DYNAMIC MODELLING OF PROCESSES IN RIVERS AFFECTED BY PRECIPITATION RUNOFF

Judith L. Jacobsen

Lyngby 1997
IMM–PHD–1997–30
ATV Erhvervsforskerprojekt EF 520

IMM

Preface

This thesis have been prepared as a joint venture between the Department of Mathematical Modelling (IMM), Technical University of Denmark, and the engineering consultancy firm, PH-Consult, Ordruphøjvej 4, DK–2920 Charlottenlund, as part of the requirements for the degree of Ph.D. in engineering.

This is an industrial research project funded equally by PH-Consult and the Academy of Technical Sciences (ATV).

The thesis encompasses an inter-disciplinary study between time series analysis and stochastic modelling, on one side, and appraisal of water quality and urban impact from precipitation runoff, on the other. The work is focused on the incorporation of apriori knowledge about the physical system in all the phases involved with the identification of models.

The main contribution to this field is aimed towards the development of stochastic models for water quality. A minor application of statistical methodology to ease the calibration of deterministic models have also been carried out. The applications of the methods used and the industrial relevance is presented along with the results. A more thorough report on the business aspects regarding this project is given in the Technical Report *Technological management of the transfer of knowledge within stochastic modelling* (Tech-report IMM-24).

Acknowledgements

I wish to express my gratitude to all who have contributed with support in different ways towards this research.

First of all, I wish to thank my supervisors, Assoc. Prof., Ph.D. Henrik Madsen, IMM, DTU, Prof. Poul Harremoës, Department of Environmental Science (IMT), DTU and PH-Consult and Assoc. Prof., Jens J. Linde-Jensen, Department of Applied Civil and Environmental Engineering, DTU (formerly DIA-B). They deserve much gratitude for their help and guidance, as well as their enthusiastic and inspiring encouragement throughout my work on this dissertation.

I also wish to thank my colleagues at both PH-Consult and IMM, especially past and present members of the time series group and the environmental statistics group. Also thanks to my ATV group, who along with the two formerly mentioned groups, have helped keeping my spirits high. I am grateful for the network I have acquired through the industrial Ph.D. program and for the many good discussions regarding our specific status in transferring knowledge between university and industry. Special thanks are due to ATV for helping provide this network through their new course in technology management, as well as their overall support.

The following deserves special acknowledgement for providing data and advice:

- Solar radiation for the cases of Harrestrup Å and Usserød Å: Department of Agricultural Sciences, Section of Soil and Water and Plant Nutrition, The Royal Veterinary and Agricultural University, Copenhagen, Denmark.
- Data for Harrestrup Å:

The Danish Meteorological Institute, DMI, and Ballerup Municipality.

- Data and advice for Gryde Å:
  Assoc. Prof. M. Erlandsen of Institute for Biostatistics, Århus University, Denmark and Biologist T.M. Iversen, director of the National Environmental Research Institute (NERI), Silkeborg, Denmark.
- Personal contributions:
  Prof. K. Sand of University of Copenhagen, Lab. for Fresh Water Biology, Hillerød, Denmark; for sharing his vast knowledge of the biological system in small rivers.
  Professor P.C. Young, director for Center for Research on Environmental Systems and Statistics (CRESS); for his interest and for sharing his vast knowledge of systems analysis, specifically with regards to dissolved oxygen in rivers.

Last, but not least, I am grateful to my new husband Håkon and our son Derek for their patience and their tolerance of the hardships they have had to endure because of my enthusiasm with this project.

Symbols

This section provides a list of the most often used symbols, as well as a list over the most common abbreviations.

Notation

In this list of symbols and the related variables, t denotes time, while x denotes space and T denotes temperature.

For the units the following symbolism is used: L: Length; M: Mass; T: Time; E: Energy. [–] denotes that a unit cannot be supplied, or that the symbol has no unit.
Notice that the unit of time is denoted T, as is the variable: temperature. Also organic matter, expressed by biological oxygen demand is denoted by L, as is the unit of length. It should be very clear to discern which one is used at the time. With these exceptions, it is attempted that no symbol carries more than one meaning, although this is often the case in literature.

In order to provide an easy overview, only frequently used symbols are included here. Throughout the report, it is aimed that a symbol is introduced as necessary when used.

Roman letters

Symbol	Description	Unit
$A(x,t)$	Cross sectional area	$[L^2]$
$C(f)$	Cumulative periodogram	$[T^{-1}]$
$C(x,t)$	Dissolved oxygen concentration	$[ML^{-3}]$
$C_m(x,t,T)$	Saturated oxygen concentration	$[ML^{-3}]$
e_t	Measurement error	$[-]$
f	Frequency	$[T^{-1}]$
g	Gravitational constant	$[LT^{-2}]$
$h(x,t)$	Water level	$[L]$
$I(f)$	Periodogram	$[T^{-1}]$
$I(t)$	Solar radiation	$[EL^{-3}T^{-1}]$
I_o	Slope (in $^0/_{00}$)	$[-]$
$L(x,t)$	Organic matter	$[ML^{-3}]$
K_{1c}	Oxygen consumption,	$[MM^{-1}L^{-1}]$
K_{1l}	Degradation of organic matter	$[MM^{-1}L^{-1}]$
$K_2(h,T)$	Reaeration constant	$[T^{-1}]$
K_3	Change in organic matter, dep. of DO	$[MM^{-1}]$
K_b	Oxygen consumption from the riverbed	$[T^{-1}]$
K_l	Degradation constant	$[T^{-1}]$
$K_L(h,T)$	Adsorption coefficient	$[LT^{-1}]$
$P(I,t,T)$	Photosynthesis	$[ML^{-2}T^{-1}]$
P_m	Max. photo synthesis	$[ML^{-3}T^{-1}]$
$P_r(t)$	Precipitation (rain)	$[L]$
$R(T)$	Respiration	$[ML^{-3}T^{-1}]$
R_w	Hydraulic radius	$[L]$
$Q(x,t)$	Water flow	$[L^3T^{-1}]$
$v(x,t)$	Velocity	$[LT^{-1}]$
V_F	Sheer stress velocity	$[LT^{-1}]$
$V(x,t)$	Volume	$[L^3]$
w_t	Wiener process for model noise	$[-]$

Bold $\mathbf{I}$ is the unit matrix, while capital (non-bold) I denotes solar radiation. Langley, Ly, is an old unit for solar radiation used in Paper **C**. $Ly = cal/cm^2 = 41.868kJ/m^2$.

Greek letters

Symbol	Description	Unit
α	Coefficient in photosynthetic function	$[-]$
β	Linear coefficient in photosynthetic function	$[ML^{-3}E^{-1}]$
η	Coefficient in photosynthetic function	$[L^{-1}]$
θ	Temperature correction factor	$[-]$
$\boldsymbol{\theta}$	Parameter vector	$[-]$
κ	Coefficient for precipitation impact	$[ML^{-3}]$
τ	Time constant for low-pass filter	$[T]$
ω	frequency	$[T^{-1}]$

Abbreviations

Abbreviations are also explained where introduced.

Abbrev.	Description
ATV	Academy of Technical Sciences
BOD	Biological oxygen demand
CSO	Combined Sewer Overflows
CTLSM*	Continuous Time Linear Stochastic Modelling
DO	Dissolved oxygen
NER	Nearly Equal Ripple (filter)
WPA+	Water Pollution Act+
WTP	Waste water treatment plant

* Notice that this program has later been adapted to handle non-linear models as well. + In Danish: Vandmiljøplanen.

Summary

In this thesis, models for the dynamics of oxygen and organic matter in receiving waters (such as rivers and creeks), which are affected by rain, are developed. A time series analysis framework is used, but presented with special emphasis on continuous time state space models. Also, the concept of model identifiability is attended. For estimation of the parameters in the models the maximum likelihood method is used and the Kalman filter employed to evaluate the likelihood function. In the case of non-linear models, the extended Kalman filter is used. To evaluate the models, various residual analysis methods and model validation tools are employed.

To develop the water quality model, including hydraulic relations and the states of oxygen and organic matter, the qualitative concepts of the physical, biological and chemical models are introduced. The model types used in this thesis are stochastic models. Most of the models are based on measurements from one measuring station, though models based on the linear reservoir description applied to measurements from two measuring stations are also considered. Both time varying and time invariant models are employed. In the models, the oxygen dynamic complex includes reaeration, photosynthesis, respiration and degradation of organic matter.

The effect of pre-filtering data is investigated, as is various functions for photosynthesis as a function of solar radiation. For the degradation of organic matter delayed reactions have been studied. In most models, precipitation in the form of rain have been included to study the impact from this.

Finally, the future and industrial perspectives are presented, along with a list of suggestions for future research related to the subjects considered in this thesis.

Resumé (in Danish)

I denne afhandling er der udviklet modeller for dynamikken af ilt og organisk stof i recipienter (specielt mindre vandløb), som er påvirket af regn. Rammerne for de tidsrække analytiske aspekter er presenteret med opmærksomheden særligt henledt på kontinuert tids tilstandsmodeller. Identificerbarhed af modellerne er også kort beskrevet. Til estimation af parametrene i modellerne bruges maximum likelihood metoden, hvor Kalman filteret bruges til evaluering af likelihood funktionen. I de ikke-lineære tilfælde bruges det udvidede Kalman filter. For at evaluere modellerne, bruges forskellige residual analyser samt nogle statistiske model validerings værktøjer.

Med henblik på at udvikle vandkvalitetsmodellen, inklusive de hydrauliske relationer, samt tilstandene for ilt og organisk stof, introduceres de kvalitative koncepter for de fysiske, biologiske og kemiske modeller. Modellerne i afhandlingen er stokastiske modeller. De fleste er baseret på målinger fra een målestation, men også modeller baseret på en lineær reservoir model med målinger fra to målestationer er foreslået. Både tidsvarierende og tidsinvariante modeller er brugt. I modellerne for iltindholdets dynamik er der inkluderet geniltning, fotosyntese, respiration og nedbrydning af organisk stof.

Effekten af at pre-filtrere data er undersøgt, samt forskellige funktioner for fotosyntese som en funktion af solindstråling. Nedbrydningen af organisk stof er blevet undersøgt med henblik på det forsinkede iltforbrug. I de fleste modeller er regn inkluderet for at undersøge dens påvirkning på vandløbet.

Til sidst er fremtidige og industrielle perspektiver opridset, sammen med en liste af forslag til videre forskning indenfor de emner, der er undersøgt i denne afhandling.

Contents

B Modelling the Transient Impact of Rain Events on the Oxygen Content of a Small Creek.

C Modelling the Dynamical Effect of Solar Radiation on the Oxygen Content of a Small Creek.

D Estimation of a Low Pass Filter for Solar Radiation Data.

In press: Proceedings of the 9.th Conference of the European Consortium for Mathematics in Industry, ECMI (1997). **141**

E A Method for Automatic Calibration of Parameters for an Integrated Model.

Submitted to: Water Resources Research (1997). **153**

F A Stochastic Model for Two-station Hydraulics Exibiting Transient Impact.

Submitted to: The Watermatex conference (1997). **167**

Part I

Theory and background

Chapter 1

Introduction

Water is the basis of all life. Water to drink and for transportation purposes. Not only transporting people and goods, but also as a transport medium for unwanted wastes. Looking at historical human settlements, one will always find them located in close proximity of water. If the water source dried up or could not be used, perhaps because of pollution, the settlements moved away.

In Denmark, where drinking water (mostly groundwater) can be transported from wells outside the cities, clean rivers and streams are instead considered a quality recreational advantage, especially in urban areas. This may seem a luxury compared to parts of the world where clean drinking water (and enough of it) is the foremost concern. However, there is no doubt that all over the world there is a global recognition of water quality being one of the 21. century's biggest environmental concerns (Shanahan and Somlyódy (1995)) and already Young and Beck (1974) recognized that solutions must be found by technological innovation. However, the work in this thesis is directed towards Danish rivers as the literature mainly concerns American, English and German rivers.

Oxygen, measured as dissolved oxygen, is one of the most universal environmental factors to evaluate water quality in rivers and streams. The level of oxygen determines the kind of aquatic fauna that may exist, and thereby the trophic level of the ecosystem. The higher the trophic level, the higher order of animal life and the larger the bio-diversity.

In this chapter, the background and purpose for the present dissertation is given, along with a short historical overview of the major contributions to the investigation of the oxygen dynamics of rivers. The historical overview does not pretend to be complete in any way. It is only intended to outline the historical developments. Finally, an outline of how to read this thesis is given.

1.1 Background

In Denmark, the Water Pollution Act (WPA) enacted in 1987 has set standards for stationary waste water discharge that have resulted in investments of more than Dkr. 15 billion. A combined sewage system has built-in storage basins to take some of the load when the sewage system cannot handle the extra strain during heavy rainfall. Rainfall is a powerful transient burden of the entire system, that may cause discharge of oxygen consuming components resulting in lowered oxygen levels and possibly dead fish. Thus, the WPA-standards eventually may result in even further investments in renewal and restoration of the sewage system, as well as in the monitoring of the receiving waters.

The essence of the problem involves three different components: **1**) The drainage system, **2**) the sewage treatment plants and **3**) the receiving waters. Each of these components are equally important and holistic treatments of the integrated system are surfacing. While the first two components have been the subject of much research, the receiving waters have not yet been submitted to such thorough investigation.

Within recent years, markedly improved measuring methods and equipment have resulted in still larger amounts of data. Thus, a pronounced need for statistically correct treatment of data have surfaced. At the same time the development in data treatment systems (e.g. computers) have supplied the possibility for much improved treatments. Presently, very large amounts of data are being collected at enormous expenses, while very small amounts have been spent on the data treatments that could lead to new knowledge. As yet, few statistically sound and consistent methods to treat these data have been developed. This, despite the fact that these data are very important to evaluate the WPA, as well as future initiatives for a better management and protection of the environment. Also, the quality in

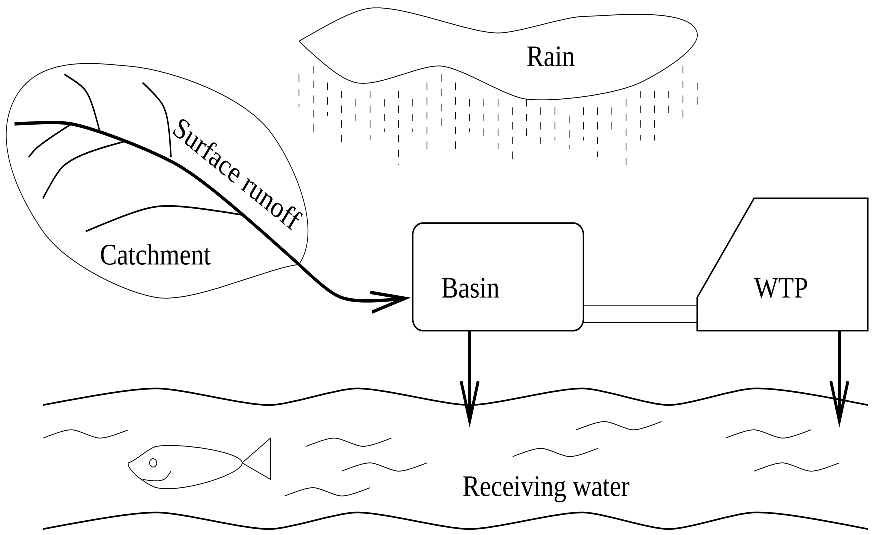


Figure 1.1: Integrated urban runoff system with a catchment, from where surface runoff is directed to a storage basin and then on to the waste water treatment plant (WTP) or directly into the receiving water.

monitoring programs is inconsistent at best and methods to assure a high quality in the painstakingly collected data needs to be developed as well.

1.2 Purpose and goal

The purpose of initiating this research project was a desire to develop stochastic models for the processes in rivers affected by the impact from rain events. To reach that goal, more experience would be gained with data treatments based on statistically sound principles. The host academic environment had the specialised expertise from extensive research with stochastic modelling of a diverse collection of physically based systems. Statistical methods in general was also to be applied where it could be advantageous to the other areas of business in the host company, though this was not the main goal. For example to ease the calibration procedure of the deterministic models.

It has been recognized for some while that stochastic models have something new to offer as a supplement to the traditional deterministic models the company has built their business upon. The purpose of the study was to investigate the advantages of stochastic models, while concentrating on models for the oxygen dynamics in streams.

This is where a jointly funded Ph.D. project works as a link to speed up the technology transfer between university research and practical use of this research. In return, the university partner obtains insight into the implementation of their theories.

1.3 Historical overview

Streeter and Phelps (1925) studied the pollution and natural purification of the Ohio river and presented the first basic theory of the oxygen dynamics in polluted rivers. The model was based on a mass balance, but only included oxygen consumption from degradation of organic matter and oxygen replenishment from reaeration. The river was assumed to be homogeneous, totally mixed, without dispersion and with a stationary supply of organic matter. Photosynthesis and respiration was not accounted for.

The next major contribution came from Odum (1956) who added photosynthesis and respiration to the oxygen mass balance. Respiration was considered constant, though at the time this was argued to possibly be dependent on the oxygen tension in the polluted waters, or repressed during photosynthetic action. It was also proposed that the respiratory metabolism of a polluted stream might be proportional to the content of organic matter. The double curve analysis was presented as a tool to calculate the gross primary production between two measuring stations. Owens and Marris (1964) and McDonnell and Weeter (1972) conducted laboratory research into which factors influenced the respiration, including temperature, plant types and the oxygen concentration.

Dobbins (1964) fractionized the terms for dissolved oxygen and organic matter, represented by biological oxygen demand (BOD) into sedimentation of organic matter and diffuse organic loads. It was proposed that the degradation constant might be increasing when the degradation process was adapting to a new load of BOD. Both Dobbins (1964) and Thackston and Krenkel (1969) investigated the reaeration coefficient as a function of hydraulic characteristica and its dependence on temperature, while

Hornberger and Kelly (1975) determined atmospheric reaeration based on Odums productivity analysis. Camp (1965) concentrated on measuring BOD and the resuspension of sedimented organic matter. Furthermore, he presented the oxygen sag curve as the curve where the degradation of organic matter had reduced the oxygen concentration to its minimum, somewhere downstream from the source. He stated that the assumption of steady-state conditions would introduce greater errors than would the inaccuracies in the determination of rate constants.

1.4 Outline of the thesis

In this section the organization of the thesis is outlined with a reading suggestion.

The thesis is based on a collection of papers, which are included at the end. To tie the work in the various papers together, the first part of the thesis is a summary of the theory behind the applications, as well as an overview of the methods used. As this work is inter-disciplinary, the theory given strive to be thorough without pretending to be a tutoring book, yet easily understandable for various groups of professionals.

In Chapter 2, the time series analysis concepts, relating to the work is given as a reference frame for the estimating procedures used throughout this thesis. Also, the Grey box modelling concept is introduced and the advantages of this concept is explained. Finally in this chapter, an application is presented of a simple statistical method to automatically calibrate the parameters of a deterministic model. Parts of this chapter may seem very simple for the experienced time series analysts, but is included to give an adequate background for the work reported in the included papers for readers who are less familiar with the time series concepts.

In Chapter 3, the qualitative concepts behind the hydraulics and oxygen dynamics are introduced. The reader who was very familiar with the previous chapter may have a greater need to go into detail with this one. On the other hand, the reader with a hydrological background may want to skip some parts of this chapter.

Chapter 4 presents the different cases that have been worked with and summarizes the results gained from models for the three different rivers.

The rivers are very different: They range from a rural, natural river, to a concrete lined urban stream of mostly surplus water during rain.

In Chapter 5 a discussion is given of the overall results from the work with this thesis and a number of suggestions for further work in this area. Also the industrial perspective and the applications for the company are discussed here.

Finally, the conclusion is presented in Chapter 6, after which the included papers are placed.

Chapter 2

Time series analysis concepts

In this chapter some relevant general concepts from time series analysis is presented as an introduction for the reader who is primarily well versed in the water quality issues of this thesis. However, it also serves as a description in more detail, than is generally possible in a paper or article of often pre-specified length.

First, it is described how data are sampled prior to the modelling effort and whether or not a filter has been applied. Then the state space model is introduced for linear time invariant, linear time variant and non-linear systems. Before the description on how to estimate the parameters in the models, some considerations of the model at hand is explained, including whether or not it is identifiable. Finally, residual analysis and model validation is described. Terminology and notation is naturally introduced where they occur.

2.1 Sampling and filtering of data

Before a set of parameters can be estimated for a given data set, the choice of sampling time should be considered. If the sampling interval is too long, some interesting dynamics may be lost, while too much attention might be drawn to otherwise insignificant dynamics if the sampling instants are too close together. Thus, the purpose of considering the sampling time and to concentrate on the horizon of main interest, is to avoid loss of information from the data due to sampling.

In a continuous time model, the parameters will generally depend on the input to the system, the sampling instant and the pre-sampling filter (if one is used). Problems arise if aliasing occur (see e.g. Bloomfield (1976)). Aliasing is when high frequencies in the unsampled data are seen as low frequency variation in the sampled signal because the spectral density of the continuous time signal is folded into the Nyquist interval. In order to avoid or reduce the aliasing problem a pre-filter has to be used before sampling.

2.1.1 The Nyquist interval

Consider two signals; one with a frequency higher than the other; say if one with the frequency $\omega_1 = \pi/T$ and the other with $\omega_2 = 3\pi/T$. ($\omega = 2\pi f$, where f is the frequency). These two signals could be sampled with the same sampling time, T ($f = 1/T$), which then makes the two signals indistinguishable, and hence aliases of each other. The concept is illustrated in Figure 2.1 in the time domain, and shows two sinusoidal signals with a sampling that makes it impossible to conclude which signal the sampling is from.

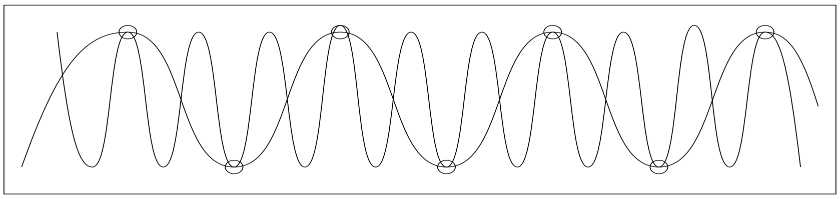


Figure 2.1: An example of the aliasing problem in the time domain. The lower frequency signal is generated using $\sin(\omega_1 t)$, while the other (higher frequency) signal is generated by $\sin(\omega_2 t)$. The small circles denotes the sampling points.

If a continuous time signal is zero outside the so-called Nyquist interval $[-\pi/T, \pi/T]$, then there is a unique relation between continuous time and discreet time signals. Mathematically, the Nyquist interval and the aliasing problem is best explained in the frequency domain. Consider a sampled signal $x_s(t) = x(t)j(t)$, where $j(t)$ is a series of indicator functions, indicating the sampled values, and let $x(t)$ be given in the frequency domain by $X(\omega)$. By folding we obtain:

$$\begin{aligned} X_s(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} J(\nu)X(\omega-\nu)d\nu \\ &= \frac{1}{T}\int_{-\infty}^{\infty} \delta(\nu - n\omega_0)X(\omega-\nu)d\nu \end{aligned} \tag{2.1}$$

where $J(\omega) = \delta(\omega - n\omega_0)$ $(\omega_0 = 2\pi/T)$ is the Fourier transformation of $j(t) = \delta(t - nT)$ (see Madsen (1989)). Thus, the sampled signal is in the frequency domain given by:

$$X_s(\omega) = \frac{1}{T}\sum_{n=-\infty}^{\infty} X(\omega - n\omega_0) \tag{2.2}$$

where T is the sampling time and ω_0 is the sampling frequency.

Hence, the sampled signal in the frequency domain is represented by a periodic function with the period ω_0. If $X(\omega)$ contains values outside the interval $[-\omega_0/2, \omega_0/2] = [-\pi/T, \pi/T]$, these will not be distinguishable in $X_s(\omega)$ from values inside this interval. The frequency $\omega = \pi/T$ is called the Nyquist frequency and may be utilized to decide the sampling time T, as the sampling frequency should be at least twice as large as the bandwidth of the signal (Shannon's theorem).

