the time of travel interpreted as $t = x/v$, (T)

X - the distance downstream (L)

2.7 BOD-DO River Models

2.7.1 General introduction of BOD-DO river models

BOD-DO river models deal with the oxygen household conditions of the river, by considering some of the main processes that affect dissolved oxygen (DO) concentrations of the water.These models are of basic importance since aquatic life, and thus the existence of the aquatic ecosystem, depend on the presence of dissolved oxygen in the water.

All river water quality models, and thus the BOD-DO models, can be derived from the general basic water quality model equation (Eq. 2.19). For some details of this derivation procedure see the Chapter on Basic Theory.

The main process that affect (deplete) the oxygen content of water is the oxygen consumption of microorganisms, living in the water, while they decompose biodegradable organic matter.

This means that the presence of biodegradable organic matter is the one that mostly affect the fate of oxygen in the water. There are internal and external.sources of such biodegradable organic matter. Internal sources include organic matter that stem from the decay (death) of living organisms, aquatic plants and animals (also termed "detritus", or dead organic matter).

Among external sources anthropogenic ones are of major concern and this includes waste water (sewage) discharges and runoff induced non-point source or diffuse loads of organic matter.

In the models biodegradable organic matter is taken into consideration by a parameter termed "Biochemical oxygen demand, BOD". BOD is defined as the quantity (mass) of oxygen consumed from a unit volume of water by microorganisms, while they decompose organic matter, during a specified period of time. Thus BOD_5 is the five day biochemical oxygen demand, that is the amount of oxygen that was used up by microorganisms in a unit volume of water during five days "incubation" time in the respective laboratory experiment. Thus the unit of BOD is mass per volume (e.g $g0_2/m^3$, which equals mgO_2/l).

Another main process in the oxygen household of streams is the process of reaeration, the uptake of oxygen across the water surface due to the turbulent motion of water and to molecular diffusion. This process reduces the "oxygen deficit" (D) of water, which is defined as the difference between saturation oxygen content and the actual dissolved oxygen level.

These two counteracting processes are considered in the traditional BOD-DO model (Streeter and Phelps, 1925) in the mathematical form that can seen in the "Graph window" on chapter "The traditional BOD-DO model", the "Oxygen-sag curve".

2.7.2 The traditional BOD-DO model, the "oxygen-sag curve"

In this model the decomposition of biodegradable organic matter is expressed as the "first order" decay of BOD (termed here L) in function of the time (where time is the time of travel t =x/v) by Eq 2.30 and 2.31.
The oxygen line, the oxygen sag curve, is written for the oxygen deficit D is such a way that oxygen consumed by microorganisms adds to the oxygen deficit, while the process of aeration (or reaeration; the uptake of oxygen across the water surface due to turbulence and molecular diffusion) reduces this deficit (Equations 2.32 and 2.33).

In these equations the initial conditions, e.g $L = L_0$ and $D = D_0$ at x =0 ($t = t_0$) should be calculated using the "Dilution equation" (Eq 2.22). The substitution of waste water and river parameter values is relatively straight forward in the case of calculating L_0 (Eq. 2.34), while for calculating D_0 first the initial oxygen concentration should be calculated (Eq. 2.35) and the result of this should be subtracted from the saturation DO concentration to achieve D_0 (Eq.2.36).

The saturation dissolved oxygen concentration of the water is temperature dependent, and the respective values can be obtained either from tables published in the relevant literature or from experimental expressions. In this teaching aid we will use the latter method in the form of Equation 2.37 (Wang et. al, ref. Gromiec, 1983):

The oxygen sag curve has a critical point where the dissolved oxygen content of water is the lowest, that is when the oxygen deficit is the highest. The time of travel (or the corresponding downstream distance) can be expressed by finding the minimum of the sag curve. It is obtained in the form of Eq. 2.38 for t_{crit},, Eq. 2.39 for x_{Crit}, and Eq. 2.40 for D_{crit}. Thus the critical dissolved oxygen concentration is obtained as the difference between saturation oxygen concentration and the critical oxygen deficit (Eq. 2.41).
For the practical use of the above simple model equations one should find, estimate, the values of the two model parameters K_1 and K_2.
There are two basic ways of estimating values of the reaction rate parameters:

1. If one has in-stream measurement data of DO and BOD then one can calibrate the model, by fitting the calculated curves to the measured ones. This can be easily done for BOD (for K_1), expressing K_1 from Eq. 2.31; but the value of reaeration coefficient K_2 can be found only by trial-error model simulations.

2. If it has not access to measurement data then one can estimate model parameters using formulae and tables published in the relevant literature.

The value of the reaeration coefficient K_2 depends, eventually, on the hydraulic parameters of the stream and a large number of experimental formulae have been presented in the literature along with reviews of these literature equations (Gromiec, 1983, Jolankai 1979, 1992). These expressions deviate from each other, sometimes substantially. It has been developed a special equation on the basis of a number of literature published equations that give, the value of K_2 in function of flow velocity v and stream depth H, by simply averaging the coefficient values. Thus the obtained formula is Equation 2.42. For the estimation of the value of K_1 the Table 2.1 of Fair (ref. Jolankai, 1979) can be used, when knowing the value of K_2 can be used. This Table expresses the ratio $f= K_2/K_1$ in function of the verbally described hydraulic condition of the stream. Both the reaeration coefficient K_2 and especially the decomposition rate coefficient K_1 depend on the ambient (water) temperature. For this latter the most widely accepted formula is Eq. 2.43

One should note that reported literature values of K_1 and K_2 vary over wide ranges of which, for this teaching aid programme, we will consider the following domain:

K_1 - 0.1 - 1.7 day^{-1}

K_2 - 0.2 - 1.2 day^{-1}

If we discretize this domain at 0.1 day^{-1} steps one can obtain the variation of the f $=K_2/K_1$ ratio. From this table one should not adopt values of f lower than 0.5 or higher than 5 .0.

2.7.3 The BOD decay model

The BOD decay model describes the decomposition of biodegradable organic matter is expressed as the "first order" decay of BOD (termed here L) in function of the time (which is the time of travel along the stream t =x/v).