2.1.2 Filtering

Having decided to limit a set of data to frequencies within an interval of relevance to the modelling that is to be performed, the next decision is which kind of filter to use. The filter could either be applied to the data before the estimation procedure, e.g. a pre-filtering technique, or be inherent in the estimation procedure, e.g. estimated along with the parameters. To model the oxygen dynamics, the diurnal variation is of major interest and there is not expected to be any time constants of interest less than one hour. Hence, we may wish to filter away the frequencies higher than half an hour. A part of the high frequency variation might be a result of stochastic fluctuations, not necessarily representative of the dynamics in the system.

In Paper **C** it was investigated which effect the application of a nearly ideal low pass filter would have on the accuracy on the parameter estimates of a model describing a very simple model for the oxygen dynamics. A Nearly Equal Ripple (NER) filter (Kaiser and Reed (1977)) was used on the solar radiation data. These were measured every five minutes and showed very high frequency variations, along with diurnal variation. Eight different descriptions of photosynthesis as a function of solar radiation were tested to find the best description. The application of the filter was successful in the sense that it leads to more precise parameter estimates.

In Paper **D** it was decided to include the estimation of a parameter in a filter in the overall estimation procedure. This would enable the system to be used on-line. The results in terms of more accurate estimates, were not as impressive as when pre-filtering with a nearly ideal low pass filter. However, the possibility for on-line monitoring is a great advantage at a time where industry has to monitor their waste water and its effect on the receiving waters. Enormous investments are placed in monitoring programs with more and more sophisticated measuring equipment. Therefore, it would be desirable to use estimation methods, which can be used to extract as much information as possible from the data.

2.1.3 Summary of sampling and filtering

It is imperative that the system to be modelled is well known before rejecting high frequencies as noise. Some of these variations **could** be an unknown dynamic that might be of interest. In the model used in Paper **C** it is argued that since the solar radiation is measured as a point measurements (obtained using one single solarimeter) and the photosynthesis only were expected to react to the low frequency variation, it might lead to a more precise estimation. This was in fact shown in that paper.

Sub-sampling without the use of a filter (picking out values at equidistant intervals) works well if data are free of high frequencies, i.e. all the frequencies, which cannot be distinguished from their aliases. This was the case in Papers **A**, **B** were the solar radiation data had already been sampled by an averaging filter before we obtained the data. Since it is important to filter all data by the same filter, the other dataseries used to estimate parameters in those models were also filtered by averaging data the same way.

2.2 State space models

A state space model is an internal parametric representation for the relation between the input and the output signals of a system. The continuous time state space model allows for direct physical evaluation of the parameters because it allows for the use of physical information to obtain a reasonable model structure.

The variables of a system, such as the oxygen and organic matter provides the state vector for a given time $t > t_0$. Together with future input and noise values, the correct state vector describes the future evolution of the system. In a deterministic system the noise term is zero and the future values of the state vector can be predicted exactly from the number of initial values and future input. A model formulation based on deterministic descriptions and physical relationships of a given system will typically be a set of coupled differential equations. To these are added stochastic terms to account for the uncertainties and discrepancies between the model and the true system, as well as measuring errors.

In stochastic systems, at a given time t all information from the past is contained in the actual value of the state. Hence, the state vector describe a Markov process.

Stochastic differential equations are governed by an irregular stochastic process, such as Gaussian white noise. The equations are written symbolically as nth order stochastic differentials. which describes an Itô process (Melgaard (1994), Øksendal (1995)):

$$d\mathbf{x}_t = \mathbf{f}(t, \mathbf{x}_t, \mathbf{u}_t)\, dt + \mathbf{G}(t, \mathbf{x}_t, \mathbf{u}_t) d\mathbf{w}_t \qquad t \geq 0 \tag{2.3}$$

where $\mathbf{f}$ is a deterministic function of time t and state $\mathbf{x}$ (called a drift term), $\mathbf{u}$ is the input vector and $\mathbf{G}(t, \mathbf{x}_t, \mathbf{u}_t)$ is the diffusion coefficient. $\mathbf{w}_t$ is an n-dimensional standard Wiener process. Eq. (2.3) is a general state space description.

2.2.1 The linear time invariant model

The simplest model is linear, time invariant and described by the system equation:

$$d\mathbf{x}_t = \mathbf{A}(\boldsymbol{\theta})\, \mathbf{x}_t\, dt + \mathbf{B}(\boldsymbol{\theta})\, \mathbf{u}_t\, dt + \mathbf{G}(\boldsymbol{\theta})\, d\mathbf{w}_t \tag{2.4}$$

where $\mathbf{x}$ is an n-dimensional state vector, $\mathbf{A}$ is an $n \times n$ matrix, which characterizes the dynamical behaviour of the system and $\boldsymbol{\theta}$ is the parameter vector. $\boldsymbol{B}$ is an $n \times m$ matrix, specifying how the m-dimensional input signals $\boldsymbol{U}$ enter the system. Notice that linearity in the parameters is not required. The observation equation, taken at discrete time instants t_k, is:

$$\mathbf{Y}_k = \mathbf{C}(\mathbf{u}_k, \boldsymbol{\theta})\,\mathbf{x}_k + \mathbf{D}(\mathbf{u}_k, \boldsymbol{\theta})\,\mathbf{u}_k + \mathbf{e}_k \tag{2.5}$$

where $\boldsymbol{D}$ is a matrix, which accounts for input variables that directly affects the output. None of the models in this thesis include such phenomena, hence this matrix is zero and thus excluded in the following.

This model was used in Paper **F** and later extended to include time varying terms, which is reported in Paper **G**. The model investigated in Paper **F** is a simple hydraulic model, modelling the water level by a lumped parameter linear reservoir model.

For linear and time invariant models, the state variables may be eliminated. Hence, the transfer function, or external form, is obtained (see e.g. Box and Jenkins (1976), Madsen and Holst (1996)).

2.2.2 The linear time varying model

Consider the linear time varying stochastic differential equation

$$d\mathbf{x}_t \quad = \quad \mathbf{A}(\mathbf{u}_t, \boldsymbol{\theta}, t)\,\mathbf{x}_t\,dt + \mathbf{B}(\mathbf{u}_t, \boldsymbol{\theta}, t)\,\mathbf{u}_t\,dt + \mathbf{G}(\boldsymbol{\theta}, t)\,d\mathbf{w}_t \tag{2.6}$$

The model may typically be formulated as a linear model, but with time varying coefficients varying according to an input signal. The observation equation is:

$$\mathbf{Y}_k \quad = \quad \mathbf{C}(\mathbf{u}_k, \boldsymbol{\theta}, t_k)\,\mathbf{x}_k + \mathbf{e}_k \tag{2.7}$$

Another approach leading to this type of model, is a linearization of the general differential equation (2.3) around some reference point $\mathbf{x}^*$. In this case the matrices are calculated by:

$$\mathbf{A}(\mathbf{u}_t, \boldsymbol{\theta}, t) = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\Big|_{x=x^*} \qquad \mathbf{B}(\mathbf{u}_t, \boldsymbol{\theta}, t) = \frac{\partial \mathbf{f}}{\partial \mathbf{u}}\Big|_{x=x^*} \qquad \text{etc.} \tag{2.8}$$

Linear time varying models are presented in Papers **C**, **D** when investigating the effects of pre-filtering and the estimation of a filter. Paper **G** also presents a more complete water quality model with both oxygen and organic matter, with the photosynthesis as a function of the solar radiation.

2.2.3 Nonlinear models

Nonlinear models may be dependent on the input as well as the states. Using Eq.s (2.7)–(2.8), but extending the matrices in (2.8) to be dependent on the state vector, we obtain:

$$d\mathbf{x}_t = \mathbf{A}(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\theta}, t)\, \mathbf{x}_t\, dt + \mathbf{B}(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\theta}, t)\, \mathbf{u}_t\, dt + \mathbf{G}(\boldsymbol{\theta}, t)\, d\mathbf{w}_t \qquad (2.9)$$

were it should be noticed that $\mathbf{G}(\boldsymbol{\theta}, t)$ is not a function of $\mathbf{x}$.

In Papers **A, B** the delayed degradation of sedimented organic matter was investigated by the theory for oxygen diffusion in the sediment. This is generally acknowledged to be a half order model, meaning that a square root of the oxygen state was included. Hence, this is a non-linear model.

2.3 Systems considerations

Before the actual modelling is started many considerations should be done. Some of these concern whether the system is

- Persistently excited?
- Observable?
- Structurally identifiable?

If a non-identifiable model is specified, the methods used for estimation will not converge. The first question is whether the inputs are able to persistently excite the model, so that the applied estimation procedure is able to distinguish between different parameter values. That a system is *observable*, means that the states involved must either be directly observable or reconstructable by the other observed states. If two structures have the same joint density function, they are said to be observationally equivalent (Harvey (1989)). If there exist a unique model structure, the model is *structurally identifiable* (Walter and Pronzato (1990)). An autonomous sub-structure in the model complex that cannot be identified from the information, in the main model is not structurally identifiable. Thus, different models will yield equally good fits, i.e. there is no uniquely best set of parameters (Beck et al. (1993)). Åstrøm and Bohlin (1965) defines a process as completely identifiable if the maximum likelihood estimates of the parameters are consistent. It is only possible to observe a dynamic system if it is *persistently* exited, since the response is obscured by noise. I.e., the question to ask is whether it is possible to distinguish between different models for a given estimation method? Thus, the experimental conditions must assure that data contains enough information. These last two concepts could also be denoted explicit and implicit observability, respectively.

These concepts can be exemplified by the complex dynamics regarding the oxygen dynamics in receiving waters. The state for organic matter is *observable* even though it is not measured, because there is a linear relationship to oxygen relating to the degradation of organic matter (see Papers **A, B, G**). In the photosynthetic process, there are several complicated processes regarding the physiology of the plants, besides the dependence of the solar radiation. If these processes do not depend on the solar radiation, the temperature or the other variables in the systems complex (e.g. if they depend on plant size, form or structure), this sub-system is not *structurally identifiable*. This is illustrated in Figure 2.2, where one system contains a sub-system, which is not identifiable.

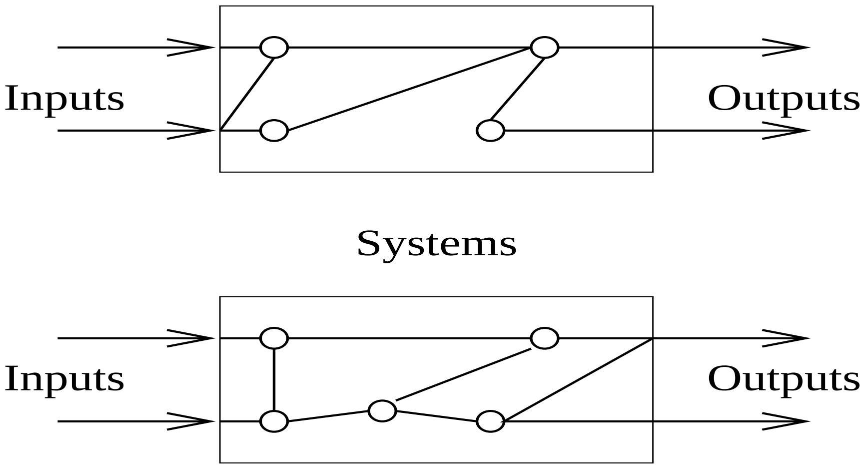


Figure 2.2: The concept of model structure identification. (Adapted from Jakeman et. al 1993). The figures illustrate that there may be a sub-system, which is not identifiable.

The storm sewer overflows with its large concentration of organic matter, most probably keeps the systems *persistently exited*, e.g. these inputs are expected to facilitate variations with respect to the interaction between oxygen and organic matter that are observable to a degree that cannot be explained by measuring errors or other chance disruptions.

2.4 Structural identifiability of models

The problem of structural identifiability arises from the fact that for a given transfer function model, there exists an infinite number of possible state space model representations. Hence, it is in general necessary to impose some constraints on the state space model in order to provide a unique relation between the parameters in the transfer function and the state space model. These constraints also concern the incremental covariance of the innovations and the continuous time Kalman gain (see Melgaard and Madsen (1993b) for more detail). Here, for simplicity only the theory for the deterministic part of the stochastic differential equation will be presented. With this method, questions regarding parameterization and structure may be answered before data collection is started.

In a stochastic system, such as the linear state space model described in Section 2.2.1, Eq. (2.4), the question of structural identifiability concerns the parameters of the $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ matrices. If they are uniquely determined, the system is structurally (or globally) identifiable.

One way to check this is to consider how the linear stochastic differential equation should be transferred into a general form where $\boldsymbol{Y}(t)$ from the measurement equation is uniquely determined. This may be done by using the Laplace transformation. For linear and time invariant systems, the Laplace transformed output, is determined by the Laplace transformed impulse response function $\mathbf{G}(s)$ and the input vector $\mathbf{U}(s)$:

$$\mathbf{Y}(s) = \mathbf{G}(s)\,\mathbf{U}(s) \tag{2.10}$$

The transfer function $\mathbf{G}(s)$ is obtained by Laplace transforming the matrix equation of the linear state space model:

$$\frac{d\boldsymbol{X}}{dt} = \boldsymbol{A}\,\boldsymbol{X} + \boldsymbol{B}\,\boldsymbol{U} \tag{2.11}$$

$$\boldsymbol{Y} = \boldsymbol{C}\,\boldsymbol{X} \tag{2.12}$$

By Laplace transforming (2.11) we obtain:

$$(s\boldsymbol{I} - \boldsymbol{A})\boldsymbol{X} = \boldsymbol{B}\boldsymbol{U} \tag{2.13}$$

Then by combining (2.12) and (2.13) we can eliminate $\boldsymbol{X}$:

$$\boldsymbol{Y}(s) = \boldsymbol{G}(s)\boldsymbol{U}(s) \tag{2.14}$$

where

$$\boldsymbol{G}(s) = \boldsymbol{C}(s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{B} \tag{2.15}$$

In order to identify the parameters in the linear state space model, it is sufficient to identify the parameters for the transfer function $\boldsymbol{G}(s)$ (Söderström and Stoica (1989)). Eq. (2.15) is the estimateable transfer function and the model parameters in Eq.s (2.11) – (2.12) should be determined uniquely as functions of the parameters of the transfer function for the model to be identifiable.

2.4.1 An example of identifiability

The concept of structural identifiability is illustrated in the following example, which is an extension of the stochastic, hydraulic two-station model suggested in Paper **F** and carried out in Paper **G**. The system is a third order linear stochastic differential equation for a linear reservoir model with two reservoirs.

The state space model written in matrix notation is:

$$\begin{bmatrix} dh_{out} \\ dh_2 \\ dh_1 \end{bmatrix} = \overbrace{\begin{bmatrix} -a & a & 0 \\ 0 & -a & a \\ 0 & 0 & -a \end{bmatrix}}^{\boldsymbol{A}} \overbrace{\begin{bmatrix} h_{out} \\ h_2 \\ h_1 \end{bmatrix}}^{\boldsymbol{X}} dt + \overbrace{\begin{bmatrix} 0 & 0 & c \\ 0 & b & c \\ a & 0 & c \end{bmatrix}}^{\boldsymbol{B}} \overbrace{\begin{bmatrix} h_{in} \\ \frac{\Phi P_r}{A_{surf}} \\ \frac{1}{A_{surf}} \end{bmatrix}}^{\boldsymbol{U}} dt \tag{2.16}$$

where, for ease of calculations (and because the meaning of the parameters is not the issue here), we have introduced:

$$a = \frac{1}{R_h}; \qquad b = A_c; \qquad c = Q_c \tag{2.17}$$

With the observation equation (2.12): $\boldsymbol{Y} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \boldsymbol{X}$ Eq. (2.15) become:

$$\boldsymbol{G} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} s+a & -a & 0 \\ 0 & s+a & -a \\ 0 & 0 & s+a \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 & c \\ 0 & b & c \\ a & 0 & c \end{bmatrix} \tag{2.18}$$

Now, if $(s\boldsymbol{I} - \boldsymbol{A}) = \tilde{\boldsymbol{A}}$, then $\tilde{\boldsymbol{A}}^{-1}$ can be found by:

$$\tilde{\boldsymbol{A}}^{-1} = \frac{Adj(\boldsymbol{A})}{\det \boldsymbol{A}} \tag{2.19}$$

Then, after some tedious calculations, we obtain:

$$\boldsymbol{G} = \frac{\begin{bmatrix} a^3 & +ab(s+a) & c(s+a)^2 + ac(s+a) + a^2c \end{bmatrix}}{(s+a)^3} \tag{2.20}$$

which may be written more conveniently as:

$$\boldsymbol{G} = \frac{\begin{bmatrix} \tilde{\alpha}_{11} & \tilde{\alpha}_{12} & \tilde{\alpha}_{13} \end{bmatrix}}{\tilde{\beta}_{11}} \tag{2.21}$$

Eq. (2.20) should be compared to what it is actually possible to observe in Eq. (2.21). Thus we have:

$$\begin{aligned}
\tilde{\alpha}_{11} &= \overbrace{a^3}^{p_3} \\
\tilde{\alpha}_{12} &= \overbrace{ab}^{p_2} s + \overbrace{a^2b}^{p_3} \\
\tilde{\alpha}_{13} &= \overbrace{c}^{p_1} s^2 + 3\overbrace{ac}^{p_2} s + 2\overbrace{a^2c}^{p_3} \\
\tilde{\beta}_{11} &= \overbrace{1}^{q_0} s^3 + \overbrace{3a}^{q_1} s^2 + \overbrace{3a^2}^{q_2} s + \overbrace{a^3}^{q_3}
\end{aligned}$$

All the terms with s^3, s^2, etc. are compared for the numerator and the denominator, respectively. I.e.:

$$\overbrace{\begin{aligned} 0 &= p_0 \\ c &= p_1 \\ ab + ac &= p_2 \\ a^3 + a^2b + 2a^2c &= p_3 \end{aligned}}^{\text{numerator}} \qquad \overbrace{\begin{aligned} 1 &= q_0 \\ 3a &= q_1 \\ 3a^2 &= q_2 \\ a^3 &= q_3 \end{aligned}}^{\text{denominator}} \tag{2.22}$$

I.e.:

$$\begin{aligned} c &= p_1 && (2.23) \\ ab+3ac &= p_2 \rightarrow b = \frac{p_2-3ac}{a} && (2.24) \\ 3a &= q_1 \rightarrow a = \frac{q_1}{3} && (2.25) \end{aligned}$$

Based on the above equations, it can be seen that it is possible to uniquely identify a, b, and c. Hence, the linear stochastic state space model with the $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ matrices from Eq. (2.16) is identifiable.

The concept of identifiability is more complicated for non-linear models. An obvious approach is to linearize the system around some equilibrium point and use the methods available for linear systems. However, identifiability of a linearized system, does not automatically ensure identifiability of the complete non-linear system (see e.g. Ljung and Glad (1994)).

2.5 Maximum likelihood estimates

In this section it is described how the parameters in a linear stochastic differential equation can be estimated, using discrete time measurements and the maximum likelihood method. In order to obtain the likelihood function the matrix $\boldsymbol{Y}^*(t)$ is introduced:

$$\boldsymbol{Y}^*(t) = [\boldsymbol{Y}(t), \boldsymbol{Y}(t-1), \ldots, \boldsymbol{Y}(1), \boldsymbol{Y}(0)]' \qquad (2.26)$$

where $\boldsymbol{Y}^*(t)$ is a matrix containing all the observations up to and including time t. $\boldsymbol{\theta}$ denotes a vector of all the unknown parameters - including the unknown variance and covariance parameters.

The likelihood function is the joint probability density of all the observations assuming that the parameters are known, i.e.:

$$\begin{aligned} L'(\boldsymbol{\theta}; \boldsymbol{Y}^*(N)) &= p(\boldsymbol{Y}^*(N)|\boldsymbol{\theta}) \\ &= p(\boldsymbol{Y}(N)|\boldsymbol{Y}^*(N-1), \boldsymbol{\theta}) p(\boldsymbol{Y}^*(N-1)|\boldsymbol{\theta}) \\ &= \left(\prod_{t=1}^{N} p(\boldsymbol{Y}(t)|\boldsymbol{Y}^*(t-1), \boldsymbol{\theta})\right) p(\boldsymbol{Y}(0)|\boldsymbol{\theta}) \qquad (2.27) \end{aligned}$$

where successive applications of the rule $P(A \cap B) = P(A|B)P(B)$ is used to express the likelihood function as a product of conditional densities.

The maximum likelihood method is used for parameter estimation for the models in Papers **A**, **B**, **C**, **D**, **F**, **G**. All equations are given in Papers **A**, **D**, **F**, though only the main equations in Paper **A**, as this paper also included the equations for going from continuous to discrete time (see Section 2.5.1) and the equations for the Kalman filter (see Section 2.6).

The maximum-likelihood estimate (ML-estimate) is the set of estimated parameters $\hat{\boldsymbol{\theta}}$, which maximizes the likelihood function. Since it has not been possible to optimize the likelihood function analytically, a numerical method, such as e.g. the quasi-Newton method has to be used. The numerical details of these algorithms can be found in Madsen and Melgaard (1991).

An estimate of the uncertainty of the parameters, $\boldsymbol{D}$ is obtained by the fact that the ML-estimator is asymptotically normally distributed with mean $\boldsymbol{\theta}$. The variance matrix is unknown, but by using Cramer Rao's inequality a lower bound is obtained, i.e.:

$$\boldsymbol{D} = \boldsymbol{H}^{-1} \tag{2.28}$$

where the matrix $\boldsymbol{H}$ is given by

$$\left\{\hat{h}_{lk}\right\} = -E\left[\frac{\partial^2}{\partial\theta_l\partial\theta_k}\log L(\boldsymbol{\theta}; \boldsymbol{Y}^*(N))\right] \tag{2.29}$$

An estimate of $\boldsymbol{D}$ is obtained by equating the observed value with its expectation and applying

$$\{h_{lk}\} \approx -\left(\frac{\partial^2}{\partial\theta_l\partial\theta_k}\log L(\boldsymbol{\theta}; \boldsymbol{Y}^*(N))\right)_{|\theta=\hat{\theta}} \tag{2.30}$$

i.e. equating the parameter vector $\boldsymbol{\theta}$ by it's estimate $\hat{\boldsymbol{\theta}}$. Hence, the uncertainties of the parameter estimates are found using the observed curvature of the log-likelihood function, evaluated at the final estimates.

Since the parameter estimates are asymptotically normal distributed, a t-test of the parameter significance can be performed.

2.5.1 From continuous to discrete time

The discrete time model corresponding to the continuous time model given in Section 2.2.1, Eq. (2.4) is obtained by integrating the differential equations through the sample interval $[t, t+\tau[$.

When the measurements used for the estimation have been filtered, the filter dynamics should be included in the model description to assure a correct system specification. In cases where the measurements have been low pass filtered, e.g. by simple averaging, it is advantageous to calculate the input as linear through the sample interval. This was the case in Papers **A**, **B**, **G**. In natural systems it is not reasonable to belive that input is constant through the sample interval (it is in fact only normal in man-made I–O inputs). Therefore, the sample interval was linearized in all estimations.

The equations for transfering the continuous time descriptions to discrete are given in Paper **A**, with the appropriate extensions for the linear (non-constant) sampling time.

2.6 The Kalman filter

In order to calculate the one-step prediction (to evaluate the likelihood function) for the state of the system along with formulas for updating (or reconstructing) this estimate, the Kalman filter was used in all the estimations. The evaluation of the maximum likelihood estimate by the Kalman filter via the prediction error decomposition was first introduced in 1965 (Harvey (1989)), but not used much until later when more computer power became available. Ljung and Caines (1979) showed the asymptotic normality of prediction error estimators, such as the Kalman filter.