In Equation 2.31 the initial conditions, e. g $L = L_0$ at $x = 0$ ($t = t_0$ should be calculated using the "Dilution equation" (Equation 2.12 and 2.34).

More information is available in the "Basic theory", the "General description ofBOD-DO river models" and the "General description of the traditional oxygen sag curve".

$$\frac{dL}{dt} = -K_1L \qquad \text{--}2.30$$

$$L = L_0e^{-K_1t} \qquad \text{--}2.31$$

Where

L- BOD in the water (M, usually $g0_2 / m^3$)

L_o- initial BOD in the stream (below waste water discharge), see also Eq. 2.34 (M, usually g o_2/m^3)

K_1 - is the rate coefficient of biochemical decomposition of organic matter (T^{-1}usually day-$^{-1}$)

t- is the time, that is the time of travel in the river interpreted as t =x/v, where x is the distance downstream of the point of effluent discharge (T, usually days)

2.7.4 The dissolved oxygen model

The traditional dissolved oxygen model describes the fate, the "sag", of the dissolved oxygen in the river as influenced by the decay of biodegradable organic matter and the reaeration process (across the water surface). In Equation 2.4 the initial conditions, e.g $D=D_0$, $L=L_0$ at $x=0$ (at $t=t_0$) should be calculated using the "Dilution equation" (Equation 2.22, 2.34 and 2.35).

$$\frac{dD}{dt} = K_1L - K_2D \qquad \text{-----------------------}2.32$$

$$D = \frac{K_1L_0}{K_2-K_1}\left(e^{-K_1t} - e^{-K_2t}\right) + D_0e^{-K_2t} \qquad \text{------------}2.33$$

Where

D- is the oxygen deficit of water ($g0_2/m^3$), see also equations 2.36 and 2.37.

L- BOD in the water ($g0_2/m^3$)

D_0- is the initial oxygen deficit in the water (downstream of effluent outfall) ($g0_2/m^3$), see also equations 2.35 and 2.36

L_0- is the initial BOD concentration in the water ($g0_2/m^3$), (downstream of effluent discharge), see also Eq 2.34

K_1- is the rate coefficient of biochemical decomposition of organic matter (T^{-1}usually day^{--1})

K_2 - is the reaeration rate coefficient (T^{-1}, usually day^{-1})

t- is the time, that is the time of travel in the river interpreted as $t=x/v$, where x is the distance downstream of the point of effluent discharge; and v - is the mean flow velocity of the river reach in concern. (T)

LI

2.7.4.1 The "dilution equations" for BOD and DO

Considering a river and an effluent discharge of steady state conditions (with flows and concentrations not varying in time) and assuming instantaneous full cross-sectional mixing of the sewage water with the river water the initial concentration C, downstream of an effluent outfall can be calculated by the dilution equation (Eq. 2.22), which stems from the balance equation of in- and outflowing fluxes written for the section of the discharge point (e.g. back-ground river mass flux plus pollutant discharge mass flux equals the combined mass flow downstream of the point of discharge). This equation is used very frequently in simple analytical water quality models for calculating the initial concentration of pollutants. This two dilution equations compute the initial concentration of BOD and DO in the river downstream of a point source sewage discharge, with the assumption of instantaneous mixing.

$$L_0 = \frac{L_s q_s + L_b Q_b}{Q_b + q_s} \quad \text{------------------------------2.34}$$

$$DO_0 = \frac{DO_s q_s + DO_b Q_b}{Q_b + q_s}$$

-----------------------------2.35

Where

L_0 - is the initial concentration of BOD in the river, downstream of the effluent discharge point (ML^{-3}, e.g. mgO,/l);

L_b- is the background concentration of BOD in the river, (ML^{-3}, e.g. mgO_2/l);

L_S - is the BOD content of the waste water, (ML^{-3}, e.g. mgO_2/l);

DO_0 - is the initial concentration of dissolved oxygen in the river, downstream of the effluent discharge point (ML^{-3}, e.g. mgO_2/l);

DO_b - is the background concentration of dissolved oxygen in the river, (ML^{-3}, e.g. mgO_2/l);

DO_S - is the dissolved oxygen content of the waste water, (ML^{-3}, e.g. mgO_2/l);

Q_b - discharge (rate of flow) of the river upstream of the effluent outfall, ($L^3 T^{-1}$, e.g. m^3/s);
q_s - the effluent discharge, ($L^3 T^{-1}$, e.g. m^3/s);

2.7.4.2 The initial oxygen deficit equation

This set equations is used to calculate the initial oxygen deficit of the water downstream of a point source sewage discharge as compared to the saturation dissolved oxygen concentration, which latter is temperature dependent.

The initial oxygen deficit is calculated by subtracting the initial oxygen concentration D_0 from the saturation oxygen concentration DO_{sat} (Equation 2.36).

The saturation dissolved oxygen concentration of the water is temperature dependent, and the respective values can be obtained either from tables published in the relevant literature or from experimental expressions. In this teaching aid we will use the latter method in the form of Equation 2.37 (Wang et. al, ref. Gromiec, 1983).

$$D_0 = DO_{sat} - DO_0 \,, \quad [mgO_2/litre] \qquad \text{------------------------------2.36}$$

$$DO_{sat} = 14.61996 - 0.4042T + 0.00842T^2 - 0.00009T^3 \qquad \text{--------------------------------2.37}$$

Where

DO- is the initial concentration of dissolved oxygen deficit in the river, downstream of the effluent discharge point (ML^{-3}, e.g. mgO_2/l);

DO_0- is the initial concentration of dissolved oxygen in the river, downstream of the effluent discharge point (ML^{-3}, e.g. mgO_2/l); see also Eq. 2.35;

DO_{sat}- is the saturation oxygen concentration of water (ML^{-3}, e.g. mgO_2/l);

T- is the water temperature (^0C).