2.6.1 Linear time invariant models

If the matrices **A**, **B** and **G** in Eq. (2.4), Section 2.2.1 (if **G** is applicable) are time invariant, it is possible to find an explicit solution for the prediction of mean and covariance of the state vector and observations, by integrating the equations over the time interval $[t_k, t_{k+1}[$ and assuming that $u_t = u_k$ in this interval. This implementation of the Kalman filter thus involves the calculation of the exponential of a matrix, which may be done once and for all for a given set of parameters in a time invariant model (see Madsen and Melgaard (1991) for details).

2.6.2 Linear time varying and non-linear models

For linear models the Kalman filter provides the exact solution for the filtering problem. Another possibility in the more general case, is a linearization of the non-linear differential equation:

$$d\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t, \mathbf{u}_t, \theta, t)dt + \mathbf{G}(\theta, t)d\mathbf{w}_t \tag{2.31}$$

with the observations $\mathbf{y}_k$ taken at discrete time instants, t_k, as a function $\mathbf{h}$ of time, state, input and parameters

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{u}_k, \theta, t_k) + \mathbf{e}_k \tag{2.32}$$

where $\mathbf{e}$ is a Gaussian white noise process independent of $\mathbf{w}$, and $\mathbf{e}_k \sim N(\mathbf{0}, \mathbf{S}(\theta, t_k))$ around some reference point as indicated in Eq. (2.8), Section 2.2.2. Calculating this reference trajectory by integrating the state equation after setting the noise equal to zero, is called the linearized Kalman filter methods.

Paper **A** shows the equations for the Kalman filter. The extended model suggested in Paper **F** is a linear time varying model, which is then described in Paper **G**.

Missing input observations due to e.g. failure in the measurement system is handled conveniently by the Kalman filter algorithm. If the observation at time t is missing, new information is not obtained and hence, the state $\boldsymbol{x}_t$ and $\boldsymbol{P}_t$ is not updated; i.e.:

$$\hat{\boldsymbol{x}}_{t|t} = \hat{\boldsymbol{x}}_{t|t-1} \qquad \hat{\boldsymbol{P}}_{t|t} = \hat{\boldsymbol{P}}_{t|t-1} \tag{2.33}$$

The recursions of the Kalman filter are obtained by again using the updated estimations for $\boldsymbol{x}_t$ and $\boldsymbol{P}_t$ to make predictions for $t+1$.

2.6.3 The extended Kalman filter

The Kalman filter is applied to linear state-space models. In many natural systems often non-linearities are present. However, it is possible with the extended Kalman filter to estimate parameters in non-linear noisy models. The extended filter works by making local linear approximations around the state at each time instant. This is a better choice than the linearized

Kalman filter. Principally, there is no difference between parameter estimates and state estimates (Ljung (1979)). For the initial presentation and a more thorough treatment of the extended Kalman filter, see Jazwinski (1970).

The model in Papers **A, B** is non-linear, as is one of the models tried in Paper **G**, while most of the models tried out in Paper **C** only included non-linearities in the parameters.

In order to make the integration of the prediction formula, computationally feasible and numerically stable for stiff systems (i.e. systems with very different time constants), the time interval $[t_k, t_{k+1}[$ is sub-sampled and the general differential Eq. (2.31) is linearized using the state estimate at the given sub-sampling time. The concept is shown in Figure 2.3.

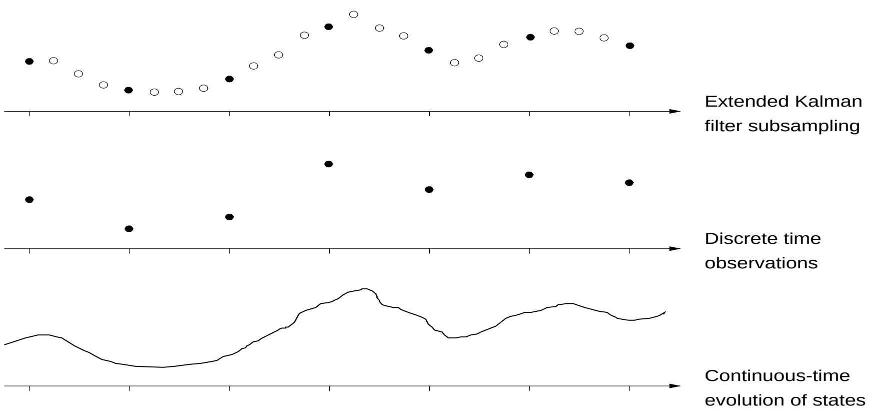


Figure 2.3: The continuous time and discrete time scales considered in the extended Kalman filter. The open dots are the sub-samples, at which time the differential equation is linearized using the state estimate.

The observations $\mathbf{y}_k$ are taken at discrete time instants, t_k (see Eq. 2.32). In this case the extended Kalman filter is used as a first order approximative filter. Being linearized about $\hat{\mathbf{x}}_t$ the state and covariance equations have a structure similar to the Kalman filter equations for linear systems. Hence, it is possible to reuse the numerically stable routines implemented for the Kalman filter (for more detail, see e.g. Melgaard (1994), Madsen and Holst (1996)).

The matrix $\mathbf{C}$ is obtained from the linearization of the measurement equation (Eq. 2.32):

$$\mathbf{C}(\hat{\mathbf{x}}_{k|k-1}, \mathbf{u}_k, \theta, t_k) = \frac{\partial \mathbf{h}}{\partial \mathbf{x}}\Big|_{x=\hat{x}_{k|k-1}} \tag{2.34}$$

and $\mathbf{A}$ is obtained from the linearization of the system equation,

$$\mathbf{A}(\hat{\mathbf{x}}_t, \mathbf{u}_t, \theta, t) = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\Big|_{x=\hat{x}_t} \tag{2.35}$$

The prediction of the output is given by:

$$\hat{\mathbf{y}}_{k+1|k} = \mathbf{h}(\hat{\mathbf{x}}_{k+1|k}, \mathbf{u}_{k+1}, \theta, t_{k+1}) \tag{2.36}$$

The formulas for predicting the mean and covariance of the state-vector are then given by:

$$\begin{aligned}
\frac{d\hat{\mathbf{x}}_{t|k}}{dt} &= \mathbf{f}(\hat{\mathbf{x}}_{t|k}, \mathbf{u}_t, \theta, t)) \ , \qquad t \in [t_k, t_{k+1}[& (2.37)\\
\frac{d\mathbf{P}_{t|k}}{dt} &= \mathbf{A}(\hat{\mathbf{x}}_{t|k}, \mathbf{u}_t, \theta, t)\, \mathbf{P}_{t|k} + \mathbf{P}_{t|k}\mathbf{A}^T(\hat{\mathbf{x}}_{t|k}, \mathbf{u}_t, \theta, t) & \\
&+ \mathbf{G}(\theta, t)\mathbf{G}^T(\theta, t) \ , \qquad t \in [t_k, t_{k+1}[& (2.38)
\end{aligned}$$

where $\mathbf{A}$ is given by Eq. (2.35). For the state propagation the equation become:

$$\begin{aligned}
\frac{d\hat{\mathbf{x}}_t}{dt} &= \mathbf{f}(\hat{\mathbf{x}}_j) + \mathbf{A}(\hat{\mathbf{x}}_j)\{\hat{\mathbf{x}}_t - \hat{\mathbf{x}}_j\} \ , \qquad t \in [t_j, t_{j+1}[& (2.39)\\
&= \mathbf{A}(\hat{\mathbf{x}}_j)\, \hat{\mathbf{x}}_t + \{\mathbf{f}(\hat{\mathbf{x}}_j) - \mathbf{A}(\hat{\mathbf{x}}_j)\, \hat{\mathbf{x}}_j\} \ , \ t \in [t_j, t_{j+1}[& (2.40)
\end{aligned}$$

where $[t_j, t_{j+1}[$ is one of the sub-intervals of the sampling interval $[t_k, t_{k+1}[$, assuming the sampling interval has been divided in n_s sub-intervals. In these derivations only the state dependency is given for simplicity. Eq. (2.40) is a linear ordinary differential equation which has the exact solution

$$\begin{aligned}
\hat{\mathbf{x}}_{j+1} &= \hat{\mathbf{x}}_j + (e^{\mathbf{A}(\hat{\mathbf{x}}_j)\tau_s} - \mathbf{I})(\mathbf{A}(\hat{\mathbf{x}}_j)^{-1}\mathbf{f}(\hat{\mathbf{x}}_j)) & (2.41)\\
&= \hat{\mathbf{x}}_j + (\mathbf{\Phi}_s(\hat{\mathbf{x}}_j) - \mathbf{I})(\mathbf{A}(\hat{\mathbf{x}}_j)^{-1}\mathbf{f}(\hat{\mathbf{x}}_j)) & (2.42)
\end{aligned}$$

where $\tau_s = t_{j+1} - t_j = \tau/n_s$, and τ is the sampling time. Correspondingly the equation for the state covariance become:

$$\mathbf{P}_{j+1} = \mathbf{\Phi}_s(\hat{\mathbf{x}}_j)\mathbf{P}_j\mathbf{\Phi}_s(\hat{\mathbf{x}}_j)^T + \mathbf{\Lambda}_s(\hat{\mathbf{x}}_j) \ , \tag{2.43}$$

The algorithm to solve Eq.s (2.37) and (2.38) is to use $\hat{\mathbf{x}}_{k|k}$ and $\hat{\mathbf{P}}_{k|k}$ as starting values for (2.42) and (2.43) and then perform n_s iterations of (2.42) and (2.43) simultaneously. This algorithm has the advantage of being numerically stable for stiff systems and still computationally efficient, since the fast and stable routines of the linear Kalman filter can be used.

Many of the models tried out in Paper **C** was suggested by Cosby and Hornberger (1984). However the implementation of the extended Kalman filter described above is quite different from the one used in that paper. Cosby and Hornberger (1984) obtained the final parameter estimates by averaging the recursive estimates. This requires a reasonable prior specification of noise covariances for the state variables. However, Ljung (1979) stated that divergence of the parameter estimates might be traced to misspecification of the state noise covariance. And in general the problem is to give a specification of the state noise covariance. It may also be caused by using inadequate start-values.

Beck and Young (1975b) found that the algorithm for the extended Kalman filter was too dependent of a good choice of initial parameter estimates. This presumably lead to the development of another recursive estimation method based on the Instrumental Variable technique in both single and multi variable form (Young and Whitehead (1975), Whitehead and Young (1975)). This was claimed to posess superior convergence proporties with greater effiency.

Here, parameters are estimated directly thus avoiding to extend the state vector and the intermediary recursive estimates. As the noise covariance are estimated along with the parameter estimates, the potential of misspecification of noise is avoided. The extended Kalman filter is solely used to obtain the one-step predictions, which enters the likelihood function and the method is rather simple to use.

2.7 Residual analysis and models validation

The final step of model building is to check if the model describes a set of measured observations satisfactory, and if the residuals are independent realizations of white noise. If the model is not satisfactory, the checking may often provide a hint to an alternative improved model. A typical validation is to check the model on a set of data, that has not been used for estimation, e.g. checking whether the variance of the prediction error is of the same order. Unfortunately, most data sets used in this thesis were not of a length that permitted using only half of the data for estimation and the other half for validation. This was partly due to the importance of using a set that included the impact from the rather rare rain events.

The model is furthermore evaluated statistically by testing each of the parameters with a t-test on the hypothesis that the parameter is equal to zero. The equations given in Section 2.5, Eq.s (2.28)–(2.30), were used to estimate the variance of the parameter estimates and these variances then serve as a basis for calculating the t-test values.

There are a number of statistical tests available for checking if the independence assumptions are violated and the program used for modelling (CTLSM) includes several of these diagnostics. However, it must not be forgotten to carry out a graphical check in the form of a visual inspection of the prediction and simulation errors.

2.7.1 Auto- and partial auto-correlation functions

The auto-correlation function $\hat{\rho}(k)$ is a measurement of the correlation between the present observation and the past observations, with all the time lags taken into account. It can be interpreted as the memory in the system. If the residuals ϵ_t is white noise, then

$$\hat{\rho}(k) \in N(0, 1/N) \qquad \text{(approx.)} \tag{2.44}$$

Based on this the 2σ (an approximative 95 % confidence interval) limits may be calculated, thus checking if the individual values are significantly different from zero.

The partial auto-correlation function represents the additional correlation contained in the observation k time lags before the present observation, given the observations in the intermediate lags. This can be used to identify the model order, i.e. how many state variables should be included. For the creek reported in Paper **A**, **B** the partial auto-correlation function contributions from lag 1 and 2, indicating a model with two states (but not which two). Given the information that in receiving waters oxygen depletion is caused by the degradation of the organic matter, it was a prudent choice to include a state for organic matter, expressed as BOD (Biological Oxygen Demand). This is not easily measured and here the advantage of the state space modelling formulation is clearly seen: States that are not measured, can be taken into account.

2.7.2 Cumulative periodogram

The cumulative periodogram is based on the fact that the variation of white noise is equally distributed on all frequencies. For the frequencies $f_i = i/N$, where $i = 0, 1, 2, \ldots, N/2$, the periodogram may be calculated for the residuals, ϵ_t as:

$$\hat{I}(f_i) = \frac{1}{N}\left[\left(\sum_{t=1}^{N/2} \epsilon_t \cos 2\pi f_i t\right)^2 + \left(\sum_{t=1}^{N} \epsilon_t \sin 2\pi f_i t\right)^2\right] \tag{2.45}$$

where $I(f_i)$ indicates how much of the variation from the residuals ϵ_t that can be attributed to f_i, i.e. a frequency domain description of the variation of ϵ_t.

The scaled cumulative periodogram $C(f_i)$ is obtained by:

$$\hat{C}(f_j) = \frac{\sum_{i=1}^{j} \hat{I}(f_i)}{\sum_{i=1}^{N/2} \hat{I}(f_i)} \tag{2.46}$$

which is a non-decreasing function, defined for the frequencies f_i.

The total variation for N observations equals $N\sigma_\epsilon^2$. Therefore the theoretical periodogram for white noise is $I(f_i) = 2\sigma_\epsilon^2$ and hence the theoretical cumulative periodogram $C(f_i)$ for white noise is a straight line from origo to (0.5,1). A confidence interval around the straight line can be calculated as a Kolmogorov-Smirnov test (see Melgaard (1994)). Figure 2.4 shows the cumulative periodogram, $C(f_i)$ for three different situations: A successful model, **a**), and two unsatisfactory models, **b**) and **c**). Example **b**) is typical of a model without a sufficient description of low frequency variations, while **c**) shows an example where improvement can be made by extending the high frequency description.

Examples of the cumulative periodogram for sets of residuals are shown in Papers **B**, **G**.

Besides the above mentioned test, the models reported in this thesis were validated by analysis of the cross correlation for the residuals and the input. For a sufficient model, these should show no dependence, i.e. the cross correlations should be close or equal to zero.

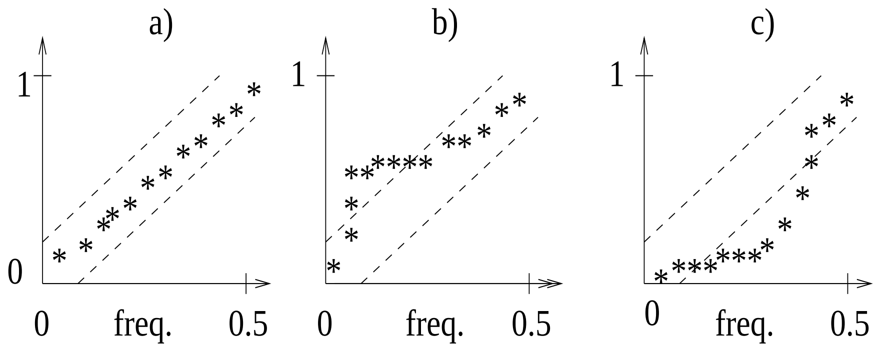


Figure 2.4: The cumulative periodogram for one successful and two unsatisfactory models.

2.7.3 Summary for model checking

In this section several methods have been presented, which can be used in the final modelling check. The assumption about the Wiener process for the stochastic terms was tested by testing if the residuals can be assumed to be white noise, since integration of a Wiener process leads to white noise. This is carried out in the time domain by using the autocorrelation function and in the frequency domain by the cumulative periodogram. Some of these methods have properties that make them useful to indicate the next step towards a satisfactory model. Thus a list of methods may include the following:

- Test the parameters if they are significantly different from zero and check the correlation matrix between the estimated parameters.
- Test if the residuals can be accepted to be white noise and if they are independent of the input.
- Compare measured and simulated output, and the one-step prediction error.
- Compare estimated physical parameters with those recorded in similar investigations and any prior information about the system.

If the model meets these criteria, a validation on a separate set of data should be carried out if possible. The final acceptance can not be based purely on objective tests. The user will make a subjective decision based on the purpose and the intended use for the model. The model accuracy of a model intended for control, for instance, will depend on the expenses (monetary and humanitarian) the damage from a failure of early warning will inflect on the users. In general, for a model intended for control, the short time prediction is the most important. It has been said that "in general, models are no good; however, some are useful". To this may be added: "if they are applied correctly".

2.8 Deterministic and grey box modelling

Grey box modelling of water quality was suggested already in Young and Beck (1974), Beck and Young (1975a), Young and Whitehead (1975) and Thyssen (1980) who advocated the use of what they called *internally descriptive* or *databased mechanistic* models, which are very similar to what is here called Grey-box models. The models have in common that the existing knowledge of the processes is combined with the statistical knowledge of the uncertainty in both the model structure and in the measurements. Jamieson et al. (1972) used a deterministic rainfall runoff model combined with a stochastic model for the model discrepancy and in effect argued for the grey box concept. Whitehead and Young (1975) used a similar concept in a water quality model combining a deterministic streamflow forecasting model, and a stochastic black box model for rainfall runoff. These examples essentially comprises what we call a grey box model, though in discrete time as opposed to continuous time, as considered here. In fact, in part two of the discrete time stream flow and water quality model, in the discussion of water quality models Whitehead et al. (1979) and Whitehead et al. (1981) mentions discrete time grey box models as an alternative to the Instrumental Variable method employed.

When modelling dynamical systems, such as the oxygen level in a river, the concept of so-called grey box modelling have some very attractive properties. This concept combines prior physical, biological and chemical knowledge about a given system with information from data. Hence, it becomes possible to describe the total variation by introducing stochastic terms and thus describe the system through a set of coupled stochastic differential equations. That the models include parameters well known by e.g. biologists facilitates communications with other professionals. Inclusion of stochastic terms is important in the modelling of complex systems where errors in measurements may be dominant. With respect to dynamic models for the level of oxygen in rivers, the knowledge of cause and effect is fairly complete. But variations, due to e.g. impact from rain events, makes it necessary to focus on stochastic models. Better information is not necessarily more precise information and uncertainty is not ignorance – it is knowledge. Seemingly precise forecasts may hide uncertainties and thus provide less information than a forecast showing a range of possible values.

When calibrating deterministic (called physically based in Beven (1989)) models, only **selected** events (for instance when it rains) are used in the final modelling, even though there may be data from a longer period (in order to include the events needed). Stochastic modelling, however, use **all** the data collected to identify the model and estimate the parameters of this model. In reality, a monitoring program never yield perfect data. Errors occur in the equipment, as well as vandalism (either intentional or not) from bypassers and playing children. Statistical methods may be taken into use to repair the data so the entire series may be used (see e.g. Jacobsen and Voss (1994a)), or a sufficient segment may be taken out and used. The significant difference is that the final time series of data includes the interesting events, as well as the periods in between these events.

Another problem with some deterministic models is the large number of parameters, where a much reduced set of parameters may describe the system equally well. Beven (1989) discusses the limitations of physically based models and mentions a model with 2400 parameters! With a stochastic model, a much simpler lumped conceptual model might be much more useful. Beck (1991) cites Botkin (1977): "If small is beautiful and big is ambiguous, then middlesized is meaningful.

2.8.1 The grey box concept

To understand the grey box concept, a short introduction to the three model types most frequently used in the modelling of dynamical systems is useful. These are divided by colours: Black box, white box and grey box.

Traditionally *black box* models are identified on the basis of the data only. The parameters may not have a physical interpretation and such models may not be representative of the general dynamics and physics of the system. Examples of these are auto regressive models (AR), moving average models (MA) and combinations of these (ARMA). A deterministic model is sometimes called a *white box* model. It is purely based on known deterministic equations for a system. Equations for complex systems are often combined with detailed equations for smaller sub-systems, which often have been subjected to various approximations or is a result of laboratory research. It is therefore difficult to quantify the uncertainty of the total model.

The grey colour suggests that this model is a mix of the two previously described models. This kind of model combines available knowledge and known equations of the system with statistical modelling tools. Based on this, model parameters are estimated and these have physical meaning and may readily be interpreted. The three model types are illustrated in Figure 2.5.

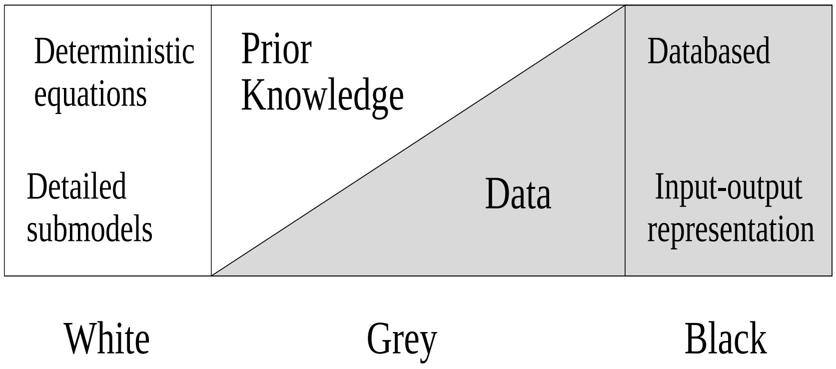


Figure 2.5: The grey box modelling concept. The shade of grey for a given model may be expressed as the relationship between the amount of knowledge and data available.

The difference between the databased mechanistic models of Young and Beck (1974), Beck and Young (1975a), Young and Whitehead (1975) and the grey box modelling in this thesis, lies mainly in the approach. With the databased mechanistic modelling method, the model structure is inferred from purely databased investigations, such as transfer functions. These are confronted by reality and additional or more complex hypotheses are only incorporated when deemed necessary (Beck et al. (1993)). With grey box methods, a well-known deterministic model or hypothesis is proposed and uncertainties are added. Then if some of the model elements are clearly redundant, they are eliminated. Referring to Figure 2.5, the two approaches come from opposite sides, so to speak. However, they also differ with respect to the time frame. The databased mechanistic models are discrete time models, while the grey box models are continuous time. This is important, since prior physical knowledge is often given as continuous time relations.

With any system there are always uncertainties and phenomena that have not yet been detected or fully realized. For natural systems, such as rivers, meteorological effects are one example of uncertainty. Therefore, combining available knowledge with information from data is an optimal approach towards understanding a given system. The experience obtained by using the grey box modelling concept is of major importance and have been applied in all included papers, except Paper **E**. The importance of the grey box concept apply not only to modelling the dynamics of the oxygen content in a river, but also in modelling other complex dynamical systems (e.g. building components (Hansen et al. (1996)) and the heat dynamics of greenhouses (Melgaard et al. (1994)). Eventually models for the entire integrated system could be developed.

2.8.2 Advantages

The advantages of the using the grey box model with a continuous time formulation are that:

- Missing observations are easily accommodated (see e.g. Paper **A**).
- Available physical knowledge and statistical modelling tools is combined to estimate the parameters of a rather complex dynamical system.
- Known equations are included and used to formulate the model.
- The method facilitates a direct physical interpretation of the estimated parameters. This eases communication with biologists and other professionals.
- Non-linear and time-varying models are possible (see e.g. Papers **A, B, C, G**).
- States that are not measured, such as BOD, may be incorporated (see e.g. Papers **A, B, G**).
- The models typically have fewer parameters than discrete time models, black box models and deterministic models.

Improvements are possible by combining detailed comparison with data and the use of physical facts, as direct physical interpretation of the estimated parameters is possible.

Non-linear relationships may be taken into effect. These have typically been linearized, which adds to the overall uncertainty in the model. Grey box modelling further have the advantage that it is possible to test different model structures. Stigter and Beck (1994) use the Kalman filter application to discern between different model structures, while this is done through the hypothesis testing of the significance of the parameters in this thesis.

2.9 Output error estimation methods

There are many other ways of combining deterministic and stochastic models. Whitehead and Young (1975) coupled a deterministic model with a black box model (ARMA) and estimated the parameters recursively in discrete time. The stochastic input stemming from precipitation was modelled as a black box model and then coupled with a deterministic model for water quality and tried both on a model for water flow and on an oxygen/organic matter (DO–BOD) model on a 55 kilometer stretch of the Bedford-Ouse River system in East-England.

Winter (1988) established a receiving water quality model based almost exclusively on deterministic nitrogen fractions such as nitrate and ammonia. Along with this effort, Wittenberg (1992) has introduced a collection of differential equations for nitrogen exchanges. The purpose was to assess the effect of combined sewer overflows, (CSO). After simulating over a long time, the results of this simulation was investigated statistically.

In paper **E** an output error model have been used to calibrate the parameters in an otherwise purely deterministic model. The integrated model used in this investigation, is a combination of a simplified model for urban runoff with pollution, and a model for oxygen depletion in rivers. In both models the traditional hydrograph method (Rauch and Harremoës, 1996) of routing the water volume from element to element is used. It is similar to a deterministic program that PH-Consult is the owner of, which models rainfall: SAMBA (it is especially made for computations of pollution loads and for optimisation in real time control). However, calibrating the parameters can be a long tedious phase usually solved by heuristic procedures. Essentially, it is just an optimization routine.

The automatic calibration routine will assist in a faster and more efficient use of time when working with traditional models like these and others.

2.9.1 An automatic calibration procedure

Working with deterministic models always include a phase of calibrating the parameters of the model. An experienced operator can reduce the time spent on this task considerably, but still the time allocated to this problem is significant.

A traditional statistical method, the so-called output-error estimation method, is used to calibrate system parameters for both runoff and receiving water quality in Paper **E**. The output-error method is a well known principle (Söderström and Stoica (1989)) based on the squared difference between a measured time series and a simulated output. By minimizing this difference, it is possible to find the parameters, which provides the best fit of the model to the measured data for that situation. This is provided the model structure is well behaved, i.e. that the parameters are structurally identifiable. Applied to a deterministic model, several parameters may be calibrated automatically using a measured time series. The calibration in Paper **E** was done in two steps, since it was advantageous to use the causality between the two systems. First the runoff from the watershed was calibrated, based on data from an overflow structure, to get the runoff hydraulics correct. Then, the receiving water model was calibrated based on the overflow from the calibrated catchment and the oxygen concentration in the river.

A Monte Carlo validation of the calibration routine was carried out for a synthetic case and it was shown that even with data contaminated by a quite high level of noise, the calibration method yield results that are not significantly different from the known values (see Paper **E**).

Developing an automatic calibration routine, that may cut down on time required for this task (calibration by hand) is an example of so-called *regular technology* described in (Jacobsen (1997)), where a known task is performed in a new and different way (using a statistical method instead of a heuristic method).

2.10 Summary of time series analysis concepts

In this chapter various time series analysis concepts have been presented. Starting with the considerations of whether sub-sampling or filtering should be performed on the data, which are going to be used in the modelling effort. Then the state space model is introduced as the model type used in all the models that form the basis for this thesis. Having decided on the model type, some considerations made, as to whether the problem and the model specified is **a**) persistently exited and **b**) identifiable. After these preliminaries the methods for estimating the parameters are described and various tools for model validation is presented. Then the grey box concept is introduced with a list of the advantages of this modelling concept. Along with the presentation of the grey box model, other models, which in essence may be called grey box models are briefly presented and the differences pointed out. Finally, the output-error estimation method is outlined with a summary of an applied example.

Chapter 3

Qualitative concepts of the hydraulics and the oxygen dynamics

3.1 Introduction

Of all the water belonging to our planet, 96.5 % is comprised by oceans, while polar ice and groundwater each accounts for 1.7 % (i.e. 3.4 % total). This leaves only 0.1 % for surface and atmospheric water. Thus, the atmospheric water (the driving force of surface water hydrology) system contains only 12900 km^3 of water: less than one in 100000 of all the earths water (Chow et al. (1988)). However, the environmental engineering concerns with these 0.1 % are numerous. Some of the applications are:

- Wastewater treatment and disposal
- Pollution abatement
- Recreational use of water bodies
- Fish and wildlife protection
- Flood prediction and protection

The hydrologic cycle play an active role in the quality of water. Precipitation washes particulate materials and gases from the air and surface water flow carries with it metals and particular matter from the streets in the cities. Particles may have organic and inorganic compounds attached and surface flow from farmland and industrial areas may also seriously affect the quality of water in the receiving waters. Of these influences, the sudden impact from large rain events is the subject of much concern and therefore of prime research interest. This is not new, Whitehead and Young (1975) stated: *It is the transient violations of water quality standards that cause most problems in the short term.* However, at that time data collection was not as sophisticated as today and computer power not as readily available.

When large rain events occur, the waste water treatment plants cannot keep up with the huge amounts of combined rain and waste water. In most places, storage basins are built to detain the excess water, but it is not possible, nor economically feasible to detain all excess amounts of water. Hence, combined sewer overflows (CSO's) are led into the receiving water without any purification. Thereby organic matter is introduced into the stream and degraded by the use of oxygen, which may cause unpleasant sight or smell when the oxygen is exhausted. Oxygen is of vital importance to flora and fauna.

In Denmark about 98 % of drinking water is based on ground water. Therefore we are not dependent of clean receiving waters for that purpose. However, in an affluent society, rivers and streams are considered a quality recreational advantage, that should be protected and cared for; especially in urban areas. Thus, healthy ecosystems in rivers and streams, with abundant fish-life and bio-diversity, becomes a balance between using the rivers as receiving waters and the amount of money we are willing to spend on keeping the rivers clean for recreational facilities. The tools to make the right decisions only become available through research into the processes that governs the quality of water.

3.1.1 Model classification

Hydrologists classify models as either *physical* or *abstract* models, where physical means an actual hands-on model constructed in a reduced scale and abstract models represents mathematical models. For systems and time-series analysts, however, a physical model is often a terminology used to describe a mathematical model which incorporates physical laws when linking input to output.

Mathematical models can be either deterministic or stochastic. In deterministic models, identical input will yield identical output. A stochastic model contains random components, which means that identical input will yield different outputs, though within a specified or experimental probability distribution. Consider a deterministic model with a stochastic term to account for measurement errors. If the measurement errors are normally distributed, then the output of such a stochastic model will also be normally distributed.

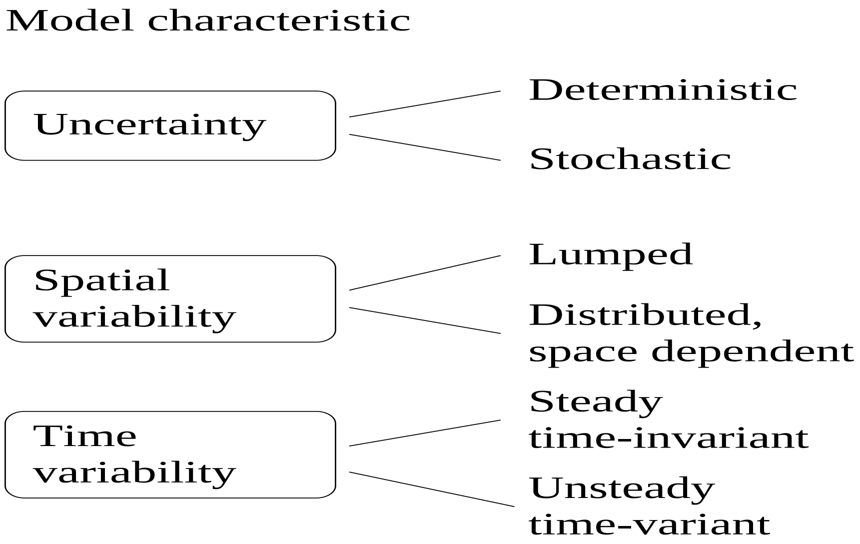


Figure 3.1: Classification of hydraulic models with respect to variability in randomness, space and time (Adapted from Ven Te Chow et al. (1988).

Having decided upon a deterministic or stochastic model, the next step is to decide whether to deal with a lumped model or a distributed model. A lumped model is a model where the system is spatially averaged or regarded as a single point (or points) in space, such as considering precipitation to fall uniformly over a watershed, thus ignoring the spatial variation. A distributed model would include the variation that takes place at various points in the catchment, as well as the variation over time at these various points, as a rain event passes the catchment.

Time-varying or steady state models constitute the final level of the model hierarchy. Chow et al. (1988) classify deterministic flow models as either steady or unsteady flow models, while stochastic models are classified as time-invariant or time-varying. However, in time series analysis there is a tradition for just one classification; that of time-invariant or time-variant. A system is time-invariant when the output at time t from an input signal at time t is the same as the output at time $t-\tau$ from an identical input at any other time $t-\tau$. I.e. the way the system processes input into output does not change with time. Mathematically this is expressed as:

$$y(t) = \mathcal{F}[x(t)] \Rightarrow y(t-\tau) = \mathcal{F}[x(t-\tau))] \tag{3.1}$$

where $y(t)$ is the output and $x(t)$, the input. Thus, time-invariance relates to system processing and a system may be time-invariant even though its input and output varies with time. A dynamic model that varies with time may thus describe either a time-variant or a time-invariant system, and this is the same for both deterministic and stochastic models.

Working with a lumped stochastic model, the system does not in general have a unique formulation. Better parameter estimates may be obtained by restricting the number of free parameters in the models by incorporating physical knowledge into the model. If one or more of the variables change with each timestep, such as a variable being dependent on one of the inputs like temperature or the solar radiation, the model is time-varying.

3.1.2 Model types in this thesis

In this thesis three model types have been investigated. The integrated model in paper **E** is a deterministic, distributed, un-steady state model. The concept of this model is shown in Figure 3.2. The figure shows how the input flow is different from the output, i.e. it is an unsteady or time varying flow. If the waterlevel is parallel to the river bottom, the flow is uniform.

Q_{in} is the flow over time t into a stretch of river of length L, and Q_{out} is the flow out of that stretch. In Figure 3.3 a grid shows how the variation occurs in time and space.

The spatial distribution is modelled by a traditional hydrograph method where the water is routed from element to element. The catchment and the receiving water body is divided into these elements for calculation purposes (Rauch and Harremoës (1996)).

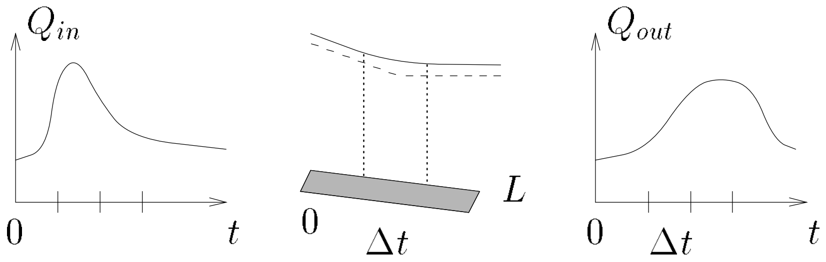


Figure 3.2: Deterministic, distributed, non-steady state model. (Adapted from Ven Te Chow et al., 1988). Since spatial variations are included, varying waterlevels may be accommodated.

The hydrologic model in paper **F** is based on a linear reservoir description and is a stochastic, lumped, dynamic model. The description includes stochastic terms for model errors, as well as for measuring errors. It does not explain the variation in space between the two measuring stations, nor does it explain the spatial variation over the catchment from the rain events. This concept is shown in Figure 3.4.

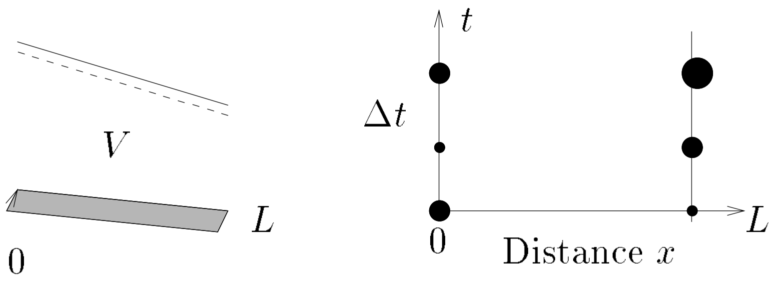


Figure 3.4: Lumped, time-varying model. (Adapted from Ven Te Chow et al., 1988). There are no dots over the distance x, as there is no spatial dependence. Volume V, rather than flow is used.

The model varies in time as shown by the unequally sized dots at distance zero and length L with no dots in between, as there is no spatial dependence. If the spatial variation is negligible or the uncertainty introduced by ignoring it is less than the overall model uncertainty, it is reasonable to use a lumped model. The flow is described by a stochastic distribution as shown in Figure 3.5. In the expansion of that model (suggested as an improvement and described in Paper **G**) the parameter for the surface area of the receiving water is time dependent, as is the oxygen dynamic processes. Hence, the extended model is a time-varying model.

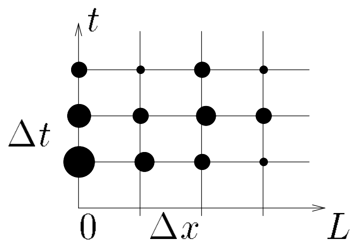


Figure 3.3: Time/space grid showing how variations occur both in time and space. (Adapted from Ven Te Chow et al., 1988).

All the other models (described in papers **A**, **B**, **C**, **D**) are stochastic, lumped, time-varying models. In all these papers the photosynthesis is a function of the measured solar radiation and the saturated oxygen concentration is a function of the temperature, except in papers **B**, **C** where the saturated oxygen concentration was given. Figure 3.5 shows how the flow, Q, belongs to a probability distribution, P(Q) that have no spatial variability. The characteristics of the distribution varies with time.

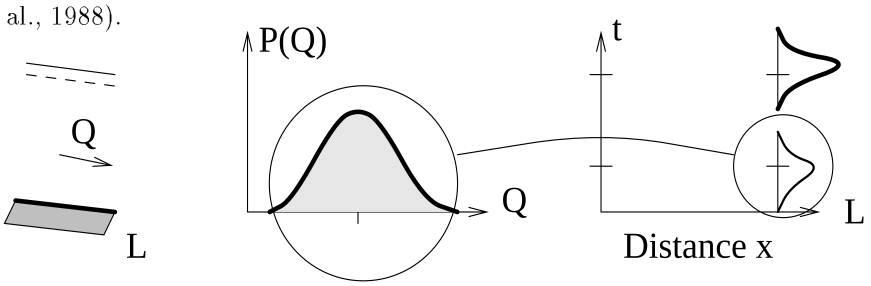


Figure 3.5: Stochastic lumped, time-varying model. (Adapted from Ven Te Chow et al., 1988). The distribution varies with time, not in space.

3.2 The hydrodynamics

Since our knowledge of hydrological phenomena is far from perfect, it is not possible to describe all hydrological processes with exact physical laws. Typically, various deterministic models, derived from laboratory studies, performed under ideal conditions, are compiled into large models, which are often overparameterised. An alternative, more simplified way of looking at these phenomena is the systems concept. Such models are constructed by relating the input to the output by an operator or *transfer function* and combining this with the knowledge one has of the system at hand. A larger system may be created by a combination of sub-systems.

A hydrologic system is defined by a control volume within well-defined boundaries, which accept water and other inputs, operate on them internally and produce one or more output. Hydrological systems are inherently random because the major input is precipitation, which is stochastic in nature.

3.2.1 Waterflow

To study water in motion, velocity is the most important factor. Already in the 15'th century, Leonardo da Vinci (1452-1519) made the first systematic studies of velocity distributions in streams.

There are two ways to mathematically express water in motion. By

- the Lagrangian method
- the Eulerian method

The Lagrangian viewpoint considers individual water particles, following each one at all times. This is the traditional approach, which focus on the spatial variation. The other point of view focus on a specific point in space and considers the motion of all water particles that pass this point as time goes by, i.e. focusing on the variation in time. This is known as the Eulerian viewpoint.

The Lagrangian method.
In this case the particle velocity is obtained by differentiating the particle's position with respect to time. The position may be represented by a vector, which in the Cartesian coordinate system is expressed as:

$$\mathbf{r}(t) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \qquad [\mathrm{m}] \tag{3.2}$$

where $\mathbf{i}, \mathbf{j}$ and $\mathbf{k}$ are the unit vectors. Differentiating Eq. (3.2) with respect to time, we obtain the velocity of the fluid particle:

$$\begin{aligned} \mathbf{v}(t) &= \frac{dx}{dt}\mathbf{i} + \frac{dy}{dt}\mathbf{j} + \frac{dz}{dt}\mathbf{k} \\ \mathbf{v}(t) &= v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k} \qquad [\mathrm{m/s}] \end{aligned} \tag{3.3}$$

The Eulerian method.
In this case the fluid particle's velocity depends on the specific point in space and time:

$$v_1 = f_1(x, y, z, t) \qquad v_2 = f_2(x, y, z, t) \qquad v_3 = f_3(x, y, z, t). \tag{3.4}$$

With this method we stand in one place and observe the motion of particles passing by this point, as opposed to the Lagrangian method in which we move along with the same velocity as a specific water particle and track its position as time passes. I.e. just as models are classified with respect to spatial distribution and time variance, so are the observations of the motion of water.

The Lagrangian viewpoint is simple to apply with unsophisticated equipment, such as any light floating object, whereas the Eulerian method requires more sophisticated measuring equipment, such as electronic observation devices. Most monitoring programs today are based on the Eulerian method.

3.2.2 Flow classification

Flow is also classified with respect to spatial distribution and time variance. The path followed by an element of a moving water particle is called a flow path or a characteristic (Di Toro (1969)). A streamline is defined as a curve where its tangent is in the direction of the fluid velocity at any point.

With the Eulerian method we describe the spatial variation of the velocity field as a function of position along a streamline and time: $\mathbf{v} = \mathbf{v}(s, t)$. A flow is uniform if the velocity does not change from point to point along any of the streamlines in the field of flow. Therefore, the streamlines depicting such flow are straight and parallel. Otherwise, there is a directional change, or speed change, respectively. The conditions of uniform and non-uniform flow are expressed mathematically by:

$$\frac{\partial \mathbf{v}(x,t)}{\partial x} = 0 \quad \text{uniform flow}, \qquad \frac{\partial \mathbf{v}(x,t)}{\partial x} \neq 0 \quad \text{non-uniform flow}. \tag{3.5}$$

A flow classification based on the variation of velocity with respect to time at a given point in a field of flow are mathematically defined as follows:

$$\frac{\partial \mathbf{v}(x,t)}{\partial t} = 0 \quad \text{steady flow}, \qquad \frac{\partial \mathbf{v}(x,t)}{\partial t} \neq 0 \quad \text{un-steady flow}. \tag{3.6}$$

When he total derivative with respect to time equals zero, we have stationarity:

$$\frac{dv}{dt} = 0 \qquad \text{i.e.} \qquad \frac{\partial \mathbf{v}(x,t)}{\partial x} + v\frac{\partial \mathbf{v}(x,t)}{\partial t} = 0 \tag{3.7}$$

Systems transformation

An input, $Q_{in}(t)$, into a model is transformed to an output, $Q_{out}(t)$, from a model via a transformation equation, or the system performs a transformation of the input, $Q_{in}(t)$, into the output, $Q_{out}(t)$ represented by:

$$Q_{out}(t) = \Omega \, Q_{in}(t) \tag{3.8}$$

where Ω symbolize the transfer function between the input and the output. The transfer function can either be a constant, an algebraic or a differential equation. Hence, Ω is either a linear, an algebraic or a differential operator.

$$\Omega = \frac{Q_{out}(t)}{Q_{in}(t)} = C \tag{3.9}$$

$$\Omega = \frac{Q_{out}(t)}{Q_{in}(t)} = a + b^2 \tag{3.10}$$

$$\Omega = \frac{Q_{out}(t)}{Q_{in}(t)} = \frac{1}{1+kD} \tag{3.11}$$

where D is the differential operator. If k is a function of either the input or the output, then we have a non-linear system.

Surface water

When a rain event passes over a hydrologic catchment a number of surface water processes take place. First, surface storage (ponds etc.) begins to fill, followed by water infiltration into the soil and soil moisture storage. The surface storage and the infiltration is covered by initial loss and the time it takes, until the water actually starts its surface flow, is called the detention time. Figure 3.6 shows the concepts of the surface runoff.

The time it takes water to flow from the furthest part of the catchment to the outlet is called the *time of concentration*. Outlets are typically rivers, lakes and the sea. Also flow from springs may contribute to surface

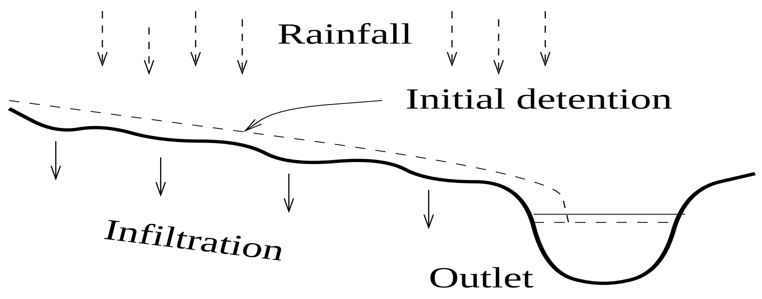


Figure 3.6: Surface flow over a catchment produced by excess rain over infiltration. (Adapted from Horton, 1945).

water flow. In urban areas large amounts of the excess water is caught by combined sewer systems and reaches the outlet as treated water from the waste water treatment plant, or untreated as overflow from storage basins. Often the water is routed through sewers quite far away from the area over which it fell. If extraction of groundwater is extensive, this may lead to occasional dry-outs of shallow rivers and streams.

The hydrograph

A unit hydrograph is the unit impulse response function of a linear hydrological system. If there are several impulses over time, a plot of the streamflow vs. time is a hydrograph for that time period.

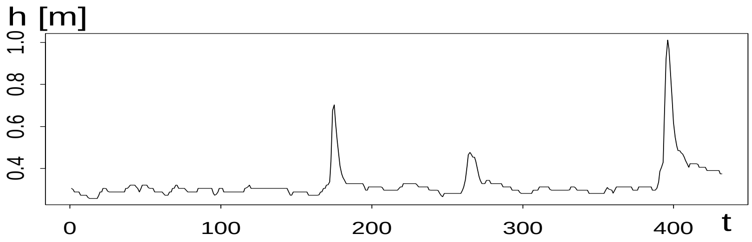
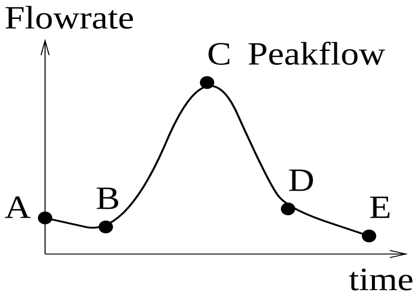


Figure 3.7: The water level of Usserød Å. A typical example of a hydrograph, though here the water level is shown instead of water flow.

A perennial (the time period is a whole year) hydrograph is typical of Danish rivers and streams where neither snow-melt, nor long dry spells occur. There are snow-melt of course, but it is almost negligible compared to regular precipitation in form of rain. However, in relation to the modelling of water quality, expressed by the levels of oxygen, the primary interests is focused on summer data as this is where the lowest levels of oxygen usually occur. Dry spells as a consequence of an arid climate is not a concern in Denmark. The spikes in the hydrograph are caused by rain events which influence the flow indirectly as increased runoff from the waste water treatment plant and overflow from detention basins. Also direct runoff contributes to the extra flow. Direct influence of the rain falling directly upon the river surface is of lesser importance, because of the small size of the rivers and streams in Denmark. The baseflow is the more slowly varying flow when there is no effect from rain. It may include infiltration from groundwater as subsurface flow.