2.7.4.3 Oxygen sag curve and its critical values

The oxygen sag equation has a critical point where the dissolved oxygen content of water is the lowest that is when the oxygen deficit is the highest. The time of travel (or the corresponding downstream distance) can be expressed by finding the minimum of the sag curve. It is obtained in the form of Equation 2.38 for t_{crit}, Equation 2.39 for x_{Crit}, and Equation 2.40 for D_{crit}. Thus the critical dissolved oxygen concentration is obtained as the difference between saturation oxygen concentration and the critical oxygen deficit (Equation 2.41).

This set of four equations is used to computes the lowest dissolved oxygen concentration (highest oxygen deficit) in the river water downstream of a single source of sewage water along with the corresponding time of travel and downstream distance. For more details please see the "Basic theory", the "General description of BOD-DO river models" and the "General description of the traditional oxygen sag curve".

$$t_{crit} = \frac{1}{K_2 - K_1} \ln \frac{K_2}{K_1} \left(1 - \frac{D_0 (K_2 - K_1)}{L_0 \, K_1} \right) \qquad \text{------- } 2.38$$

$$x_{crit} = v \, t_{crit} \qquad \text{--}2.39$$

$$D_{crit} = \frac{K_1}{K_2} L_0 e^{-K_1 t_{crit}} \qquad \text{------------------------}2.40$$

$$DO_{crit} = DO_{sat} - D_{crit} \qquad \text{--------------------------------}2.41$$

Where

t_{crit}- the critical time of travel(time during which the water particle arrives to point of lowest DO concentration in the stream);

D_0 – is the initial concentration of dissolved oxygen deficit in the river, downstream of the effluent discharge point (ML^{-3});

L_0 - is the initial concentration of BOD in the river, downstream of the effluent discharge point (ML^{-3});

K_1- is the rate coefficient of biochemical decomposition of organic matter, the BOD decay rate, (T^{-1}, usually day^{-1})

K_2 - is the reaeration rate coefficient, the rate at which oxygen enters the water from the atmosphere, (T^{-1}, usually day^{-1})

X crit - the critical distance downstream of the point of effluent discharge (the point of lowest DO concentration) (L);

V- is the average flow velocity of the river reach in concern ($L\,T^{-1}$);

D_{crit} - is the critical (highest) oxygen deficit in the water, along the river, (ML^{-3}, e.g. $mgO_2/1$)

DO_{crit} - is the critical (lowest) dissolved oxygen concentration of the water (ML^{-3}, e.g. $mgO_2/1$)

DO_{sat} - is the saturation oxygen content of water, see also equation 2.37 (ML^{-3}, e.g. $mgO_2/1$);

K_2 estimation and its equation

This equation is used for the estimation of the value of the reaeration rate coefficient K_2 in function of the flow velocity and flow depth. Note that this equation have been "generated" for the purpose of this study and thus it differs from the many other formulas offered by the relevant literature.

The value of the reaeration coefficient K_2 depends, eventually, on the hydraulic parameters of the stream and a large number of experimental formulae have been presented in the literature along with reviews of these literature equations (Gromiec, 1983, Jolankai 1979, 1992). These expressions deviate from each other, sometimes substantially. Thus the obtained formula is Equation 2.42.

This equation is used for the estimation of the value of the reaeration rate coefficient K_2 in function of the flow velocity and flow depth. Note that this equation have been "generated" for the purpose of this study and thus it differs from the many other formulas offered by the relevant literature.

Note that mean velocity v and average flow depth H can not be arbitrarily selected (the more so since you have already specified river flow Q before). As Q =B*H*v, the river width B is defined by the other three variables. It calculates the width to depth ratio B/H and gives a warning signal when it indicates irregular channel dimensions. Moreover, velocity v and flow depth H are interrelated by hydraulic relationships, such as the Chezy or the Strickler-Manning formula. Selecting this latter the ratio $v/H^{2/3}$ remains constant for a given river reach (of constant channel slope and channel roughness). The user might wish to consider this, by manually checking that this ratio remains constant for the same example, when either v or H is altered. The authors have also incorporated a control which keeps this ratio within realistic literature ranges of the combination of channel slope and channel roughness, by releasing a warning signal when this range is exceeded.

$$K_2 = 2.148\ v^{0.878}\ H^{-1.48}$$

----------------------2.42

where

K_2- is the reaeration rate coefficient, the rate at which oxygen enters the water from the atmosphere, (day^{-1})

v - is the average flow velocity in the river reach, (m/s)

H - is the average depth of flow over the river reach, (m)

Temperature correction formula for K_1

Both the reaeration coefficient K_2 and especially the decomposition rate coefficient K_1 depend on the ambient (water) temperature. For this latter the most widely accepted formula is Equation 2.43.

This equation is used for the correction of the value of BOD decomposition rate coefficient K_1 in function of the water temperature. Note that this formula has been selected for this study from among the many others offered by the relevant literature. Also note that in algorithm that guides the selection of the value of K_1at 20^0C temperature in function of the type and size of the river and of the already calculated value of K_2. For more details please see the, the “General description of BOD-DO river models” and the “General description of the traditional oxygen sag curve”.

$$K_{1(T)} = K_{1(20^oC)} 1.047^{(T-20)}$$

------------------------------2.43

Where

$K_{1(T)}$ - is the value of rate coefficient K_1 at water temperature T ^0C

$K_{1(20\,C)}$ - is the value of rate coefficient K, at water temperature T=20 ^0C

T - water temperature (T^0C)

Ratio $f=K_2 / K_1$ in function of verbally described hydraulic condition of the stream

Description of water body	**Range of $f=K_2 / K_1$**
Large slow river	1.5-2.0
Large river of medium flow velocity	2.0-3.0

Table 2.1

2.7.5 Expanded, modified, BOD-DO river models

In addition to the decay of organic matter and the process of reaeration, discussed under the above headings, there are many other processes in a stream which affect the fate (the sag) of the dissolved oxygen content. These processes are, without claiming completeness, as follows:

Physical processes:

Effects of dispersion (mixing), spreading, mixing, diluting pollutants, thus reducing BOD (and increasing aeration, a process that is to be included in the reaeration rate coefficient K_2);

Settling of particulate organic matter, that reduces in-stream BOD values;

Chemical, biological and biochemical processes:

Effects of benthic deposits of organic matter (e.g the diffuse source of BOD represented by the decay of organic matter that had settled out earlier onto the channel bottom);

Sinks and sources of oxygen due to the respiration and photosynthesis of aquatic plants [macrophytes, phytoplankton (algae) and attached benthic algae];

oxygen consumption by oxidizing biochemical processes, such as nitrification.