Hydrograph components:

AB Baseflow recession
BC Rising flowrate
CD falling flowrate
DE Baseflow recession
Peak flow occur after the rain has stopped.

Figure 3.8: Components of the streamflow hydrograph during a storm.

Assumptions
The unit hydrograph is a simple linear model that can be used to derive the hydrograph resulting from any amount of excess rainfall. Chow et al. (1988) have listed the following basic assumptions related to this model:

1. The excess rainfall has a constant intensity within the effective duration. Hence, the rain durations should be small.
2. The excess rainfall is uniformly distributed throughout the whole drainage area. Hence, the area should be small.
3. The base time of the duration of direct runoff resulting from an excess rainfall of given duration is constant. This depends on the method of baseflow separation.

4. The ordinates of all direct runoff of a common base time are directly proportional to the total amount of direct runoff represented by each hydrograph, i.e. the principles of superposition and proportionality.
5. The hydrograph is considered unique for a given watershed, and invariable with respect to time.

For most practical purposes these approximations are satisfactory. Stochastic models include terms for model uncertainty to make up for such approximations. This is the advantage in relation to deterministic models. Furthermore, estimations performed in stochastic models include the uncertainty of the parameters.

3.2.3 A linear reservoir model

A linear reservoir model is a model where the output, Q, is linearly related to the storage, V, by a storage constant, k, which has the dimension of time, because V is a volume and Q is a flow rate: $V = k\,Q$ (Chow et al. (1988)). Using the continuity equation, the storage rate is equal to the difference between the input and the output:

$$\frac{dV}{dt} = Q_{in} - Q_{out} \qquad V = A_{surf}\,h \tag{3.12}$$

where A_{surf} is the surface of the river-reach considered and h is the water level. This is demonstrated for a very simplified hydraulic system in Paper **F**. That paper had a suggestion of a further extension of the model, which was later carried out along with a model for the oxygen dynamics. This is reported in Paper **G**.

The linear reservoir model is a lumped parameter model where the dynamics (e.g. hydraulics alone or the water quality model including the oxygen content of the water) of the watershed (or river segment) may be represented by n identical linear reservoirs with the same time constant k. If one unit volume inflow is routed through these n reservoirs, then the response is the unit hydrograph. Each of the reservoirs can be characterized by a state. It is assumed that the watershed response does not change from event to event. Hence, it is also a time-invariant system. The principles of proportionality and superposition applies as to any linear system.

As with most models this is an approximation. Actual hydrologic data are not truly linear but varies with respect to the length of time between rain events because of the earlier mentioned initial losses depending on soil wetness etc. Viewing water level or flow data, such as those in Figure 3.7 in the previous section, clearly show that a run-off phase when the level declines is much slower than when the level rises. In the terminology of Tong (1990), such data do not exhibit reversibility and should not be expected represent a linear system.

If Eq.s (3.12) are formulated as the transfer functions in the systems transformation section, we have

$$\frac{dV}{dt} = Q_{in}(t) - Q_{out}(t) \tag{3.13}$$

$$k\,\frac{dQ_{out}}{dt} + Q_{out}(t) = Q_{in}(t) \tag{3.14}$$

Hence, we obtain Eq. (3.11). If the differential operator, $1 + k\,D$, has n real roots: $-1/k_1, -1/k_2, \ldots, -1/k_n$, then the system is described in a cascade containing n linear reservoirs in a series with storage constants k_1, k_2, ..., k_n, respectively. Here, in the modelling of the water quality dynamics, the same time constant k will be used for all the reservoirs. The impulse response function (introduced in the time series section) corresponding to routing a unit volume through the linear reservoir for time-lag $l = t - \tau$ is obtained via the convolution integral:

$$\begin{aligned}
u_1(l) &= 1/k \exp(-l/k) \\
u_2(l) &= t/k^2 \exp(-l/k) \\
u_3(l) &= \frac{t^2}{2\,k^3} \exp(-l/k) \\
\cdots &= \cdots \\
u_n(l) &= \frac{1}{k\,\Gamma(n)} \left(\frac{t}{k}\right)^{n-1} \exp(-l/k) \qquad (3.15)
\end{aligned}$$

where $\Gamma(n) = (n-1)!$. Starting from rest, $Q_o = 0$, $u_1(l)$ becomes the outflow from the first reservoir and the inflow to the second, while $u_2(l)$ becomes the outflow from the second reservoir and inflow to the third and so on (Chow et al. (1988)).

The last equation (3.15) mathematically equals the Gamma distribution, also known as the waiting time distribution X $\in$ G(n,k), where n is the shape parameter and k the scale parameter. For $n = 1$ (only one reservoir) we obtain the exponential distribution.

The statistical explanation is that as a model of the hydrograph, the Gamma distribution models the time it takes for a unit of water to arrive at the outlet. Hydrologists calls it the retention time distribution.

The first moment $M_1 = k\,n$ of the Gamma distribution represents the time-lag of the centroid of the area under the instant unit hydrographs. Thus, the use of linear reservoirs represents the storage effect of a hydrologic system, resulting in a time-shift between the centroid of the inflow and that of the outflow, as given by: $M_{1,Q_{in}} - M_{1,Q_{out}} = n\,k$ where M_1 are first moments.

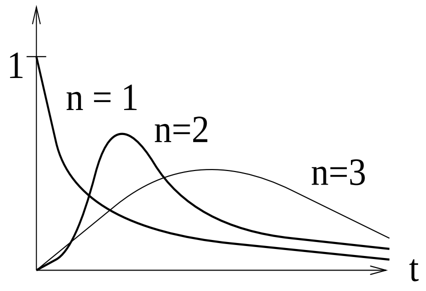


Figure 3.9: The Gamma distribution for different values of n and $k = 1$.

3.3 The oxygen dynamics

Many of the dominant mechanisms which determine the variations of the oxygen level in small streams and rivers are quite well known and described deterministically (see the historical overview, Section 1.3. The results from these important works must be included in a modelling effort to successfully assess the water quality expressed through the oxygen concentration, measured as dissolved oxygen. Oxygen dynamics involves transient components from the natural diurnal and annual oxygen oscillation, caused by photosynthesis and respiration from plants and the impact during rain.

All the included papers (with the exception of Paper **F**) describes the oxygen dynamics in more or less detail. However, Paper **A** describe the physical and chemical factors in more detail than the others, though Paper **C** describe photosynthesis as a function of solar radiation in more depth, while Paper **B** is more concerned with the transient impact from rain. Temperature dependence is treated in Papers **B**, **C**, **D**. In Paper **G**, the oxygen dynamics are modelled by a linear reservoir model formulation.

3.3.1 Dissolved oxygen

The level of oxygen determines the kind of aquatic flora and fauna that can exist, and thereby the trophic level of the ecosystem. The maximum oxygen saturation in fresh water is about 10 mg/l at 10^o C and fish need between 4 and 6 mg/l, depending on the evolutionary stage of the fish. The most demanding fish habitat with respect to oxygen is salmon spawning grounds where the oxygen concentration should not get below 8 mg/l. With the minimum oxygen concentration at night 2 to 4 mg/l less than the daily maximum, this leaves little room for oxygen demanding processes, such as degradation of organic matter.

The processes in the oxygen dynamics can be divided into physical, biological and chemical factors.

Physical factors

- Reaeration
- Temperature
- Precipitation

Reaeration is when an absorption process of oxygen from the air will strive to maintain equilibrium with the concentration of dissolved oxygen in the water. The absorption is assumed to be a first order process (Thackston and Krenkel (1969), Simonsen (1974)):

$$K_2 = \frac{K_L}{h} = const\ \frac{V_F}{h} = const\frac{\sqrt{g\ R_w\ I_o}}{h} \qquad [d^{-1}] \qquad (3.16)$$

$$K_L = 2.9 \cdot 10^{-4}(1+\sqrt{\boldsymbol{Fr}})\sqrt{gI_o} \qquad [m/d] \qquad (3.17)$$

$$\sqrt{\boldsymbol{Fr}} = \frac{v}{\sqrt{g\ h}} \qquad (3.18)$$

K_2 is the reaeration constant, K_L is the adsorption constant and h is the water level. V_F is the sheer stress velocity and the gravitational constant, g = 9.82 m/s^2. I_o is the gradient of the riverbed and R_w is the hydraulic radius, equal to the cross sectional area divided by the wetted perimeter. $\boldsymbol{Fr}$ is Froudes number, and v is stream velocity. Froudes number is also dependent on the water depth, but is considered constant to keep it simple. This relationship was used in the models reported in Paper **A**. Dobbins (1964) stated that Froudes number was less than 0.5 for most natural streams.

It is seen from Eq.s (3.16) and (3.18) that stream velocity and slope of the stream bottom, i.e. hydrology, is a considerable factor. Hornberger and

Kelly (1975) report an example of estimating the reaeration using a productivity analysis. Empirical relationships for reaeration as functions of flow and depth is found in Boyle and Scott (1984), Winter (1988), Wittenberg (1992) for one English and several German rivers. The German findings, however, were only valid for rivers between 25 and 70 m wide, i.e. much larger rivers than we have in Denmark.

Temperature in a small river or stream varies seasonally over the year, as well as diurnally and influences the saturated oxygen concentration and the constants in the reaeration, respiration and photosynthetic processes. Typically the temperature dependencies are modelled by exponential functions of the Arrhenius type based on 20 ^{o}C:

$$K(T) = K(20)\theta^{T_1 - 20} \tag{3.19}$$

where the water temperature, T, is measured in degrees Celsius and θ is the correction factor. This means that a 10 degree raise in temperature doubles the process rate. Temperature corrections for various parameters are listed in Table 3.1.

Constant	θ	References
K_2(T)	1.015–1.047	Dobbins (1964)
R(T)	1.0713	Simonsen (1974)
P(I,T)	1.035	Thyssen (1980)

Table 3.1: Temperature correction values for the reaeration constant, K_2, the respiration, R, and the photosynthesis, P.

Streeter and Phelps (1925) found a θ-value of 1.016 for the reaeration for the Ohio river and Thyssen (1980) found θ-values for the reaeration constant between 1.015 and 1.0241 for a small creek in Denmark, while Spildevandskommiteen (1985) suggests 1.025 in Denmark. Wright and McDonnell (1986) estimated $\theta = 1.08$ for respiration and θ-values for photosynthesis between 1.21–1.26 for an American creek. Erlandsen and Thyssen (1983) estimated the respiration coefficient in a study of Gryde river, in Denmark. Data from this river is investigated in Papers **C, D** where temperature corrections were used. In Paper **B**, with a different dataset, however, no improvements were found and the corrections therefore not used. Temperature dependency may be negligible if data are from a period where stable conditions prevail (especially in the summer) and the water temperature remains fairly constant (Beck and Young (1975a)).

The saturated oxygen concentration, C_m, is dependent on temperature and the pressure from the atmosphere (see e.g. Thyssen (1980) for an overview and Spildevandskommiteen (1985)). It is usually approximated to a 2. or 3. order polynomial as a function of temperature as listed in Table 3.2.

Equation: $C_m =$	References
Fitted tabular values:	Hunter (1978)
$14.5294 - 0.3735771\ T + 0.004973443\ T^2$	
Fitted equilibrium data:	Beck and Young (1975b)
$14.54 - 0.39\ T + 0.01\ T^2$	

Table 3.2: Saturated oxygen concentration equations.

The tabulated values were used in Paper **A**, while the simple function used by Beck and Young (1975b) was utilized in Papers **B, G**. The data used in Papers **C, D** included saturated oxygen. Temperature will naturally be highly correlated with the solar radiation.

Precipitation Precipitation covers rain, hail and snow, plus the mixes between any of these. However, in all applications in this thesis only rain is accounted for. This is partly due to the fact, that the lowest oxygen concentrations occur in summer, thus all the data used are measured in the summertime. Rain data is collected from tipping bucket collectors from the Danish Meteorological Institutes network of measurements. Each tip contains 0.2 mm of rain and the unit of the registered tips is intensity $[\mu\, m/s]$. A thorough explanation of the collecting system and the measuring methods may be found in Bramslev (1989) and in Jacobsen and Voss (1994b). A rain event is defined as raining period where the time between tips is less or equal to one hour. Within a rain event the tips are also stored cumulated.

Precipitation onto and evaporation from the surface of the stream itself will be neglected due to the very small surface. Precipitation run-off though, may be of significance, thus accounting for some direct dependency and resulting in a delayed oxygen consumption. Rainfall is a powerful transient burden of the entire system.

Chemical and biological factors

- Photosynthesis
- Respiration
- Degradation of organic matter
- Biomass

Photosynthesis is a process where plants convert light energy, in the form of measured solar radiation, into chemical energy. Very simplified, this can be written:

$$CO_2 + H_2O \overset{Light\ \ energy}{\rightarrow} CH_2O + O_2 \tag{3.20}$$

Carbon dioxide, CO_2, and water produces organic matter represented by CH_2O, and oxygen, O_2. Not all the wavelengths of the measured global solar radiation can be utilized for photosynthesis. The photosynthetic active spectral area is very close to that of visible light: 400-700 nm. Wright and McDonnell (1986) reports the fraction of total radiation active in phosynthesis as 46 % of the total solar radiation, while Hornberger et al. (1976) reported values for the photosynthetic efficiency between 0.16–0.60. Some biologists claim that only the green colour corresponding to wavelength λ = 665 nm. is used. The spectral distribution changes with the depth of the water, as well as the strength of the sunlight. For photosynthetic production of oxygen in small rivers and streams, as is common in Denmark, it is reasonable not to account for this fact.

Photosynthesis has been the subject of considerable interests and have been studied by Odum (1956) who stated that "The highest primary production rates are in the recovery zones of streams polluted with organic wastes". Camp (1965) introduced the oxygen-sag curve as the consequence of discharge of organic matter into the river. He found that "The rate of supply by photosynthesis is considerably higher than the rate by atmospheric reaeration".

More research was later done by Sand-Jensen (1983), Thyssen et al. (1983), Erlandsen and Thyssen (1983), Jeppesen and Thyssen (1984), Cosby (1984), Cosby et al. (1984), Cosby and Hornberger (1984), Erlandsen and Sørensen (1984), El-Shaarawi et al. (1988), Sand-Jensen (1989)). Production from photosynthetic activity takes place both in the water phase via phytoplankton (algae floating free in the water) and on the riverbed via macrophytes (large growing plants). The process is usually considered dependent on solar radiation and plant mass only, though the the rate and magnitude is very plant specific (Sand-Jensen (1983)). There are some speculations that the rate slows down in the afternoon due to changes in the CO_2 levels (Sand-Jensen (1989), Sand-Jensen (1995)). Recently, (Portielje et al., 1996) have shown that the cumulative primary production (photosynthetic production) is positively related to the level of external nutrient input for both benthic algae and macrophytes.

Often light is assumed to be the only limiting factor (especially in shallow streams), and for simplicity, the production of oxygen is assumed proportional to the light intensity: $P(I) = \beta\, I$. This approximation was used in Papers **A, B, G** where it worked satisfactorily, probably because of the low waterlevel. If the stream is overshadowed by trees or other vegetation, the photosynthetic production will usually be linear. In Paper **C**, for a different river, it is shown that this is not a very good approximation. Eight different functions for solar radiation were tried out and it was found that a Michaelis-Menten type relationship gave a much better description.

$$\begin{aligned} P(I) &= \frac{P_m}{P_m/E_o I + 1} \\ &= P_m \frac{I}{P_m/E_o + I} \end{aligned} \qquad (3.21)$$

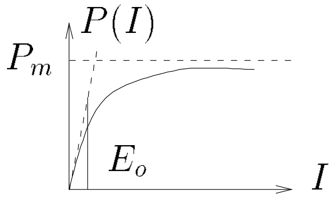


Figure 3.10: Photosynthesis described as a Michaelis-Menten type reaction.

where P_m is the maximum photosynthetic rate, E_o is the slope of the initial photo-synthetic curve as the light intensity approaches zero and the ratio E_o/P_m is the light intensity at which the line tangent to the initial slope intersects the line tangent to the maximum photosynthetic rate. In the Michaelis-Menten equation, this ratio corresponds to the 'substrate concentration' where the rate of growth is exactly half of the maximum.

Respiration from flora and fauna in rivers is a process which generate the energy stored by the photosynthetic process (Eq. (3.20)):

$$O_2 + CH_2O \rightarrow CO_2 + H_2O + Energy \qquad (3.22)$$

The organic matter uses oxygen while carbon dioxide and energy is produced. The process is usually considered constant (see e.g. Hornberger and Kelly (1972)), though there are some speculations that it is dependent on the CO_2 (corresponding to the photosynthesis) and the oxygen content in the water (Owens and Marris (1964), McDonnell and Weeter (1972), Sand-Jensen (1983), Sand-Jensen (1989)). Odum (1956) argued that the respiration possibly was dependent on the oxygen tension in polluted waters, or repressed during photosynthetic action, and suggested that the respiratory metabolism of a polluted stream is proportional to the content of organic matter. On the other hand, the research in Thyssen et al. (1990) suggests that respiration is either linearly or exponentially dependent of oxygen. (Portielje et al., 1996) argues that photo respiration leads to higher oxygen consumption rates at high light intensities.

3.3.2 Organic matter

Streeter and Phelps (1925) founded the theory behind this source of pollution in rivers. Degradable organic matter is not in itself harmful for aquatic life, as it serves as an energy source for heterotrophic organisms, but mineralization removes oxygen from the water. Organic matter is very difficult to measure on-line as it is usually expressed as the biological oxygen demand over 5 days: BOD_5. Camp (1965) pointed out that "Water temperature vary so much during the day, that it is pointless to determine BOD in the laboratory at constant temperature" and attempted to construct equipment for continuous recording of BOD. The exchange between oxygen and organic matter takes place in the water phase as soluble organic matter and on the riverbed as particular matter.

The classical theory regarding decomposition of organic matter is based on a stationary contribution and the assumption that the processing of organic matter can be described kinetically as a first order process:

$$\frac{dL}{dt} = -K_l L \tag{3.23}$$

where K_l is the degradation constant.

However, during the transient impact from rain the assumption of a stationary contribution is violated. Because of the variations in loads, flow, temperature and in imperfect mixing, Camp (1965) felt that the assumption of steady-state will introduce greater errors than will the determination of rate constants.

The oxygen demand associated with rain takes place as both an immediate and a delayed process. This follows as the organic matter is extracted into an absorbed fraction that is degraded directly by the micro organisms and a fraction that is adsorbed to the sediment. In the sediment, the organic matter is bound and not decomposed until later.

If this delayed degradation of the organic matter may be described as the oxygen diffusion in sediment, Harremoës et al. (1989) has shown that this is a half order process, thus resulting in a term with the square root of the oxygen concentration. This was tried out in the models reported in Papers **A, B, G** but with little success.

Biomass is a term that covers bacteria, animals and plants. Algae and plant growth in small rivers and streams increases the roughness and the hydraulic radius, thereby decreasing the velocity. In Thyssen et al. (1990) extensive biological measurements have been conducted and assessed on small streams in Denmark, but some of the parameters estimated in the various models turned out to be very specific for specific streams. After the plants have reached their peak of growth, they die and are degraded by the use of oxygen, which can be debilitating for other aquatic life. Wright and McDonnell (1986) investigated a macrophyte biomass model as a function of the phosforus loading on an American river and also Winter (1988), Wittenberg (1992) have done extensive research of biomass models. However, the latter two have concentrated on German rivers, which are often so large that the processes in them are characteristic to that of lakes (e.g. with separate layers and little movement). In the rivers modelled in this thesis, models for biomass have not been considered to be of relevance, because of modest plant growth.

Boyle and Scott (1984) have shown that benthic algae and biofilm contribute significantly to oxygen depletion in slate or concrete-line rivers such as the river modelled in Papers **A, B**. And recently, Portielje et al. (1996) have shown that a layer of Benthic algae is the main location of primary production. Thyssen (1980) stresses that the role of benthic algae is of much more importance in smaller rivers and streams, where also the microfilm may have an impact.

3.3.3 Summary of the oxygen dynamics

If a river is unpolluted, dissolved oxygen levels will be near saturation and it will be characterised by a diverse community of plants and animals. The ecosystem can be viewed as a balanced cycle. With the photosynthetic process, solar energy is stored as chemical energy in the organic material, where oxygen is produced and carbon dioxide consumed. The reverse process of respiration and decomposition depletes the oxygen and releases carbon dioxide. Higher temperature causes higher total respiration and low reaeration due to high oxygen saturation. As respiration takes place during the entire day, and photosynthesis only in the presence of light, the oxygen concentration will be lowest just before sunrise. Conversely, during the afternoon, super saturated conditions may occur as plant photosynthesis is at its peak. However, high light intensities may result in higher oxygen consumption rates. Respiration is often assumed constant, but may vary with respect to the type of plantlife and the levels of CO_2 and oxygen. As a consequence of the above mentioned processes, large diurnal oxygen variations may occur.

Overflows from the sewage system will result in high levels of both dissolved and particular organic matter. This has the impact of higher oxygen consumption from degradation, both immediately and delayed, and may cause unpleasant odour or turbid water. However, low oxygen levels and high levels of organic matter are not always correlated. Very large oscillations may occur far downstream, from a pollution source, due to super-saturated daytime conditions and anaerobic conditions at night.

3.4 Water quality models

In section 3.1.1, one model classification was that of deterministic and stochastic models as two opposites in a spectrum. However, it has been argued in the grey box model discussion in Section 2.8 that a better, or more realistic model is a combination of prior knowledge coupled with stochastic terms. By incorporating physical knowledge into the model, the number of parameters may be restricted, which may lead to better parameter estimates. Most included Papers **A, B, C, D, F, G** are related to grey box models, while the only statistical element in Paper **E** lies in the method used to calibrate an otherwise completely deterministic model.

A mathematical model for water quality contains the following elements:

- **States** or variables, such as oxygen, organic matter, algae.
- **Forced** functions: Exogenous variables such as water flow, temperature, various pollutants and solar radiation.
- **Mathematical** functions: Equations for degradation, exchange between water and sediment, photo synthesis and saturated oxygen concentration.

The state variables and various functions are described in section 3.3, as well as in Paper **A**.

Whitehead and Young (1975), Whitehead et al. (1981) put forth the following properties that a good model for oxygen and organic matter should possess:

- It should be a truly dynamic model, capable of operating with a time-varying system.
- It should be as simple as possible, while still retaining the dynamics of the system behaviour.
- It should include reasonable mathematical approximations of the physico-chemical properties of the system.
- It should account for the inevitable errors and be verified against real data.

To describe un-steady, distributed flow in a river, requires the one-dimensional Saint Venant equations developed by Barre de Saint Venant in 1871. These are the conservation of mass and the momentum equation derived in e.g. Chow et al. (1988).

$$\frac{\partial Q}{\partial x} + \frac{\partial A}{\partial t} = q \tag{3.24}$$

Only the conservation of mass equation is shown here, as the momentum equation will not be dealt with due to difficulties with stochastic partial differential equations. If the average cross sectional velocity is constant down the river we obtain the continuity equation:

$$v\frac{\partial A}{\partial x} + \frac{\partial A}{\partial t} = q \tag{3.25}$$

The general mathematical model for transport, distribution and exchange of a non-conservative substance, such as oxygen, includes

- Transportation, local and convective.
- Distribution with respect to time and space.
- Exchanges, such as production and consumption.

3.4.1 Transport processes

The transport of oxygen due to motion of the water is called convection, while the transport due to the concentration gradient is called diffusion. Transport with both convection and diffusion is known as dispersion.

Convection

The total change in the state of a particle (substance concentration, temperature, etc., during flow, equals the sum of contributions from the local and the convective state-change. I.e. for a non-conservative substance, such as oxygen, C (Bjørnø and Hartig (1984)):

$$\frac{dC(x,t)}{dt} = \underbrace{\frac{\partial C(x,t)}{\partial t}}_{local} + \underbrace{v\frac{\partial C(x,t)}{\partial x}}_{convective} \tag{3.26}$$

Hence, the total differential contains changes as a function of both time and space. The local changes, i.e. the partial derivative with respect to time, is comprised of diurnal and seasonal oscillations over the year, as well as decomposition as time passes, but in the same spot. The convective changes is comprised of the decomposition that takes place as a function of distance from e.g. an overflow. In a Eulerian observation scheme, only information from one specific point is obtained as time passes. I.e. the convective term equals zero and it is the partial differential with respect to time that is actually observed.

Diffusion

Simple diffusion is characterized by Fick's law: A substance moves from regions of higher concentration to regions of lower concentration and the rate of motion is proportional to the concentration gradient:

$$\Phi = k\nabla C \qquad [mg/(m^2\,s)] \tag{3.27}$$

where Φ is flux per unit time, k is a proportionality constant $[m^2/s]$ and ∇ the concentration gradient. Fick's law is based on molecular transport, and states that a substance tends to equalize its distribution. However, molecular diffusion often appears to be negligible.

Dispersion

Consider dissolved oxygen in the water at time $t = 0$ and place x_0. If the substance both follows the flow and is spread with diffusion, then:

$$\frac{d}{dt}C(x,t) = k\nabla C(x,t) \tag{3.28}$$

$$\Updownarrow$$

$$\frac{\partial}{\partial t}C(x,t) + v \cdot \frac{\partial}{\partial x}C(x,t) = k\nabla C(x,t) \tag{3.29}$$

since we have assumed that the fluid flows at a constant rate v only in the x direction. Thus, the change in concentration of an element of water moving along the direction of the stream must equal the diffusion.

3.4.2 Assumptions and approximations

Solutions to deterministic partial differential equations have been developed with numerical solutions, such as finite difference and finite element methods. However, to the best of my knowledge, such methods have yet to be developed for stochastic partial differential equations, as there are both statistical and numerical problems that must be solved first. Therefore, some approximations are necessary to model stochastic partial differential equations. The models developed here are based on the Eulerian viewpoint, in which only the total state can be considered.

- Only one-dimensional changes along the direction of flow in the rivers are considered. I.e. the volume of the river is assumed completely mixed, transversally and vertically.
- The convection term during sudden impact from overflows cannot be separated from the total variation, but is assumed to be lumped with the local variation.
- The theory for degradation of organic matter is based on the kinetics of first order degradation.

The assumption of lateral mixing is reasonable after a distance of about one to four km from overflow structures (Harremoës and Malmgren-Hansen (1990)). This assumption also renders obsolete the dispersion terms.

In situations with a sudden impact from e.g. an overflow from a large rain event, the convective term becomes important. For a stationary, uniform river, this term is zero and the total differential in Eq. (3.26) equals the partial differential of the local change. Only Lagrangian based observations cover the spatial variation. Hence, there is not enough information in observations from an Eulerian measuring program to separate convective from local transport.

The difference in the model from the real world is covered by the stochastic terms for model uncertainty. This is why this model type is so versatile and grey box modelling so attractive.

3.4.3 The one station model

Due to the difficulty in setting up a monitoring program and keeping measuring stations in good working order, older measuring programs were often based on data from only one measuring station. This is the case with the data sets in Papers **A, B, C, D**. In Papers **A, B** the data is from a large measuring program with several measuring stations, though only data from one was used. The reason for this is that the river (Harrestrup Å) runs through a highly populated urban area and the electronic set-up to collect the data was new and advanced. This resulted in a lot of data fall-outs of either one or several of the variables. Also, this river is sometimes nearly dry in prolonged dry weather periods. Therefore this body of receiving water was modelled with data from only one measuring station.

The model for the urban river was developed to take account of delayed degradation of sedimented organic matter, which is attempted by the description for oxygen diffusion in the sediment. This is described by an oxygen consumption constant multiplied by the square root to the oxygen concentration. It was hypothesised that this might be connected to rain events and therefore this term was multiplied with the precipitation term. Whether this is actually physically interpretable is not certain. However, it was statistically significant. It was also tried out on the two-station model in Paper **G**, where it was not significant. However, using the square root to the oxygen state was no success. Instead, the square root of the average concentration of oxygen was used. This creek was characteristic of an urban waterway with slabs of concrete lining the creek bottom. It was the bacterial layer on these slabs that it was attempted to model this way. The water depth, h was included, as this varied highly from very low levels during dry spells to the high water levels when masses of combined sewer overflows entered the river. In Papers **A, B** the photosynthesis, was modelled linearly: $P(I) = \beta I$ as suggested by Cosby (1984). Temperature corrections was attempted, but no improvement was found.

The river modelled in Papers **C, D** was from a small Danish creek: Gryde Å, which is a natural country river with no overflows. So here the term of sedimented organic matter was not included. Also the water level was fairly constant in the period of the measurements, so the reaeration parameter was considered constant and no depth dependence incorporated. Temperature corrections were incorporated and the best function for photosynthesis was found to be the Michaelis-Menten type relationship, using prefiltered data (see Eq. (3.21) in section 3.3.1).

3.4.4 The two-station model

With better electronics and the possibility to monitor the data quality via computer, data from two or more measuring stations are increasingly becoming a reality. Earlier data collections from more than one measuring station had the disadvantage of not being sampled to include diurnal variations (see e.g. Young and Beck (1974), Beck and Young (1975a), Whitehead and Young (1975)).

By modelling data from two measuring stations, a description of the dynamics of the stretch between the two stations is obtained. If there are stationary conditions and no influence from sudden impacts, then this model may be generalized. There must be sufficient distance between the two stations, such that dynamics one wishes to model is indeed is observable. However, if the stations are too far apart, the assumption of stationarity is not fulfilled. Hence, there is a delicate balance between dynamics vs. stationarity.

The two-station model for the hydraulics based on a linear reservoir model is described in Paper **F**. With reference to this model, an extended hydraulic model, as well as a model for the oxygen dynamics for that river, were developed and reported in Paper **G**.

In lieu of data from two or more measuring stations, a distributed deterministic model, such as the one forming the basis of Paper **E**, could be used to simulate input further down a river. These simulated values might be used in a stochastic two-station model that could aid in the understanding of the processes over a longer stretch of a river.

3.5 Summary of qualitative concepts

In this chapter model classifications have been described, along with an emphasis on the models used in the work for this thesis. Then, the hydrodynamics and different viewpoints from which to study waterflow was presented. The oxygen dynamics with a description of the processes that govern the oxygen content of the water in a river was introduced and the mechanics of a water quality model was explained. Finally, the approximations necessary to build a water quality model was described along with short introductions to models from one and two measuring stations, respectively. The oxygen dynamics is a very complex array of sub-models for different processes. These will therefore be summed up in the following section.

3.5.1 Summary of sub-models for oxygen-related processes

In the previous sections it has been shown that the complex of oxygen dynamics includes several sub-models for single processes.

Photosynthesis as a function of solar radiation have been investigated in Papers **C, D**. In Paper **C** eight different functions are tried out on data, which is pre-filtered, while in Paper **D** a filter is estimated along with the best function found in Paper **C**. The estimated filter does not yield as impressive results as the use of a low-pass filter prior to estimation. However, the system with the estimated filter has the advantage that it can be used on-line.

Temperature dependency was expected for the functions of reaeration, respiration and photosynthesis. This was investigated in Papers **B, C, D**, but used only in Papers **C, D** as no improvement was found in Paper **B**. The distinction here is the use of different datasets. More work could be done with data from other rivers, as well as other datasets from the same rivers.

Oxygen consumption from the riverbed was tried out in Papers **A, B, G** with little success. Here, much more work is warranted and specific experiments should be undertaken to obtain more knowledge of these processes in natural rivers and streams.

Chapter 4

Cases and the data used

Three different rivers were used in the work reported in this thesis.

1. Harrestrup Å in the greater Copenhagen area, on the island of Zealand.
2. Gryde Å in a rural area on mainland Jutland.
3. Usserød Å in a suburban area near farmland areas, Zealand.

The Harrestrup Å and Usserød Å are used as receiving waters and receive combined sewer overflows (CSO's) when large rain events results in more water than the sewers or waste water treatment plants can handle. Gryde Å was included as this was interesting from a comparison point of view and because it had solar radiation measured right by the river and not several kilometers away, as was the case with the other two. This provided the possibilities to try out different functions for photosynthesis as functions of solar radiation.

Figure 4.1 shows a map of Denmark and the approximate location of the rivers; numbered to correspond with the list above.

4.1 Harrestrup Å

Harrestrup Å is a river, that is heavily influenced by it's urban surroundings. Baseflow is very small, the riverbed is lined with concrete slabs and during large rain events, torrents of water are released into the river. The

Figure 4.1: Maps of Denmark with Jutland (J) on the left and the island of Zealand enlarged on the right. The numbers correspond to the list with the rivers on the preceding page.

surrounding community hopes to eventually be able to restore this river for use as a recreational area.

A set of data from one measuring station have been used in Papers **A, B** to model oxygen and organic matter. The model for oxygen was designed to take into account the delayed degradation of sedimented organic matter via a description for oxygen diffusion in sediment. A part of the data used in this investigation is depicted as scaled measurements in both papers.

4.1.1 Summary of results for Harrestrup

Paper **A** is concentrated on explaining the physical and chemical factors and on presenting the estimation method in detail. On the other hand, Paper **B** is concentrated on the model for the photosynthesis and the temperature corrections of some of the parameters. The numeric results were the same, as the ultimate model proved to be identical in the two investigations. However, since the parameter for respiration were not significantly different from zero in a t-test in Paper **B**, it was tried to set this parameter equal to zero. This provided much more accurate parameter estimates for the remaining parameters.

The overall results for Harrestrup Å were:

- A linear relationship for photosynthesis as a function of solar radiation was adequate.
- Temperature corrections for the parameters were not necessary.
- Respiration was not identified.

A comparison between the measured and the simulated oxygen are shown in Paper **A**, while residuals and the results from the residual analysis is shown in Paper **B**.

Estimation in a model with observations from two measuring stations have been attempted several times. The last investigation used a linear reservoir set-up, that worked very well for the third river (Usserød), but it was not able to capture the dynamics or model even the hydraulics satisfactorily. Much more work is needed for this river, as it is quite unique. It is speculated that the inclusion of reservoirs, to model the catchment further upstream from the measuring stations, might yield some results.

4.2 Gryde Å

Gryde Å is a middle-sized, meandering river, located in rural surroundings on the Danish Jutland peninsula. There is no organic matter from waste water effluent, except maybe some spillage from local farms, though there may be some nutrients from the surrounding fields. It is dominated by large macrophytes.

Five measuring stations were set up meticulously (Thomsen and Thyssen (1979)) in the late 1970's and other data sets from this river have since been the subject of many studies in the eighties (Thyssen (1980), Thyssen et al. (1983), Erlandsen and Thyssen (1983), Jeppesen and Thyssen (1984), Cosby (1984), Cosby et al. (1984), Cosby and Hornberger (1984), Erlandsen and Sørensen (1984)) and then revisited in the nineties (Thyssen et al. (1990), Pedersen (1994)). All studies, presented in the above mentioned papers, are modelled on data from one station and only one state is used. This, despite the fact that preliminary investigations show that two variables should be included. It is not clear, however, what the second state should be. There are no sewer overflows into this river, but diffuse runoff from the agricultural surroundings may contain nitrogen.

The data used is shown in Paper **C** and the low-pass filter for the data is described. Eight different models for photosynthesis as a function of solar radiation was tried out using both the unfiltered and the pre-filtered data. Model equations for the oxygen dynamics are identical, except for the photosynthesis function, where the one found to give the best description in Paper **C** was used in Paper **D**. In Paper **D** attention is directed at the estimated filter.

4.2.1 Summary of results for Gryde Å

In this dissertation, data from one of the measuring stations have been used in Paper **C** to investigate eight different descriptions for photo synthesis as a function of solar radiation, $P(I)$, and it was shown that pre-filtering the data led to more accurate parameter estimates. The best function found in Paper **C** was used in Paper **D** to estimate a filter as well as the parameters of $P(I)$. The estimation of a filter did not yield as impressive results with respect to more accurate parameter estimates as using a nearly ideal low-pass filter to pre-filter the data. However, when considering that this enables on-line estimation, which might be of use for monitoring purposes, the results were satisfactory.

The overall results for Gryde Å were:

- A linear relationship for photosynthesis as a function of solar radiation was not adequate. A Michaelis-Menten type relationship was found to give the best result along with a logarithmic function.
- Temperature corrections for the parameters were used, as this gave a significant improvement.
- A low pass filtering of the data, including the solar radiation, leads to more accurate parameter estimates.

4.3 Usserød Å

Usserød Å is a river located about 25 km north of the city of Copenhagen. It starts at Sjælsø (a lake) and ends at another river, Nive Å, which runs into the sound between Denmark and Sweden. It is surrounded by suburban communities and farmland.

Data is collected in an extensive monitoring program as a basis for an environmental assessment of the river. There is a waste water treatment plant about midway, which uses the river as receiving water for its treated water. Besides this, there are several CSO's that discharge untreated water when large rain events taxes the capacity of the system.

The data used here was received from two measurement stations, 2191 meters apart:

- Station 302 Kokkedal – The upstream station.
- Station 301 Nive mølle – The downstream station.

The period covers a full 9 days: Aug. 21. '96 – Aug. 29. '96. Roughly 500 meters downstream from station 301 is a water mill protected from large debris flowing in the river by a gate. This may become clogged and result in the water being stowed upstream.

The following data were used:

- Water level, h $[cm]$ from stations 301 and 302 in the creek. Measured by a pressure transducer.
- Temperature ,T $[°C]$ from stations 301 and 302 in the creek. Measured by a standardized electronic censor.
- Dissolved oxygen, C $[mg/l]$, from stations 301 and 302 in the creek. Measured by a standardized electronic censor.
- Precipitation, P_r $[mm/min]$ from about 2.5 km upstream from station 302, near the waste water treatment plant. Measured by a standard tipping bucket rain gauge.
- Solar radiation, I $[W/m^2\ min]$ from Højbakke gård meteorological station about 15 km away. Measured by a solarimeter.

Besides these time series of data, the profiles of the creek were measured very close to the measuring stations. From the profiles it was possible to find the average river surface area and average wetted perimeter as a function of the waterlevel. The latter was used in the hydraulic relations for the reaeration coefficient (see Chapter 3, Section 3.3.1).

An hydraulic two-station model based on a linear reservoir model is described in Paper **F**. This first model is very simple showing only the main idea. Furthermore it is shown that possibly a constant term to account for additional diffuse water supply should be added to the model. The suggested extension of that model was carried out and the model further extended to include the transport of dissolved oxygen and organic matter. The results from these modelling efforts are reported in Paper **G**.

4.3.1 Summary of results for Usserød Å

A hydraulic model was found in two steps. First a very simple linear reservoir model was tried out. When that did not prove adequate, the model was extended with very good results. Based on some subjective measures and on the results for one, two, three and four reservoirs, it was decided that a model based on two reservoirs was adequate. The model gave an allmost perfect fit as seen in Figure 4.2

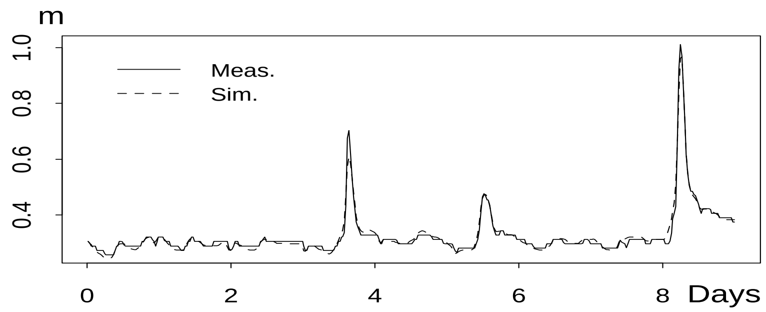


Figure 4.2: Measured and simulated oxygen at Station 301 in Usserød Å for 9 days in 15 minute intervals.

After this, dissolved oxygen and organic matter was included in the model. It was tried out whether more than two reservoirs should be used when enlarging the model. However, this gave no improvement, so it was decided to continue with two reservoirs.

Five different models for the oxygen dynamics was tried out. As the residuals did not seem to be independent, the models should be extended to adequately describe the dynamics. From the hydraulic model it seems that the methodology, using a linear reservoir model, is very promising. It has been used previously (see Whitehead et al. (1979), Whitehead et al. (1981)), but only as a discrete time model. Here, the model is in continuous time enabling a direct physical interpretation of the estimated parameters. Also, less parameters are needed compared to the discrete time model. Further work will have to be done to determine how to extend the model to obtain as impressive results, as for the pure hydraulic model. Also data should be checked more thoroughly and preferably against other sets of data from Usserød Å. If the gate, protecting the mill, has been collecting debris to an extent of corrupting the data, this may influence the results.

Photosynthesis was best described with a linear function, as using the non-linear model the parameters did not test significantly different from zero in a t-test. This, however, resulted in the degradation constant K_l not being accepted as significantly different from zero. Also, K_l was correlated with respiration, and possibly the model cannot distinguish whether oxygen is consumed from organic matter with respiration or as a first order degradation, although it seemed that the consumption was best described by respiration.

4.4 Discussion of the cases

The result of the parameter for respiration, having no significant effect in Harrestrup Å, was suspected to relate to the fact that this was a very unnatural concrete lined creek with little actual plant life. Also, the river ran almost dry during fair weather periods, as opposed to rainy periods where the water level were markedly increased. This kind of situation complicates modelling efforts.

Gryde Å is very different from Harrestrup. It is located in rural surroundings with no CSO's and thus no large impact from rain. In fact the water level was nearly constant in the observed period. However, it is not known whether it rained during that period. It is known from earlier investigations that the creek is dominated by large plants (macrophytes). Thus, it is not surprising that the respiration was significant. In Harrestrup Å it was found that a non-linear relationship between photosynthesis and solar radiation did not yield an improvement over a linear relationship. This was not the case for Gryde Å, where eight different functions were tried out. However, the water level for Gryde Å was much larger than the base-flow for Harrestrup Å. The mean water level for the two rivers were 0.73 meter and 0.30 meter, respectively, but the level was almost constant for Gryde Å, while the range for Harrestrup Å was 0.23 – 1.27 meters. Hence, it is possible that the water level is the reason behind the difference in functions for photosynthesis. Wright and McDonnell (1986) used the two-parameter model for photosynthesis on a shallow stream. However, there is no evidence of an attempt to use a linear relationship.

Usserød Å may be characterized as somewhere between the two other rivers. It is affected by transient impact from rain, but it is a more rural type river. During large rain events the waterlevel is considerably increased. This modelling work was based on two measuring stations and a linear reservoir model. After initial investigations, it was shown that two reservoirs gave an satisfactory description. The water quality model was not as impressive as the hydraulic model. More data, with more rain events should be collected and compared with the models suggested.

The best model for the oxygen dynamics was a relatively basic model where respiration was linearly dependent on the organic matter. However, there are some unanswered questions as to the degradation of organic matter which should be the focus of future attention.

All models were validated by the model diagnostic tools available in the program used for the estimation. These tools are described in Chapter 2, Section 2.7. Examples of model validation is seen in Papers **B, G**.

4.4.1 Summary of the results for all cases

It is difficult to compare the results for all three cases across the board. Not only are the rivers very different, but the datasets are also different and finally the model-types are not directly comparable. A lot of research lies ahead in collecting and preparing data-sets from different rivers (preferably more than one set for each) and then trying out the same set of models on each of them. Organic matter was only included in two of the cases. The main points are summarized in Table 4.1.

Model diagnostics for the two rivers, affected by rain and thus organic matter from CSO's, showed that there was some frequency not accounted for in modelling the oxygen dynamics. Diurnal variations are very well known and the processes well understood. The frequency not accounted for here was of a different order. In the Harrestrup Å case, this only showed up when plotting the measured and the simulated data as shown in Figure 4.3. The cumulative periodogram was completely satisfactory.

In the case of Usserød Å, the troublesome frequency was seen in the cumulative periodogram. Because of the short time period available, it was not

River	Characteristic	Main results
Harrestrup Å	Urban, shallow river heavily affected by rain.	Linear $P(I)$ and no respiration. Temperature corrections gave no improvement.
Gryde Å	Rural country river of medium depth with no effect from organic matter or rain.	Investigations for best model of $P(I)$ yielded the two-parameter model. More accurate parameter estimates were obtained by pre-filtering the solar radiation data. Temperature corrections improved the estimations.
Usserød Å	Suburban, shallow river in partly rural surroundings, affected by rain.	Two-station model, based on linear reservoirs. Linear $P(I)$ and respiration dependent on organic matter.

Table 4.1: A summary of the main results for the three rivers. $P(I)$ denotes the photosynthetic function, dependent on solar radiation.

possible to distinguish the frequency in the measured and simulated plot. However, as seen in Figure 4.4, these did show that some dynamics was not accounted for.

It is speculated whether it is the (possibly delayed) degradation of organic matter that is not described adequately. It is well known from waste water treatment plants that there is a difference between water quality during the regular week and over the week-ends. More research should be directed at identifying the frequency where the dynamics is not accounted for. To do this a longer series of data is necessary from two or more measuring stations. Also, it should be investigated for the particular river how long the distance between measuring stations should be to be able to identify the degradation of organic matter, as well as capture the delayed reaction after a heavy rain event.

It should be noted that the two cases of modelling the oxygen and organic matter dynamics are not directly comparable. The data for Harrestrup Å

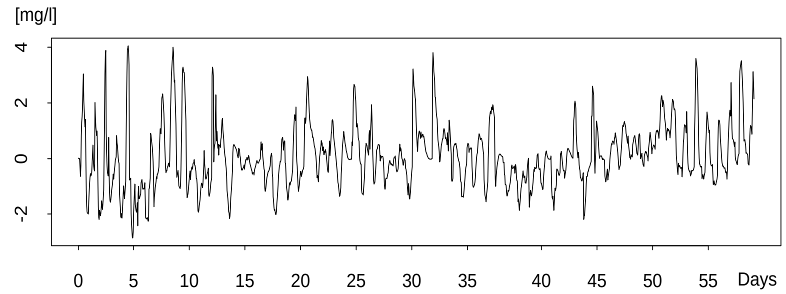


Figure 4.3: Measured minus simulated oxygen in Harrestrup Å for 60 days in hourly time-steps.

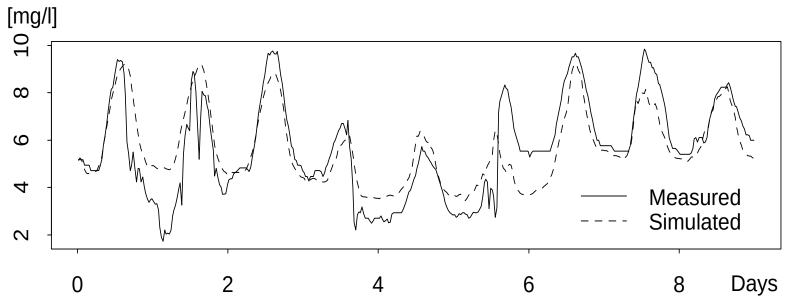


Figure 4.4: Measured and simulated oxygen at Station 301 in Usserød Å for 9 days in 15 minute intervals.

was only available from one measuring station, while there was data from two measuring stations in Usserød Å.

Chapter 5

Discussion and future perspectives

> So eine Arbeit wird eigentlich nie fertig; man muß sie für fertig erklären, wenn mann nach Zeit und Umständen das mögliche getan hat[1].
>
> *Johann Wolfgang Goethe, 1787.*

[1]"A work such as this is never really finished; one must simply declare it finished when one has done what is possible within the limits of time and circumstance"

A successful research project will usually leave the scientist with more questions, yet to be answered, than she started out with. Certainly if time and circumstance allowed, most researchers could go on for quite some time. But as the quote above indicates, one must draw a line, report the obtained results and then suggest future lines of research yet to be done.

In this chapter, a discussion of future perspectives for continuing research in the aspects covered by this thesis is suggested. Being an industrial research project, there are also industrial interests, and this perspective is also addressed.

5.1 Suggestions for future research

There are several directions that future research may take:

- Data treatment and on-line monitoring.
- Further development of the two-station model.
- Refinements of the various processes in the oxygen dynamics.