Of the many modifications of the traditional oxygen sag curve we have selected two models for inclusion in this. The criteria of selection was that the model should take many or most of the above processes into consideration (for the first model) and it should also consider longitudinally varying flow and with this non-point source external loads (for the second model). It is to be noted that the models that deal with dispersion and mixing, since such models will be separately discussed later on.

2.7.5.1 The first expanded BOD-DO model

This expanded BOD-DO model is the modification of the traditional oxygen sag curve model .The first expanded BOD-DO model, was developed by Camp (1963) and it involves the following processes in addition to the decay of organic matter (BOD decay) and reaeration:

Sedimentation of biodegradable organic matter;

Benthic oxygen demand (e.g the diffuse source of BOD represented by the decay of organic matter that had settled out earlier onto the channel bottom);

Internal oxygen source represented by the photosynthetic activity of aquatic plants. (In this case one should note that the term accounting for this process in the model is rather the balance between oxygen input via photosynthesis and oxygen consumption via the respiration of aquatic plants, since respiration is not represented by a separate term in this model).

It is also worthwhile to mention that due to the diurnal variation of light the variation of the photosynthetic oxygen source can be best represented by a periodical function of the time, as it is done in some other, more complex, models (not discussed here). There are three new parameters in this model, the sedimentation rate constant K_3 the benthic BOD; B, and the photosynthetic input of DO; P. Estimation of these parameters is rather difficult in the absence of measurement data. (measurement is also rather complicated: the white-black bottle method is used for measuring the net input of oxygen by photosynthesis; a bell-shaped device set into the bottom sediment is used for measuring the benthic oxygen demand; and sedimentation of biodegradable organic matter is indicated by the change of the slope of a straight line (in logarithmic paper) showing the longitudinal variation of in-stream BOD measurement data;- the user is advised to consult the literature for more details of these techniques, when so required). Nevertheless for the purpose of this study it can be set pre-defined ranges of these model parameter values for the calculation example, and for that only. It will, however, indicate the way how such models are used in the practice, when no field measurement data on the parameter values are available;- e.g they are used for trying to explain unaccounted differences between measured and calculated in-stream data. That is when an observed BOD DO profile can not be simulated with reasonable parameter values of K_1 and K_2 then parameters B, P, and K_3 can be used to account for unknown internal sources or sinks in a trial-error manner.

The first expanded BOD model

These equations (the differential equation and its solution) describe the decomposition of organic matter (BOD decay), its sedimentation and the benthic source of it. See also: "General introduction of BOD-DO" models, the "General introduction of the traditional oxygen sag equation" and the "General introduction of the Expanded BOD-DO models".

$$\frac{dL}{dt} = -(K_1+K_3)L+B$$

------------------------------2.44

$$L = \left[L_0 - \frac{B}{(K_1+K_3)}\right]\exp[-(K_1+K_3)t]+\frac{B}{K_1+K_3}$$

--------------2.45

Where

L – is BOD in the water ($M\ L^{-3}$, g O_2/m^3)

L_0 - is the initial BOD in the stream (downstream of the waste water discharge), see also Eq. 2.34 ($M\ L^{-3}$, g O_2/m^3)

K_1- is the rate coefficient of biochemical decomposition of organic matter (T^{-1}, usually day^{-1}).

K_3 - is the rate constant for BOD removal by sedimentation (T^{-1}, usually day^{-1});

B - is the benthic oxygen demand, the rate of BOD addition to overlying water from the bottom sediment ($M\ T^{-1}\ L^{-3}$, usually $g0_2/m^3/day$)

t - is the time of travel (t= x/v) expressed in days

The 1st expanded oxygen model

The model describes the variation of the dissolved oxygen deficit of the water with the time of travel in function of the processes of reaeration, decomposition/decay of organic matter and oxygen production by photosynthesis. For details information please see also the following topics: "General introduction of BOD-DO" models, the "General introduction of the traditional oxygen sag equation" and the "General introduction of the Expanded BOD-DO models", as "General introduction of the Expanded BOD-DO models", as well as equations 2.44 and 2.45

$$\frac{dD}{dt} = -K_2D + K_1L - P \qquad \text{-------------------------2.46}$$

$$D = \frac{K_1}{K_2-(K_1+K_3)}\left[L_0-\frac{B}{(K_1+K_3)}\right]\left\{\exp\left[-(K_1+K_3)t\right]-\exp(-K_2t)\right\}+$$

$$+\frac{K_1}{K_2}\left[\frac{B}{(K_1+K_3)}-\frac{P}{K_1}\right]\left[1-\exp(-K_2t)\right]+D_0\exp(-K_2t)$$

---------------------------2.47

Where

D- is the oxygen deficit of water (ML^{-3} ,e.g. $g0_2/m^3$), see also equations 2.36 and 2.37

D_o- is the initial oxygen deficit of water (ML^{-3} ,e.g. gO_2/m^3), downstream of the effluent discharge, see also equations 2.35 and 2.36

L – is the BOD in the water (ML^{-3} ,e.g. g O_2/m^3)
L_0 - is the initial BOD in the stream (downstream of the waste water discharge) (ML^{-3} ,e.g. g O_2/m^3), see also Eq. 2.34

K_1 - is the rate coefficient of biochemical decomposition of organic matter (T^{-1}, usually Day^{-1})

K_2 - is the reaeration rate coefficient (T^{-1}, usually day^{-1})

K_3 - is the rate constant for BOD removal by sedimentation (T^{-1}, usually day^{-1});

B - is the benthic oxygen demand, the rate of BOD addition to overlying water from the bottom sediment ($M\ T^{-1}\ L^{-3}$, usually g O_2/m^3/day)

P - is the rate of oxygen addition to water by the photosynthetic activity of aquatic plants ($M\,T^{-1}\,L^{-3}$, usually g O_2/m^3/day);

t- is the time of travel ($t = x/v$) expressed in days

2.7.5.2 The second expanded BOD-DO model

Although in this model the basic modelling concept has also been changed slightly the reader/user is kindly requested to consult also the following topics: "General introduction of BOD-DO models", the "General introduction of the traditional oxygen sag equation", the "General introduction of the Expanded BOD-DO models" and the "General introduction of the first expanded BOD-DO models".