5.1.1 Data treatment and on-line monitoring

Through this work it has become apparent that it is an art to plan, prepare and carry out a well functioning monitoring program. Besides the instrumentation and the developments of new and better sensors, there are the considerations on how to design the data collecting stations, so that they are not easily disrupted or vandalized. The maintenance must be attended to and manpower devoted to this task. Another important consideration is back-up measures. If the impact from rain is of major interest, it is important that preparations are made for alternative procedures in case of power outages and electronic equipment failure during e.g. thunderstorms.

Traditionally hydrological modelling has employed deterministic methods. Such models may contain large numbers of parameters, which may not all be identifiable. The calibration of these models increases in difficulty as the models become more complex and it should be questioned whether such complexity is really warranted (Beven (1989), Jakeman and Hornberger (1993)). Furthermore, only a fraction of the painstakingly collected data is used in such calibrations.

If there are missing data over short periods, there are methods to fill in the gaps. Furthermore, there are methods that deal with restoration and treatment of corrupted data (Martin and Yohai (1985), Jacobsen and Voss (1994a)). Just a few missing data dispersed through the data set may be dealt with through the continuous time formulation employed in this thesis. Stochastic modelling has the advantage of all the data being used in the model identification effort. Furthermore, the models will usually have a much smaller number of parameters and they are physically interpretable, facilitating communications with biologists and other professionals.

Finally, there is a whole research area devoted to the development of control schemes for water quality (see Kendrick et al. (1970), Young and Beck

(1974)). Design of experiments and assessment of identifiability combined with data storage considerations form another strong scientific line of research.

5.1.2 The two-station model

More data should be obtained from two-station set-ups to probe further into modelling measurements from two or more measuring stations. A lot of research lies ahead in collecting and preparing data sets from different rivers (preferably more than one set from each), and then developing a universal model to try out on each of them. It is important that the data are sampled and collected the same way. Two different venues are possible: one concerned with rivers influenced by rain and overflow from CSO's and one where this influence is not felt, so that the more natural oxygen dynamics in rivers may be investigated further.

The linear reservoir abstraction should be further investigated, as it is expected that this model may have some attractive features. The lumped model includes the spatial distribution without investing in measurements from several closely spaced observation stations. Via the time constant an estimate of the residence time of the river is obtained.

While the linear reservoir model complex set up for Usserød Å, as reported in Paper **G**, showed promise of a successful model, this did not work out for two-station measurements for Harrestrup Å. However, it is felt that the model may just need to be adjusted to account for the different dynamics governing this creek. It is suggested that the upstream area should possibly be included by its own set of linear reservoirs to account for the way the rain is transported before it reaches the river. In Usserød Å, this did not seem necessary, but there are not so many and so heavy overflows upstream as from the point in Harrestrup Å, where measurements from two weeks were obtained.

5.1.3 Further refinements

Despite the amount of work done on investigating the photosynthetic process, there is still much that could be done to further understand the processes of photosynthesis and respiration. Odum (1956) suggested that respiration was dependent on organic matter. This was tried out with a linear relationship, with some success but only on one dataset. On the other hand, Thyssen et al. (1990) suggests that respiration might be dependent of oxygen; either linearly or exponentially. There is certainly some evidence that these processes are plant specific (Sand-Jensen (1983), Gordon and Sand-Jensen (1990)) and thus specific for the kind of river.

The tentative results from this thesis show that the water level of the river might be decisive for which type of photosynthetic function yields the best results. With sufficient data from a number of different rivers, this could be catalogued and maybe generalized. A threshold model may be a possibility that should be tried out. Also, measurements of pH and carbon dioxide (CO_2) coupled with grey box modelling may help the biologists in determining the impact of these variables on the photosynthesis, thus resulting in better models for the oxygen dynamics.

Another important process is the delayed degradation of organic matter. It is possible that further developments of the linear reservoir model may be able to include the spatial variability through the number of reservoirs, where possibly the constant governing this process could be allowed to vary from reservoir to reservoir.

Besides the issues, which have been addressed in this thesis, there is also interest in investigating the nitrogen and ammonia influence on the oxygen dynamics (see e.g. Whitehead (1980), Winter (1988), Wittenberg (1992)). In other countries, the main focus may be on phosphorus and the acidification of rivers and lakes (see e.g. Zetterquist (1991)).

5.2 Future industrial perspectives

Several investigations have been performed that may eventually become future products of the company. In line with the aim of the industrial research project, these are still in the concept stage, rather than finished products. At this point, the results can be used by the company in more research oriented projects. After further developments (e.g. into finished software products with a user-friendly interface), these products might eventually enter the marketplace:

- An automatic calibration routine.
- Stochastic modelling, complemental to deterministic modelling.
- The possibility for design of on-line filtering and monitoring.

The automatic calibration routine will assist in a faster and more efficient use of time when working with these traditional models. The next step in the use of this method is to try it out on real data for a natural catchment. If, or rather when, new and more detailed government regulations stipulates more advanced monitoring programs the work done with the filtering of data for use in on-line monitoring will perhaps be developed further. PH-Consult, through its unique position, might be able to influence such legislation because of the research they have invested in this area.

Stochastic models have the advantage over deterministic models that the parameters estimated carries with them the dimensions of uncertainty of their value based on consistent statistics. This uncertainty gives an indication of which variables it makes sense to concentrate on and which are so uncertain that more research will have to be done on the processes which they govern. Deterministic models, on the other hand, have their strength in their detailed descriptions of the processes involved. Thus, the two types of models supplement each other and PH-Consult's knowledge and experience in both fields may result in a more holistic approach and better service to its customers.

Furthermore, with the experience PH-Consult has gained in working in different ways with real data from large monitoring programs, their service in giving advice can form the base in a well structured monitoring program. There, the following modelling (both deterministic and stochastic) may yield answers that result in a more correct identification of the problems encountered. The often very large investments that follow such monitoring may then be reduced significantly or at least be used where they really mean a difference.

Finally, the stochastic models of the receiving waters could further be developed and expanded to complete integrated models. Integrated models may be very useful in managing the environment and is in fact the new trend in environmental protection. In turn this would provide the company with a competitive edge that could become an international advantage.

Chapter 6

Conclusion

In this thesis several stochastic models have been developed for processes in rivers that are influenced by the transient impact from rain events. It is argued that these models have some attractive advantages over traditional deterministic models. These advantages may be listed as:

- In stochastic modelling, all the collected data is used to identify and estimate the parameters of the model, and missing observations are easily accommodated.
- With the grey box approach, all prior knowledge of the system is built into the model.
- The parameters identified may be interpreted physically, biologically and chemically, thus facilitating communications with biologists and other professionals.
- States that are not measured, such as BOD, may be incorporated and thus estimated.
- The models typically have fewer parameters that deterministic models.
- The estimated parameters are provided along with the uncertainty of these parameters.
- The estimated parameters give a detailed fingerprint of the condition of the river.

The problem of stochastic modelling is the computational effort. However, as computers have become more and more efficient, it is now possible to transfer the methods used and developed here to personal computers. Thus, it will be much easier for the company to include this kind of modelling in the services offered to their customers.

The stochastic models have been tried out on three different sets of data:

- Harrestrup Å, which is a heavily affected river in the outskirts of Copenhagen.
- Gryde Å, a rural country river in Jutland.
- Usserød Å, which is located in more suburban surroundings and may be characterized as a mix between the first two rivers.

The results obtained from modelling these rivers may be summarized as follows:

Pre-filtering data which removes high frequencies, which are not expected to be part of the dynamics, improved the accuracy of the estimated parameters.
Photosynthesis as a function of solar radiation seems to be linear for shallow rivers, while a two-parameter model may yield better estimates for deeper rivers.
Linear reservoir models showed great promise and should be developed further.

The estimated parameters are generally comparable to those found in the literature, when there were values to compare with. The estimated models have provided new information to the dynamics of oxygen and organic matter, expressed as biological oxygen demand.

Besides the developments of models, preliminary research was carried out on automating the process of calibrating a deterministic model using a statistical output error model. The method proved successful on a synthetic case, which was validated by a Monte Carlo method.

Finally, a number of suggestions have been presented to show the many ways this research could be continued and how PH-Consult may benefit from continued developments of the results from the present research effort.

Part II

Included papers

Included articles

Paper A: Jacobsen J.L. and Madsen H. (1996). *Grey Box Modelling of Oxygen Levels in a Small Stream.* EnvironMetrics, **7**, pp. 109–121.

Paper B: Jacobsen J.L., Madsen H., and Harremoës P. (1996a). *Modelling the Transient Impact of Rain Events on the Oxygen Content of a Small Creek.* Water, Science and Technology, **33**(2), pp. 177–187. (In: Uncertainty, Risk and Transient Pollution Events. Selected proceedings of the IAWQ Interdisciplinary Symposium.)

Paper C: Jacobsen J.L., Madsen H., and Harremoës P. (1997a). *Modelling the Dynamical Effect of Solar Radiation on the Oxygen Content of a Small Creek.* In press: EnvironMetrics. (Presented at the 6th International Conference on Environmetrics in Malaysia, Dec. 1995.)

Paper D: Jacobsen J.L., Madsen H., and Harremoës P. (1997b). *Estimation of a Low Pass Filter for Solar Radiation Data.* In press: Proceedings of the 9.th Conference of the European Consortium for Mathematics in Industry ECMI. Institute of Mathematics, Technical University of Denmark, Lyngby, Denmark.

Paper E: Jacobsen J.L., Madsen H., Rauch W., Harremoës P., and Linde-Jensen J.J. (1997). *A Method for Automatic Calibration of Parameters for an Integrated Model.* (Submitted to Water Resources Research.)

Paper F: Jacobsen J.L., Madsen H., and Harremoës P. (1997c). *A Stochastic Model for Two-station Hydraulics Exibiting Transient Impact.* (Accepted to the Watermatex conference, June 1997).

Paper G: Jacobsen J.L. (1997). *A Linear Reservoir Model of the Oxygen Dynamics for a Two-station set-up.* (Manuscript).

Paper A

Grey Box Modelling of Oxygen Levels in a Small Stream

Published in: EnvironMetrics, 7, pp. 109–121 (1996).

Abstract

Data from a stream is used to identify a dynamic model of the oxygen level as a function of solar radiation and precipitation. The time series of the oxygen level is occasionally corrupted by pumping of external water into the stream, which is of no interest to the biological processes in the stream. In this paper a grey box modelling approach, which is a statistical method taking the known physical relations into account, is used. Using this approach, an identification of a stochastic continuous time model for the oxygen level based on the discrete time data, where the corrupted data are considered as missing values, is outlined.

KEY WORDS: Oxygen dynamics; Time series analysis; Missing data; Continuous time modelling; Grey box models.

1 Introduction

The purpose of this work is to establish a stochastic model of the oxygen level in a creek, as a function of solar radiation and precipitation. Oxygen, measured as dissolved oxygen, is the most universal environmental factor when water quality in rivers and streams is evaluated. The level of oxygen determines the kind of aquatic fauna, that may exist, and thereby the trophic level of the ecosystem ((Harremoës and Malmgren-Hansen, 1990)) states that there is a pronounced need for stochastic descriptions of the different processes that influence water quality in order to obtain a better understanding of the involved phenomena.

This paper describes a grey box modelling approach to identify a continuous time linear stochastic dynamical model of the oxygen level in a small stream based on time series with missing values. The missing values are due to an occasional disturbance of the measurements, which is of no interest to the dynamics we wish to model. They are therefore identified, and considered as missing values.

Time series analysis traditionally deals with black-box models, i.e. models which are identified purely on the basis of data. It may be questioned whether such models are only representative of the data themselves, and not of the general dynamics and physics behind the data.

(Bohlin, 1984), Madsen et al. (1985), Madsen and Melgaard (1991) and Melgaard and Madsen (1993b), Melgaard (1994) describes a so-called grey box approach that combines available physical knowledge with statistical modelling tools. This is done by using well-known physical equations, and include them in the analysis, along with a set of data, which are then used to estimate the parameters, and validate or falsify the model structure.

No matter how well understood a system is, there is always uncertainties and phenomena that has not yet been detected or fully realized. Therefore, combining physical knowledge with information from data is an optimal approach towards understanding a physical system.

In Denmark, the Water Pollution Act, WPA, has set standards for stationary waste water discharge. A combined sewage system is designed with a specific hydraulic capacity, which includes an expected rate of overflow. The system has built-in storage basins to take some of the load when the sewage system cannot handle the extra strain during heavy rainfall. Overflows of polluted oxygen free water will occasionally enter the environment, with negative consequences for the recipient quality.

The essence of the problem involves three different components: The drainage system, the sewage treatment plants and the recipient. Each of these components are equally important, but while the first two have been the subject of much research, the recipient has not yet been submitted to such thorough investigation. Rainfall is a powerful transient burden of the entire system, that may give rise to discharge of oxygen consuming components, resulting in lowered oxygen levels and dead fish as some of the repercussions. Thus the WPA-standards eventually may result in huge investments in renewal and restoration of the sewage system, as well as in the monitoring of the recipients.

Many of the dominant mechanisms which determine the variations of the oxygen level in small streams and rivers are quite well known and described deterministically by e.g. Simonsen (1974) and Harremoës et al. (1989), Harremoës and Malmgren-Hansen (1990). However, there are many approximations involved, as the oxygen dynamics involves several transient components: The natural diurnal and annual oxygen oscillation, caused by photo synthesis and respiration from plants and the sudden external influence during rain. This causes so much high frequency variation (noise) that calibration of deterministic models is very difficult, and stochastic modelling has to be considered.

2 The Data and the variables

The dependent variable is the oxygen content of a small creek that run along a recreational area in Copenhagen. The input variables are temperature, solar radiation, precipitation and water depth.

Oxygen $[mg/l]$ and **temperature** $[^0C]$ data were measured online, in the middle of the stream, every 3-5 minutes, and show a marked diurnal variation.

Solar radiation measured as global radiation has a great influence on the oxygen content in a creek and therefore a clear diurnal (and annual) variation is found for the oxygen. The variation in amplitude is, however, according to Harremoës and Malmgren-Hansen (1990) and Madsen et al. (1985) random nature due to the variation of cloud cover. Solar radiation $[W/m^2]$ is measured every 10 minutes at Højbakkegård, a nearby agricultural meteorological station.

Precipitation data, given as an intensity $[\mu\, m/s]$, is collected and stored by The Danish Meteorological Institute. A thorough explanation of the collecting system and the measuring methods may be found in Bramslev (1989) and in Jacobsen and Voss (1994b)

Water depth is measured a bit down stream from the other measurements and are therefore only an approximation to the depth at the measuring station. But since it is an important parameter it has been included.

2.1 Sampling

Since the observations in the different series were not equally spaced, sampling was necessary. For the three variables of interest the hourly measurements of the solar radiation is the longest time interval and the radiation at time, t, is given as the mean over the preceding hour. It was then decided to sample all the data like the solar radiation. See Jacobsen and Voss (1994b) for a more thorough description of the sampling techniques employed.

This sampling technique suppresses a lot of the high frequency variation, giving a more smooth variation. A more correct treatment of filtering and sub sampling, in order to avoid aliasing, were considered. But since the solar radiation series were available only as averaged hourly values, it was decided to treat the other series likewise for consistency. After sampling each of the three time series contained 1460 hourly observations including 411 missing values.

3 A dynamical model for the variation of oxygen

The measurements of the input and output variables for five days are depicted together, scaled for comparison, in Fig. 1. The figure shows that

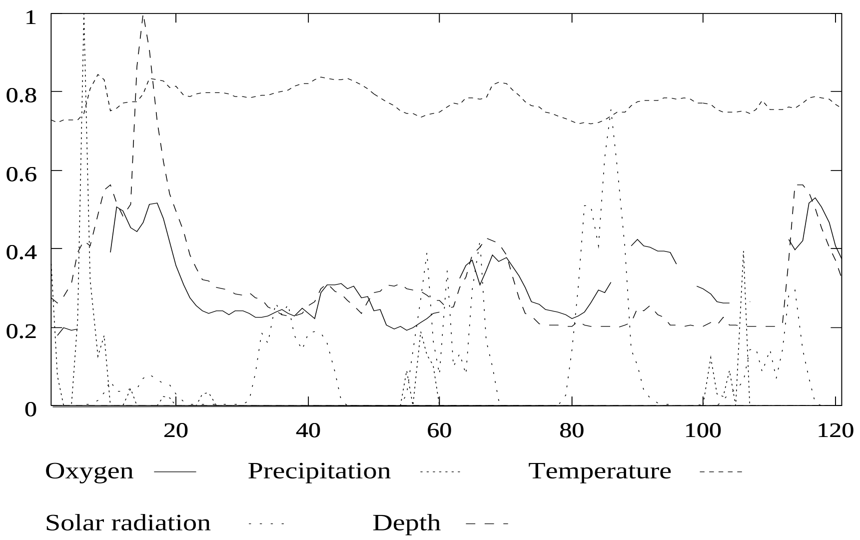


Figure 1: The scaled measurements of oxygen, solar radiation and precipitation. The oxygen measurements show the missing values.

the oxygen level is affected by variations in both solar radiation and precipitation. There seems to be a delayed response in relation to the input, specifically from a shift in the solar radiation. The response to precipitation input is not as distinct as the response to the solar radiation. The depth of the water in the stream is seen to be highly correlated with the precipitation, although a little delayed. Whenever the water depth shows an increase, so does the oxygen level.

A dynamical system, such as a stream, is very complex. All the physical processes are not known and therefore difficult to describe in detail. In this section, only a brief introduction to the physical and chemical processes and the derivations of the governing differential equations will be shown for the oxygen dynamical system. For further details, see Jacobsen and

Voss (1994b).

Auto- and partial auto-correlations revealed that a system of at least two state variables, were to be expected (Jacobsen and Voss, 1994b). When the oxygen level of a small stream is to be evaluated, the concentration of organic matter is known to be the foremost important factor. Therefore it seemed natural to have the concentrations of oxygen and organic matter as the two state variables.

The mass balance for oxygen and organic matter are thoroughly explained and derived in Harremoës and Malmgren-Hansen (1990). The mass balance for a water element is based on the assumption of total mixing and no longitudinal dispersion. The assumption of total mixing is usually fulfilled, except very close to the source of an organic matter discharge.

3.1 The differential equations

In the present complex dynamic system, where both the physical and chemical parameters, behind the oxygen levels in a creek, are interrelated, both non-linear and stochastic elements should be included in the description. A sketch of the processes involved is seen in Fig. 2.

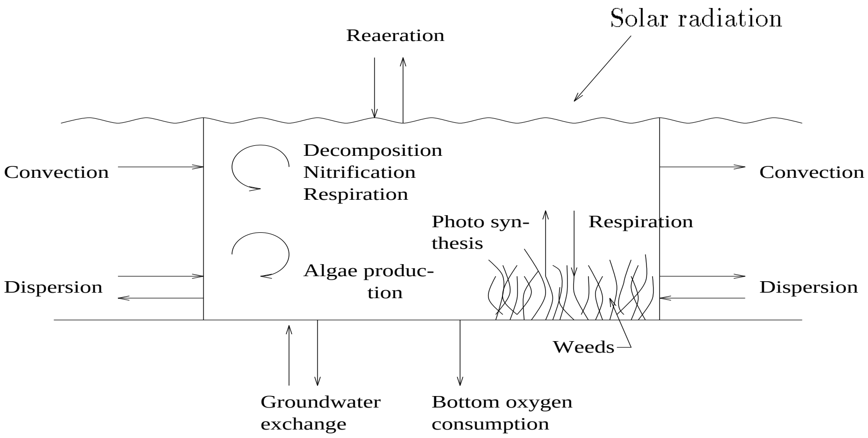


Figure 2: Overview of the processes involved with oxygen concentrations.

The following two differential equations are suggested for the state variables oxygen, C, and organic matter, L. They are highly inspired by Harremoës et al. (1989), Harremoës and Malmgren-Hansen (1990), though the precipitation terms, involving P_r, are these authors own presumptions.

$$\begin{aligned}\frac{dC}{dt} &= -\frac{K}{h\sqrt{h}}C - K_{1c}L + P(I) + \frac{K}{h\sqrt{h}}C_m(T) - R(T) \\ &\quad -\frac{1}{h}\sqrt{C}K_bP_r && (3.1)\\ \frac{dL}{dt} &= -K_{1l}L + K_3C + \kappa P_r && (3.2)\end{aligned}$$

where

$$\begin{aligned}C_m(T) &= 14.5294 - 0.3735771T + 0.004973443T^2 \qquad [mg/l]\\ R(T) &= R_{15}\theta^{T-15} \qquad [mg/l] && (3.3)\end{aligned}$$

$P(I)$ is some function of the solar radiation, I, e.g. $P(I) = \beta I$, Cosby (1984) which is used in the following. This linear function does not incorporate possible saturated or even inhibiting conditions, which are assumed to be of limited importance. The coefficients in the empirical function for $C_m(T)$, the solubility of oxygen in water as a function of water temperature, are estimated via curve fitting to a tabular given in Harremoës et al. (1989). Respiration at temperature 15^o C, R_{15} is estimated along with the other parameters in the model. Otherwise, the function is empirical with $\theta = 1.07$.

The physical explanation for the differential equations may be expressed as: The change in oxygen is depleted by a contribution from re-aeration of the measured oxygen, $\frac{K}{h\sqrt{h}}C$, but replenished from the saturated oxygen concentration, $\frac{K}{h\sqrt{h}}C_m(T)$. Degradation of organic matter from various sources in the water volume also uses oxygen, $K_{1c}L$, while photo synthesis, $P(I)$ from solar radiation produces it. Depletion of oxygen is also caused by the respiration, $R(T)$, and from the sediment at the bottom of the creek, via K_b. The change in organic matter is depleted by degradation, $-K_{1l}L$, but replenished by oxygen involved in nitrification, K_3C, which provides nourishment in the form of nutrient salts.

Photo synthesis as a function of solar radiation has been the subject of much research for lakes, but not much for small streams and rivers. It obviously depends on plants, or rather, the chlorophyll in the plants, which in turn depends on pH and nutrients in the stream.

Physical and chemical factors

The two main differential equations for the state variables, oxygen concentration and organic matter (Eq. (3.1) and (3.2)), involve both physical and chemical factors. These are described in the following.

Solar radiation, I may be influenced by solar elevation and cloud cover, possibly depending on water quality. It is assumed that there is no light saturation or inhibition. According to Harremoës and Malmgren-Hansen (1990) about 44 % of the total solar radiation lies in the photo synthetic active spectral area.

Re-aeration is an oxygen exchange from the air, striving to maintain equilibrium. Absorption of oxygen, in the water, is assumed to be a first order process:

$$\frac{dC}{dt} = \frac{K}{h\sqrt{h}}(C_m(T) - C) \tag{3.4}$$

$$K = 2.9 \cdot 10^{-4}(1 + \sqrt{\boldsymbol{Fr}})\sqrt{gI_o} \tag{3.5}$$

$$\sqrt{\boldsymbol{Fr}} = \frac{v}{\sqrt{gh}} \tag{3.6}$$

where K is a constant comprised of Froudes number, $\boldsymbol{Fr}$, the absorption coefficient, the gradient of the stream, I_o and the gravitational accelleration, g. Eq. eq:fr shows a dependence on the water depth, h but, since it is negligble and to keep it simple, Froudes number is considered constant. $C_m(T)$ is the saturated oxygen concentration at temperature, T, and C is the oxygen concentration. Harremoës et al. (1989), Harremoës and Malmgren-Hansen (1990) gives a detailed account for the re-aeration process. Stream velocity, slope of the stream bottom, i.e. water flow, are important factors. The dependence of temperature for the saturated oxygen concentration, $C_m(T)$, is tabulated in Harremoës et al. (1989) In this project a function is empirically fitted to the tabulated values. Higher temperature causes higher total respiration and low re-aeration due to high oxygen saturation.