The main differences of this modelling concept are as follows:

1. Longitudinal variation of the mass flux (the product of flow and concentration) is expressed, instead of expressing the variation of concentration with the time (of travel), thus allowing for the consideration of longitudinally varying river flow.

2. The DO equation is written for the dissolved oxygen (termed here C_{ox}) instead of the oxygen deficit D.

3. Non-point source input loads are also considered in terms of concentrations of BOD and DO in the lateral inflow (here the term lateral inflow q ($L^2\,T^{-1}$) refers to the increment of river flow Q ($L^3\,T^{-1}$) over a unit downstream distance (L) of the river, assuming uniform q values over a given river reach.

4. Photosynthesis and respiration of aquatic plants are considered separately, that is rather with their difference, e.g. (P-R).

The 2nd expanded BOD-DO model, selected for this study has essentially the same parameters as the first expanded BOD-DO model and the same limitations refer to the possibilities of parameter estimation. The differences are:

a, the parameter q, the lateral inflow, that can be relatively easily obtained from the hydrological (longitudinal) profile. For a given river reach of length x it obtained as the flow increment over the reach divided by the length of the reach.

b, Concentrations of the constituents in the lateral inflow should be either estimated by another submodel or a literature estimate of runoff concentration of the respective substance must be used.

C, Parameters of BOD decay rate K_1 and reaeration rate K_2 have slightly different values from those of the previous models (due to the difference in modelling concept and thus in the exponents of the model equations). A correction algorithm is built in the programme to facilitate conversion (not shown in the written material), so as to allow the use of the respective parameter estimation formulae and tables.

In the calculation example of this study it can be used pre-defined ranges of parameter values, within which the user may select one, so as to see their effect on the final outcome of the model simulation.

The 2nd expanded BOD model

The model equations describe the longitudinal variation (profile) of BOD in function of the decomposition process of organic matter, non-point source inputs represented by lateral inflow and a benthic source of BOD. For more details please see also the following topics: “General introduction of BOD-DO models”, the “General introduction of the traditional oxygen sag equation", the "General introduction of the Expanded BOD-DO models", and the "General description of the 2^{nd} expanded BOD-DO Model".

$$\frac{d}{dx}[(Q+qx)L] = qL_d - K_1AL + B$$

------------------------------2.48

$$L(x) = L_0F^{\beta_1} + \frac{L_d + B\phi(x)}{\beta_1}\left[1 - F^{\beta_1}\right]$$

<u>*where*</u>

$$F = \frac{Q_0}{Q_0 + qx} \quad ; \quad \beta_1 = 1 + K_1\phi(x) \quad ; \quad \phi(x) = \frac{\frac{Q_0}{q} + x}{v}$$

---------------------------------2.49

L – is the BOD in the water (ML^{-3} ,e.g. g O_2/m^3)

L_0 - is the initial BOD in the stream (downstream of the waste water discharge) (ML^{-3} ,e.g. g O_2/m^3), see also Eq. 2.34

K_1 - is the rate coefficient of biochemical decomposition of organic matter (T^{-1}, usually Day^{-1})

L_d - is the concentration of BOD in the lateral inflow to the stream, (the diffuse load components, ($M\ L^{-3}$, gO_2/m^3)

Q - is the rate of flow in the river ($L^3\ T^{-1}$, usually m^3/s)

Q_0 -is the rate of flow at the beginning of the river reach, just upstream of the waste water discharge (L^3, T^{-1}, usually m^3/s)

q - is the lateral, specific, inflow rate to the river ($L^2 T^{-1}$, usually m^2/s)

A - is the wetted cross-section area of the stream (L^2), defined as the rate of flow Q divided by the cross-sectional mean flow velocity v. (usually m^2)

B - is the benthic oxygen demand ($M T^{-1} L^{-1}$, here $gO_2/m/day$),

x - is the distance downstream along the river (L, usually in meters).

v - is the mean flow velocity along the river reach in concern ($L T^{-1}$, m/s),

The 2nd expanded dissolved oxygen model

The model describes the longitudinal variation of the dissolved oxygen content of the river as affected by point and non-point sources of biodegradable organic matter and dissolved oxygen (BOD, DO), the decomposition process of organic matter, the reaeration process and by the photosynthesis and respiration of aquatic plants.

For more details see also the following topics: "General introduction of BOD-DO models", the "General introduction of the traditional oxygen sag equation", the "General introduction of the Expanded BOD-DO models", the "General introduction of the first expanded BOD-DO models" and the "General description of the 2nd expanded BOD-DO Model".

$$\frac{d}{dx}\left[(Q+qx)C_{ox}\right] = qC_{ox_L} - K_1AL + K_2A\left(C_{ox_{sat}} - C_{ox}\right) + A(P-R)$$

--------------------------------------**2.50**

$$C_{ox}(x) = \left[\frac{L_d+B\phi(x)}{\beta_1}-L_0\right]\frac{K_1}{K_2-K_1}\left(F^{\beta_1}-F^{\beta_2}\right)+$$

$$+\left[C_{ox_{sat}}(\beta_2-1)-\frac{L_d+B\phi(x)}{\beta_1}(\beta_1-1)+\phi(x)(P-R)+C_{ox_d}\right]\frac{1-F^{\beta_2}}{\beta_2}+C_{ox_0}F^{\beta_2}$$

where .

$$F=\frac{Q_0}{Q_0+qx}\ ;\ \beta_1=1+K_1\phi(x)\ ;\ \phi(x)=\frac{\frac{Q_0}{q}+x}{v}\ ;\ \beta_2=1+K_2\phi(x)$$

-------------------------------------**2.51**

Where

C_{OX} - the dissolved oxygen concentration of water, referred to as DO in the former equations (M L^{-3}, gO_2/m^3),

$C_{ox,0}$ - is the initial dissolved oxygen concentration downstream of the waste water discharge (M L^{-3}, gO_2/m^3), see also Eq. 2.35

$C_{OX,,d}$ - is the concentration of DO in the lateral inflow to the stream, the diffuse load component, (M L^{-3}, gO_2/m^3)

$C_{OX,,sat}$ - is the saturated dissolved oxygen concentration of water termed before also as DO_{Sat} (M L^{-3}, gO_2/m^3),see also Eq. 2.37,

L_0 - is the initial BOD in the stream (downstream of the waste water discharge), see also Eq. 2.5 (M L^{-3}, gO_2/m^3)

L_d - is the concentration of BOD in the lateral inflow to the stream, (the diffuse load components, (M L^{-3}, gO_2/m^3)

K_1 - the rate coefficient of biochemical decomposition of organic matter (T^{-1}, usually day^{-1})

K_2 - is the reaeration rate coefficient (T^{-1}, usually day^{-1})

Q - the rate of flow in the river ($L^3 T^{-1}$, usually m^3/s)

Q_0 - is the rate of flow at the beginning of the river reach, just upstream of the waste water discharge ($L^3 T^{-1}$, usually m^3/s)

q - is the lateral, specific, inflow rate to the river ($L^2 T^{-1}$, usually m^2/s)

A - is the wetted cross-section area of the stream (L^2), defined as the rate of flow Q divided by the cross-sectional mean flow velocity v. (usually m^2)

B - is the benthic oxygen demand ($M T^{-1} L^{-1}$, here g O_2/m/day),

P-R - is the net difference between oxygen production by the photosynthesis and oxygen consumption by the respiration of aquatic plants ($M T^{-1} L^{-3}$, gO_2/m^3/day).

x - is the distance downstream along the river (L, usually in meters).

v - is the mean stream flow velocity in the river section investigated ($L T^{-1}$, m/s)

2.8 ANALYSIS OF WATER QUALITY BY USING STATISTICAL METHODS

The use of statistical techniques in aquatic environmental monitoring is a relatively new subject (Hamed & Rao, 1998). Statistical tests are currently considered invaluable in the design and analysis of the results of monitoring programmes (Antonopoulos *et al*, 2001; McBride & Loftis, 1994). Routine water quality monitoring programmes often serve many purposes, but are primarily aimed at assessing the environmental state of water and detecting trends (EEA, 1996). Often, monitoring programmes produce a wealth of long term time series data. Statistical tests are the principal means used in the analysis of these time series data to "...arrive at a deeper understanding of the causal mechanisms that generated it..." (Kendall *et al,* 1983). A time series generally has four components: a trend or long term movement; oscillations about the trend, of greater or less regularity; a seasonal effect and an irregular or random component (Kendall *et al*, 1983). Quantitative statements about time series data are obtained from the series by decomposing it into its individual components (Mattikalli, 1996). Of key interest in this study is the trend component.

2.8.1 TREND TESTING IN WATER QUALITY DATA

Detecting temporal trend in water quality data is a subject that is currently very topical in the water research fraternity (Hess *et al*, 2001; Hamed & Rao, 1998; Helsel & Hirsch, 1991; Hirsch *et al,* 1991).

A trend in water quality data is defined as a monotonic change in a particular constituent with time (Helsel & Hirsch, 1991), the causes of which may or may not be known. The purpose of trend testing is to investigate whether the measured values of a water quality constituent are increasing or decreasing over time in statistically significant terms (Onoz & Bayazit, 2003; Helsel & Hirsch, 1991). When checking for an increase or decrease in variable values, statistically it means determining whether the probability distribution from which they arise has changed with time (Antonopoulos *et al*, 2001; Helsel & Hirsch, 1991).

Trend detection involves statistical testing of the null hypothesis that there is no trend with time, the alternative hypothesis being that there is an overall increase or decrease in a constituent with time at some specified significance level (Onoz & Bayazit, 2003; Zipper *et al*, 1998). The significance level is the probability of incorrectly rejecting the null hypothesis when it is in fact true (McBride & Loftis, 1994; Helsel & Hirsch, 1991). The traditional default significance value usually used in statistics is 5% (Griffith *et al*, 2001; Helsel & Hirsch, 1991). A statistically significant trend is obtained when a smaller p value than the significance level (0.05) is obtained (Griffith *et al*, 2001).

When selecting the trend detection method to follow, several considerations must be made. These considerations include, taking into account the characteristics of the data itself; the type of trend expected (monotonic or step); the general category of statistical methods to employ (parametric or non parametric); adherence of data to assumptions of normality, independence, linearity, as required by various analytical procedures; as well as the objectives of the study (Darken *et al*, 2002; Qian *et al*, 2000; Hirsch *et al*, 1991).

2.8.2 Characteristics of water quality data

Water quality data exhibit many characteristics that complicate trend analysis and influence the choice of analysis procedures (Darken, 1999; Helsel & Hirsch, 1991). A mere application of analysis procedures that assume that data possess characteristics that they may in fact not, often results in the incorrect interpretation of the actual situation and therefore, add no value to the analysis at hand (Helsel & Hirsch, 1991). Helsel & Hirsch (1991), state that all statistical methods, from simple summarization methods to the more complex procedures should recognize the characteristics of most water resources data outlined below;

- No negative values are possible, the lower boundary is zero.
- Outliers, that is, those values that are considerably higher or lower than most of the data occur regularly.
- Due to the two characteristics above, positive skewness occurs. Skewness is an indication of the departure of the distribution from the normal symmetrical curve. The data are positively skewed if the curve has a longer right tail and are negatively skewed if it has a longer left tail. In many cases water quality data are positively skewed, indicating the presence of many relatively low values and a few high values (Grabow *et al*, 1999).

The measure of skewness is known as the coefficient of skewness. In perfectly normally distributed data, the coefficient of skewness is zero. Values that are greater than one, point to a degree of skewness that ought to be addressed (Grabow *et al*, 1999).

- Owing to the three characteristics above, non-normal distribution of data often occurs. Most inferential methods, or parametric methods including linear regression, are founded on the assumption of a normal (Gaussian) distribution and yet water quality data and their regression residuals are often 'distinctly non-normally distributed' (Vant & Smith, 2004; Lietz, 2000; Darken, 1999; Zipper *et al,* 1992). Tests for normality include the Shapiro-Wilk test (test statistic, W) and the Kolmogorov-Smirnov (test statistic D) among a few other tests. The Shapiro- Wilk test statistic (W) which is frequently used, ranges from zero to one, with low p-values (smaller than 0.05) resulting in the rejection of the hypothesis of normality (Grabow *et al*, 1999). Values that are greater than 0.05 suggest that there is insufficient evidence to reject the null hypothesis, and lead to the assumption that the data are indeed from a normal distribution.
- Censored data. This refers to observations that fall below or above some threshold, often occurring because of imprecision in laboratory analytic techniques when measuring minute or very large quantities (Darken, 1999). The magnitudes of water quality observations that occur beyond technically possible detection limits are often unknown. In linear regression, this amounts to incomplete information that cannot be used effectively (Darken, 1999; Zipper *et al*, 1992). Whilst the IDT (1989) have suggested that censored values can be eliminated by assigning a value of half the detection limit to each of the affected records, Cox *et al* (2005) argue that in their experience this method results in low outliers. Low outliers often result in skewness. Therefore, they propose that the detection limit itself be used instead (Cox *et al*, 2005).
- Seasonal patterns are common in water resources data, as most concentrations in surface water respond to seasonal changes in stream flow and catchment inputs (Helsel & Hirsch, 1991).
- Serial correlation also known as autocorrelation is another common feature of long term water quality data whose extent is a critical consideration. Serial correlation is a situation where water samples carry over information from one sample to the next, such that data have redundant information and less information for the same number of independent observations (Grabow *et al*, 1999). This is usually a result of monitoring the same station

at very close time intervals (McBride & Loftis, 1994). For example, data that is collected more frequently than monthly tends to lose its independence (Cox *et al*, 2005, McBride & Loftis, 1994). The power of statistical tests is a function of independent sample size. Therefore, if data lose their independence, the tests become prone to falsely rejecting the null hypothesis of no trend (Type I error), leading to a wrong conclusion of the presence of a trend (McBride & Loftis, 1994). According to Grabow *et al* (1999) the easiest way to eliminate autocorrelation, is to aggregate or average the data into longer time steps (data reduction). Alternatively, one could employ the Seasonal Kendall Test with the covariance sum method to allow for serial dependence. However, when serial correlation is negligible, tests such as the Seasonal Kendall Test may be applied in their original form (Darken *et al*, 2002).

- Covariance. This is a situation where the behaviour of the variables under study is influenced by other and uncontrolled variables. For example, the concentration of some water quality variables, such as conductivity is influenced by stream flow. Therefore, two analyses must be carried out on the data; one on the non-flow adjusted data and another on flow adjusted data where flow data are available, to determine whether a trend detected is genuine or a result of flow. Analysis of trend on flow adjusted variables eliminates the influence of flow, thereby, giving a more reliable trend (Vant & Smith, 2004; Griffith *et al*, 2001; Zipper *et al*, 1998).
- Another feature of time series data that complicates analysis is unequal sampling, which is often viewed as equal sampling intervals with missing data (Darken 1999).

In multiple station studies, multiple starting and ending dates complicate trend analysis. Therefore, in order to correctly interpret data, the water quality records must be concurrent and yet long enough to discern true trends. Hirsch *et al* (1991) state that when one is doing a study of data, for instance, from 1970 to 1985, a record that starts in 1972 or ends in 1983 need not be excluded from the study, similarly a record with a two year gap need not be disqualified from the study.

The water quality characteristics discussed above often present problems in trend analysis particularly when conventional parametric techniques such as linear regression are used (Lietz, 2000; Darken, 1999; Zipper *et al*, 1992). However, Zipper *et al*, (1992) state that in spite of the limitations identified thus far, linear regression has been observed to be capable of detecting water quality trends, particularly where those trends were well defined.

2.8.3 Types of trend

Two types of trend are considered for hypothesis testing and trend estimation in literature; these are the monotonic and step trends (Qian *et al,* 2000; Darken, 1999; Hirsch *et al*, 1991).

2.8.3.1 *Monotonic trend*

Monotonic trend occurs when a response variable changes with a concomitant change in the explanatory variable. The change is unidirectional over time and the hypothesis in this case does not specify whether this shift is continuous, linear, in one or more discrete steps, or in any particular pattern (Zipper *et al*, 1998; Hirsch *et al*, 1991). If no prior hypothesis of a time change is known, or if records from multiple stations are being analysed in a single study, then monotonic trend procedures are appropriate (Hirsch *et al* 1991).

For monotonic trend, the hypothesis that the data shifts monotonically is often tested with regression analysis of the water quality variable as a function of time (Hirsch *et al*, 1991). The alternative test is the nonparametric Mann-Kendall test for trend (Hirsch *et al*, 1991). A number of variations of the Mann Kendall test are currently in use, these include the Seasonal Kendall Test with its variations, the covariance sum test, or the covariance eigenvalue test (Darken, 1999).

2.8.3.2 *Step trend*

The step trend occurs when a variable changes from one constant level to another constant level (Darken, 1999; Hirsch *et al*, 1991). The assumption made is that the locations (means or median values) of data collected before and after a specific time are distinctly different. The trend hypothesis tested here is more specific than with the monotonic trend in that it requires that certain facts be known before examining the data (Hirsch *et al*, 1991). Step trend procedures are particularly useful in two instances; firstly, when there is a relatively large gap in time separating data into two distinct groups (Darken, 1999; Hirsch *et al*, 1991). While there is no hard and fast rule as to the length of the gap, a time gap of over a third of the study period warrants the use of the step trend procedures (Hirsch *et al*, 1991). Secondly, if an event that is likely to have resulted in change in water quality is known, then two time periods are considered, that is the before and after the incident. Examples of such events include the introduction of point or non point sources of pollution, or a ban of certain polluting activities.

Step trend techniques include the two-sample t-test with the associated estimates of change in magnitude based on the difference in sample means. The alternative is the Mann-Whitney-Wilcoxon Rank sum test and the Hodges-Lehmann estimator of trend magnitude (Darken, 1999; Hirsch *et al*, 1991).

2.8. 4 METHODS FOR DETECTING TREND

Whilst many techniques are available for trend detection and estimation in environmental data (Griffith *et al*, 2001, Hess *et al*, 2001) none are published as standard tests for assessing water quality data (Griffith *et al*, 2001). However, many have been proposed (Qian *et al*, 2000) and established through common practice (Griffith *et al*, 2001). The various trend detection tests fall under two broad categories, the parametric and the nonparametric methods (Hess *et al*, 2001; Hamed & Rao 1998; Hirsch *et al*, 1991).

2.8.4.1 Parametric methods

Parametric statistical tests are those tests that are based on estimates of statistical parameters such as mean and standard deviation. Parametric statistical methods are based on linear regression and therefore check only for linear or monotonic trend. They assume that the random variable (water quality constituent) is independent and follows the normal distribution 'or very nearly so' (Onoz & Bayazit, 2003; Grabow *et al*, 1999; Helsel & Hirsch, 1991). A normal distribution is bell shaped, but as has been asserted in the discussion above, water quality data are often skewed and therefore seldom normally distributed.

However, when the data do follow a normal distribution and are independent, parametric methods are perceived to be the most powerful tests, far more powerful than their nonparametric counterparts at any significance level (Darken, 1999; Hamed & Rao, 1998; Zipper *et al*, 1992; Hirsch *et al*, 1991). However, if the data do not follow normal distribution, the parametric tests tend to lose their power to detect trends (Hirsch *et al*, 1991). Examples of parametric tests include the ordinary least squares also known as linear regression, the t-test and the rank sum test.

Transformations have been carried out on data to make them more linear, symmetric and normally distributed (see for example Verhallen *et al*, undated; Cox *et al*, 2005) and, thus more suited to parametric methods. However, Hirsch *et al* (1991) argue that transformations are not always recommended, particularly with multiple record data.

They state that data transformations are subjective; because data can easily be manipulated until a desired or preconceived result is achieved (Hirsch *et al*, 1991).

2.8.4.2 *Linear Regression*

Simple linear regression is an important and commonly used parametric method for identifying monotonic trend in a time series (Zipper *et al*, 1992; Helsel & Hirsch, 1991). It is used to describe the relationship between one variable with another or other variables of interest. It is often performed to obtain the slope coefficient of a water quality variable on time. The slope coefficient is tested under the null hypothesis that it is equal to zero. Regression has the advantage that it provides a measure of significance based on the hypothesis test on the slope or correlation coefficient; and also gives the magnitude of the rate of change (Hirsch *et al*, 1991).

2.8.4.3 *T-Test for difference between means*

The t-test is a simple trend detection method that is often used by researchers with very basic training in statistics (Hess *et al*, 2001). It is based on whether a statistically significant difference exists between the means first half of data and last half (Hess *et al*, 2001). The mean of second half of data is simply subtracted from the mean of the first half of data. The statistical significance at 95% confidence interval is then calculated. However, Hess *et al* (2001) state that the t-test is only valid when observations are random and stress that t-tests ought to be used with extreme caution in serially correlated data.

Onoz & Bayazit (2003) used Monte Carlo simulation to investigate the relative power of the parametric t-test, in comparison with the non parametric Mann-Kendall test. They found the t-test to be slightly more powerful under normal probability distribution, decreasing in power ratio with an increase in the coefficient of skewness. However, for moderately skewed distributions, they observed the t-test to be nearly as powerful as the Mann-Kendall test. They then used the two tests on water quality data of 25 to 65 years at 107 sites in various river basins in Turkey. They found that both tests detected trend in 29 series; at two sites trends were detected by the t-test only; at two other sites by the Mann Kendall test only; and at four sites the two tests gave different results. Onoz & Bayazit (2003) then concluded that the two tests could be used interchangeably in practical applications with identical results.

2.8.4.2 Non-parametric methods

These are hypothesis tests that are distribution free (Zipper *et al* 1992), not requiring the assumption that data follow any particular distribution and not restricted to linear trend (Onoz & Bayazit, 2003). Non parametric tests demand independence of data and use rank order statistics only. They extract trend information based on the relative magnitudes of the data, but do not quantify the size of change (Hamed & Rao 1998; Helsel & Hirsch 1991). Tests including the Seasonal Kendall Slope Estimator and the Sen's slope estimator are available to quantify the trends (Libiseller, 2002; Zipper *et al*, 1992).

Non-parametric tests are thought to be more powerful than their parametric counterparts for non-normally distributed data, and are therefore useful in water resources data (Yue *et al*, 2002; Lietz, 2000; Helsel & Hirsch 1991). When dealing with normally distributed data, these tests are said to be as powerful as their parametric counterparts (Lietz, 2000; Hirsch *et al*, 1991). There are two basic non-parametric tests for trend analysis, the Mann Kendall and the Spearman's rho test (Onoz & Bayazit, 2003). These tests are discussed in subsequent sections.

The decision to use either the parametric or non-parametric tests is based on the considerations of power and efficiency (Hirsch *et al*, 1991). Power is defined as the probability to reject the null hypothesis (no trend) when it is truly false (McBride & Loftis, 1994). Therefore in order to be useful, the tests chosen should have good power to detect environmentally important trends (Griffith *et al*, 2001).

Monte Carlo simulation analyses used by Hirsch *et al* (1991) to compare the performance of both the parametric and non-parametric methods under various probability distribution conditions revealed that the non-parametric methods suffer minute disadvantages in terms of power and efficiency when data are normally distributed. However, when data depart slightly or largely from normality then modest and significant advantages are observed respectively.