Respiration, $R(T)$, from the plants in the creek uses oxygen. It is weakly dependant on the temperature, as described by Eq. (3.3). Respiration of organic matter, represented by CH_2O, is considered constant over the day. That makes it independent of the content of organic matter, plant types, the light intensity and of the oxygen-concentration itself, which may not be entirely likely, but commonly accepted as a first approximation (see

Harremoës and Malmgren-Hansen (1990)). It utilizes oxygen, O_2, produces carbon dioxide, CO_2, and generates energy, which is stored by the opposite process, the photo synthesis.

$$O_2 + CH_2O \overset{Energy}{\longleftrightarrow} CO_2 + H_2O \tag{3.7}$$

Photo synthesis, $P(I)$, is a process where plants convert light into chemical energy. In a model for photo synthetic production of oxygen, light is assumed to be the only limiting factor, and, for simplicity, the production of oxygen is assumed proportional to the light intensity. With photo synthesis depending primarily on the light intensity and respiration being almost constant, the diurnal cycle from solar radiation is clearly recognized in the oxygen variation (see Fig. 1).

Degradable organic matter in excess of what is needed to sustain life, is not in itself harmful for the fish in the stream; but mineralization removes oxygen from the water. The cutting of weeds in the stream has a large, but unknown effect in this context. The classical theory (as presented in Harremoës and Malmgren-Hansen (1990)), regarding decomposition of organic matter, is based on the assumption that the processing of organic matter, in a element of water following the stream, can be described as a first order process:

$$\frac{dL}{dt} = M_L = -K_{1l}L \tag{3.8}$$

where M_L is the production of organic matter. The degradation constant, K_{1l}, is related to the organic matter, L, as indicated by the subscript.

Sedimentation, K_s, and extraction, K_e, is particular organic matter falling to the bottom of the creek by gravity, and dissolved organic matter being extracted from the water phase, respectively. A first order removal is also assumed here; i.e. $dC/dt = (K_s + K_e)L$. These terms, as well as the degradation of organic matter, influencing the oxygen concentration, will be incorporated in one term only: $K_{1c}L$.

Nitrification uses oxygen in the biochemical oxidation of ammonia over nitrite to nitrate. Ammonia is present in waste water, run-off from farm land and decomposition of organic matter in the mineralization process. The stoichiometric coefficient, is about 4.3 $mg\ O_2$ per $mg\ NH_4^+ - N$ ($NH_4^+ - N$ is ammonium bound as nitrogen). Since the organic matter is nourished by these salts, the production of organic matter depends on oxygen through the nitrification process. This term is then included as $dL/dt = K_3 C$.

Precipitation, P_r, onto and evaporation from the surface of the stream itself will be neglected due to the very small surface. Precipitation run-off though, may be of significance, thus accounting for some direct dependency and resulting in a delayed oxygen consumption. If this dependancy is related to oxygen diffusion in the sediment, Harremoës et al. (1989) has shown that it is a half order process, thus resulting in the term with the square root of the oxygen concentration. The unknown contribution from precipitation (with an optimum delay, found to be 4 hours), in relation to the two state variables, will be incorporated with the term P_r.

4 A maximum likelihood method for parameter estimation

In this section it is shown how the parameters in a linear stochastic differential equation are estimated using discrete time measurements and the maximum likelihood method.

4.1 The model in state space form

A model formulation based on physical laws will typically be a set of coupled differential equations. To account for the stochastic fluctuations, as well as the model discrepancies from the true system, stochastic terms are added. Thus the uncertainties and lack of knowledge about the system is included in the stochastic terms.

Including stochastic terms, the differential Eq. ((3.1) and (3.2)) can be written in state space form:

$$
\begin{aligned}
\begin{bmatrix} dC \\ dL \end{bmatrix} &= \begin{bmatrix} -\frac{K}{h\sqrt{h}} & -K_{1c} \\ K_3 & -K_{1l} \end{bmatrix} \begin{bmatrix} C \\ L \end{bmatrix} dt \\
&+ \begin{bmatrix} \beta & \frac{K}{h\sqrt{h}} C_m(T) + R(T) & \frac{1}{h}\sqrt{C}\, K_b \\ 0 & 0 & \kappa \end{bmatrix} \begin{bmatrix} I \\ 1 \\ P_r \end{bmatrix} dt \\
&+ \begin{bmatrix} dw_1 \\ dw_2 \end{bmatrix}
\end{aligned} \tag{4.9}
$$

where the last term is an additive noise term, which is introduced to describe the deviation between Eq. (3.1) and (3.2) and the true variation of the state variables, C and L.

Only the oxygen concentration is measured. The recorded oxygen concentration, C_r, can be written as:

$$C_r(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} C(t) \\ L(t) \end{bmatrix} + e(t) \tag{4.10}$$

where $e(t)$ is the measurement error.

In a short notation the equations are written:

$$d\boldsymbol{X} = \boldsymbol{AX}dt + \boldsymbol{BU}dt + d\boldsymbol{w}(t) \tag{4.11}$$

and

$$\boldsymbol{Y}(t) = \boldsymbol{CX}(t) + \boldsymbol{e}(t) \tag{4.12}$$

where the matrix $\boldsymbol{A}$ characterizes the dynamical behavior of the system and $\boldsymbol{B}$ is a matrix which specify how the input signals enter the system. $\boldsymbol{X}$ is the state vector and $\boldsymbol{U}$ the input vector. Furthermore $\boldsymbol{w}(t)$ is assumed to be a process with independent increments. With the purpose of calculating the likelihood function, $\boldsymbol{w}(t)$ is restricted to be a Wiener-process with the incremental covariance $\boldsymbol{R}_1^c(t)\mathrm{dt}$. $\boldsymbol{e}(t)$ is the measurement error, assumed normally distributed white noise with zero mean and variance $\boldsymbol{R}_2$. Furthermore, it is assumed that $\boldsymbol{w}(t)$ and $\boldsymbol{e}(t)$ are mutually independent.

It is a crucial question whether the parameters of a specified state space model can be identified. If a non-identifiable model is specified, the methods for estimation will not converge. See Madsen et al. (1990) and Melgaard and Madsen (1993b), Melgaard (1994) for a further discussion on identifiability.

4.2 From Continuous to Discrete Time

Since it is assumed that the system is described by the stochastic differential Eq. (3.7) it is possible analytically to perform an integration, which under some assumptions exactly specifies the system evolution between discrete time instants.

The discrete time model corresponding to the continuous time model (3.7) is obtained by integrating the differential equation through the sample interval $[t, t+\tau[$. Thus the sampled version of (3.7) can be written as:

$$\begin{aligned} \boldsymbol{X}(t+\tau) &= e^{\boldsymbol{A}(t+\tau-t)}\boldsymbol{X}(t) + \int_t^{t+\tau} e^{\boldsymbol{A}(t+\tau-s)}\boldsymbol{B}\boldsymbol{U}(s)ds \\ &+ \int_t^{t+\tau} e^{\boldsymbol{A}(t+\tau-s)}d\boldsymbol{w}(s) \end{aligned} \tag{4.13}$$

Under the assumption that the input vector, $\boldsymbol{U}(t)$, is constant in the sample interval, the sampled version of (3.7) can be written as the following discrete time model in state space form:

$$\boldsymbol{X}(t+\tau) = \boldsymbol{\Phi}(\tau)\boldsymbol{X}(t) + \boldsymbol{\Gamma}(\tau)\boldsymbol{U}(t) + \boldsymbol{v}(t;\tau) \tag{4.14}$$

where

$$\boldsymbol{\Phi}(\tau) = e^{\boldsymbol{A}\tau} \tag{4.15}$$

$$\boldsymbol{\Gamma}(\tau) = \int_0^{\tau} e^{\boldsymbol{A}s}\boldsymbol{B}\, ds \tag{4.16}$$

$$\boldsymbol{v}(t;\tau) = \int_t^{t+\tau} e^{\boldsymbol{A}(t+\tau-s)}\, d\boldsymbol{w}(s) \tag{4.17}$$

Eq. (4.14) is the discrete form of Eq. (3.7).

On the assumption that $\boldsymbol{w}(t)$ is a Wiener process, $v(t;\tau)$ becomes normally distributed white noise with zero mean and covariance:

$$\boldsymbol{R}_1(\tau) = E\left[\boldsymbol{v}(t;\tau)\boldsymbol{v}(t;\tau)'\right] = \int_0^{\tau} \boldsymbol{\Phi}(s)\boldsymbol{R}_1^c\boldsymbol{\Phi}(s)'ds \tag{4.18}$$

If the sampling time is constant (equally spaced observations), the stochastic difference equation can be written:

$$\boldsymbol{X}(t+1) = \boldsymbol{\Phi}\boldsymbol{X}(t) + \boldsymbol{\Gamma}\boldsymbol{U}(t) + \boldsymbol{v}(t) \tag{4.19}$$

where the time scale now is transformed such that the sampling time becomes equal to one time unit.

The calculations above were carried out, assuming that $\boldsymbol{U}(t)$ was constant through the sampling interval $[t, t+\tau[$. In the following we shall, however, assume that $\boldsymbol{U}(t)$ is linear through the sampling interval. It is advantageous for cases where the measurements have been low pass filtered, e.g. by simple averaging, or generally when the input is not controlled.

The implementation of the linear interpolation is a modification to the derivations above. The modification is in the calculation of the term: $\mathbf{\Gamma}(\tau)\boldsymbol{U}(t)$ in Eq. (4.14).

Assuming that $\boldsymbol{U}(s)$ is linear through the points $(t, \boldsymbol{U}(t))$ and $(t+\tau, \boldsymbol{U}(t+\tau))$, results in the calculation:

$$\begin{aligned} \int_t^{t+\tau} e^{\boldsymbol{A}(t+\tau-s)}\boldsymbol{B}\boldsymbol{U}(s)ds &= \int_0^{\tau} e^{\boldsymbol{A}s}\boldsymbol{B}\left(\frac{\boldsymbol{U}(t+\tau)-\boldsymbol{U}(t)}{\tau}(\tau-s)+\boldsymbol{U}(t)\right)ds \\ &= \mathbf{\Gamma}_h(\tau)(\boldsymbol{U}(t+\tau)-\boldsymbol{U}(t))+\mathbf{\Gamma}(\tau)\boldsymbol{U}(t) \qquad (4.20) \end{aligned}$$

Hence Eq. (4.19) becomes:

$$\boldsymbol{X}(t+1) = \mathbf{\Phi}\boldsymbol{X}(t) + \mathbf{\Gamma}_h(\tau)(\boldsymbol{U}(t+\tau)-\boldsymbol{U}(t)) + \mathbf{\Gamma}\boldsymbol{U}(t) + \boldsymbol{v}(t) \qquad (4.21)$$

The numerical details used, are described in Madsen and Melgaard (1991).

4.3 The likelihood function

It is assumed that the observations are obtained at regularly spaced time intervals, and hence that the time index t belongs to the set $\{0, 1, 2, ..., N\}$. N is the number of observations, or rather the length of the time series. Some of the observations may be missing. In order to obtain the likelihood function we further introduce:

$$\boldsymbol{Y}^*(t) = [\boldsymbol{Y}(t), \boldsymbol{Y}(t-1), \ldots, \boldsymbol{Y}(1), \boldsymbol{Y}(0)]' \qquad (4.22)$$

i.e. $\boldsymbol{Y}^*(t)$ is a matrix containing all the observations up to and including time t. Finally, let $\boldsymbol{\theta}$ denote a vector of all the unknown parameters - including the unknown variance and covariance parameters in $\boldsymbol{R}_1^c$ and $\boldsymbol{R}_2$. The case of a missing observation, is easily handled by assuming $\boldsymbol{R}_2 = \infty$ for that time instant, and taking $Y(t)$ to an arbitrary value.

The likelihood function is the joint probability density of all the observations assuming that the parameters are known, i.e.:

$$\begin{aligned} L'(\boldsymbol{\theta}; \boldsymbol{Y}^*(N)) &= p(\boldsymbol{Y}^*(N)|\boldsymbol{\theta}) \\ &= p(\boldsymbol{Y}(N)|\boldsymbol{Y}^*(N-1), \boldsymbol{\theta})p(\boldsymbol{Y}^*(N-1)|\boldsymbol{\theta}) \\ &= \left(\prod_{t=1}^{N} p(\boldsymbol{Y}(t)|\boldsymbol{Y}^*(t-1), \boldsymbol{\theta})\right)p(\boldsymbol{Y}(0)|\boldsymbol{\theta}) \qquad (4.23) \end{aligned}$$

where successive applications of the rule $P(A \cap B) = P(A|B)P(B)$ is used to express the likelihood function as a product of conditional densities.

Since both $\boldsymbol{v}(t)$ and $\boldsymbol{e}(t)$ are normally distributed the conditional density is also normal. The normal distribution is completely characterized by the mean and the variance. Hence, in order to parameterize the conditional distribution, we introduce the conditional mean and the conditional variance as

$$\hat{\boldsymbol{Y}}(t|t-1) = E[\boldsymbol{Y}(t)|\boldsymbol{Y}^*(t-1), \boldsymbol{\theta}] \tag{4.24}$$

and

$$\boldsymbol{R}(t|t-1) = V[\boldsymbol{Y}(t)|\boldsymbol{Y}^*(t-1), \boldsymbol{\theta}] \tag{4.25}$$

respectively. It is noticed that (4.24) is the one-step prediction and (4.25) the associated variance. These can be calculated recursively by a Kalman filter.

4.4 The Kalman filter

A Kalman filter is a recursive, linear, real-time processing algorithm, which is an optimal estimate of the states of a dynamic system in a noisy environment. It is used to recursively calculate the one-step prediction for the state of the system together with formulas for updating (or reconstructing) this estimate, based on a new observation.

In the present case where the transfer of the states of the system in discrete time is described by Eq. (4.21) and the observations by Eq. (4.12), the equations for *updating* the estimate of the state $\boldsymbol{X}$, according to Madsen (1989) and Melgaard and Madsen (1993b), becomes:

$$\hat{\boldsymbol{X}}(t|t) = \hat{\boldsymbol{X}}(t|t-1) + \boldsymbol{K}_t(\boldsymbol{Y}(t) - \hat{\boldsymbol{Y}}(t|t-1)) \tag{4.26}$$

$$\boldsymbol{P}(t|t) = \boldsymbol{P}(t|t-1) - \boldsymbol{K}_t \boldsymbol{R}(t|t-1)\boldsymbol{K}_t' \tag{4.27}$$

where the Kalman gain, $\boldsymbol{K}_t$, is:

$$\boldsymbol{K}_t = \boldsymbol{P}(t|t-1)\boldsymbol{C}'\boldsymbol{R}(t|t-1)^{-1} \tag{4.28}$$

The formulas for *prediction* becomes:

$$\hat{\boldsymbol{X}}(t+1|t) = \boldsymbol{\Phi}\hat{\boldsymbol{X}}(t|t) + \boldsymbol{\Gamma}_h(\boldsymbol{U}(t+1) - \boldsymbol{U}(t)) + \boldsymbol{\Gamma}\boldsymbol{U}(t) \tag{4.29}$$

$$\hat{\boldsymbol{Y}}(t+1|t) = \boldsymbol{C}\hat{\boldsymbol{X}}(t+1|t) \tag{4.30}$$

$$\boldsymbol{P}(t+1|t) = \boldsymbol{\Phi}\boldsymbol{P}(t|t)\boldsymbol{\Phi}' + \boldsymbol{R}_1 \tag{4.31}$$

$$\boldsymbol{R}(t+1|t) = \boldsymbol{C}\boldsymbol{P}(t+1|t)\boldsymbol{C}' + \boldsymbol{R}_2 \tag{4.32}$$

The formulas require some initial values, which describes the prior knowledge about the states of the system in term of the prior mean and variance:

$$\hat{\boldsymbol{X}}(1|0) = E[\boldsymbol{T}(1)] = \boldsymbol{\mu}_0 \tag{4.33}$$

$$\boldsymbol{P}(1|0) = V[\boldsymbol{T}(1)] = \boldsymbol{V}_0 \tag{4.34}$$

The matrix $\boldsymbol{P}(t+1|t)$ is the variance of the one-step prediction of the state, $\hat{\boldsymbol{X}}(t+1|t)$, of the system.

Using Eq. (4.23) - (4.25) the conditional likelihood function (conditioned on $\boldsymbol{Y}(0)$) is calculated:

$$\log L(\theta; \boldsymbol{Y}^*(N)) = -\frac{1}{2}\sum_{t-1}^{N}\left(\log\det \boldsymbol{R}(t|t-1) + \boldsymbol{\epsilon}(t)'\boldsymbol{R}(t|t-1)^{-1}\boldsymbol{\epsilon}(t)\right) + \text{const.} \tag{4.35}$$

where

$$\boldsymbol{\epsilon}(t) = \boldsymbol{Y}(t) - \hat{\boldsymbol{Y}}(t|t-1) \tag{4.36}$$

is the one-step prediction error. It may be noted that in case of missing observations we set $\boldsymbol{R}_2 = \infty$, which gives a zero contribution to the likelihood at that time instant.

Since it has not been possible to determine an explicit expression for the ML-estimator, numerical methods have been used. (See Madsen and Melgaard (1991)).

5 Results and discussion

Different models were tried and a satisfying model found by means of varying the initial values their range and checking the diagnostic information provided by the software used (see Melgaard and Madsen (1993b)). Estimating parameters in a continuous time model has the advantage, that the parameter estimates may be directly evaluated, physically. The estimates and their standard deviations are shown in Table 1.

All parameters, except organic matter, L and respiration $R(T)$, are significantly different from zero. Organic matter, therefore seems of little significance, in itself, in this creek, however the parameters linking the two

Parameter	K_{1c}	K	K_3	K_{1l}
Estimate	$0.398 \cdot 10^{-1}$	$0.702 \cdot 10^{-2}$	0.786	0.241
Std. Dev.	$0.194 \cdot 10^{-1}$	$0.415 \cdot 10^{-3}$	0.377	$0.199 \cdot 10^{-1}$
Unit	$O_2/L\,h$	h^{-1}	mgO_2/mgL	$L/O_2\,h$

Parameter	β	K_b	R_{15}	κ
Estimate	$0.108 \cdot 10^{-2}$	$0.332 \cdot 10^{2}$	$0.134 \cdot 10^{-8}$	$-.239 \cdot 10^{4}$
Std. Dev.	$0.457 \cdot 10^{-4}$	$0.118 \cdot 10^{2}$	$0.929 \cdot 10^{-8}$	$0.120 \cdot 10^{4}$
Unit	$mgO_2/W\,10^3\,l$	$mgO_2/l\,h$	mgO_2/l	mgL/l

Table 1: Maximum Likelihood-estimates for the oxygen system, per hour, and their standard deviations.

states were significant. It seems that the depth variations of the creek, play a larger role than is commonly the case. This may be because the exchange with air turns out to be a dominant factor. In general, all the estimated parameters fall into ranges proposed in the litterature (see Harremoës et al. (1989), Harremoës and Malmgren-Hansen (1990), Simonsen (1974)).

Stochastically the model is evaluated by using tests for parameters equal to zero, as discussed above, and tests for white noise of the residuals. The tests has shown that an assumption of white noise is reasonable.

Fig. 3 shows the measured and the simulated oxygen concentrations, while Fig. 4 shows the measured values and the one step predictions for a part of the series. It is seen that the obtained model fits quite well, though, of course, not perfectly.

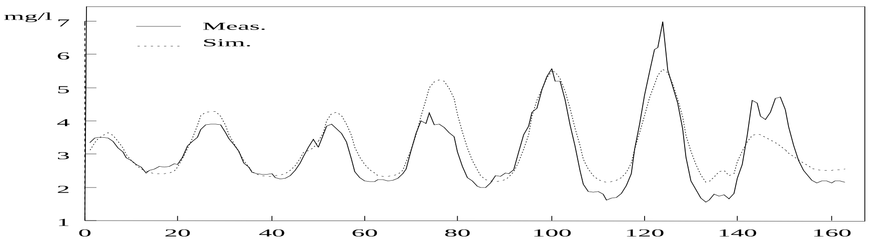


Figure 3: Measured and simulated oxygen for a part of the series.

This model had no direct time delay between the solar radiation input

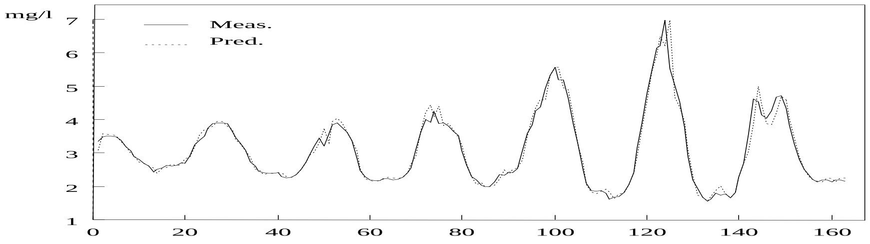


Figure 4: Measured and one step predicted oxygen for a part of the series.

and the oxygen output, and a four hour delay between precipitation input and the oxygen output. The parameter, κ, governing the precipitation is negative, suggesting that rain removes organic matter from the creek. However, this is possibly due to a dilution effect. Priorly it would be expected that oxygen is consumed when it rains, as a result of increasing levels of organic matter from sewage pipes and the precipitation itself. As it is seen in the simulation in Fig. 3, there is some variation that is not accounted for by the model.

The finding of no direct time delay between the solar radiation input and the oxygen response may just be because the time resolution is too large compared to the one hour sampling time. Still the time constants of the system will describe a phase shift between input and output variables, and hence the time difference between e.g. the maximum radiation and the corresponding maximum oxygen (as seen in Fig. 1). The rain has to be routed to the sewage system and then through that, before it finds its way to the recipient, making the four hour delay plusible.

More research will have to be done in refining the equations. A more thorough investigation would include a non-stationary mass balance, such as differential mass balance over an infinitely small element of water. Also the analysis of values from two or more stations in the stream should prove more accurate. The estimated model in the present paper may be considered as a first step towards a better modelling and understanding of the oxygen dynamics in creeks.

6 Conclusion

A continuous time linear stochastic model for the oxygen dynamics of a small creek is formulated and estimated using discrete time data, with missing values, and a maximum likelihood method. The stochastic evaluation showed, that the model was not contradicted by the statistical tests.

Purely statistical models, such as a transfer function model are useful in classifying data and describing poorly understood systems, but they generally offer little physical insight. They are usually sufficient for prediction and control applications. The advantages of the so-called black box approach is that traditional time series methods can be used for identifying the structure of the transfer function. As an alternative, physical information may be used to establish the structure of the model. Here a continuous time model has been formulated based on the known physical equations for the system and estimated using discrete time series. Thus, the model is based on the physical laws for the system, as well as the actual data, and the resulting model is called a grey box model.

With the use of the grey box approach, it was possible to obtain a model for a rather complex dynamical system using the available physical knowledge combined with statistical modelling tools. The resulting parameter estimation was physically evaluated as describing the system in accordance to what is known and expected for this kind of system.

Keeping the model formulation and parameter estimation in continuous time has several advantages. Experiments with missing observations, which most traditional time series analysis approaches cannot deal with, are easily accommodated. The model can easily be improved by a combination of detailed comparison with data and the use of physical facts, and furthermore a direct physical interpretation of the estimated parameters is possible.

The model obtained had no direct time delay from solar radiation to the oxygen concentration and a four hour time delay for the influence from precipitation. While organic matter usually is of great importance when evaluating the oxygen dynamics of a stream, it seemed to have little influence in this particular creek, while water level had more impact.

Acknowledgements

The authors wish to thank Højbakkegård for providing the solar radiation data, the department of Agricultural Sciences, Section of Soil and Water and Plant

Nutrition, The Royal Veterinary and Agricultural University, Copenhagen, Denmark. Also gratitude is due to The Danish meteorological Institute, DMI, and Ballerup Municipality for providing precipitation, and oxygen and temperature data, respectively.

Paper B

Modelling the Transient Impact of Rain Events on the Oxygen Content of a Small Creek.

Published in: Water, Science and Technology, 33(2),

pp. 177–187 (1996)

Oxygen has been monitored in a small urban creek for a two months period. The purpose of the data collection is to use statistical modelling to obtain a better understanding of the phenomena governing the oxygen concentration. The long term goal is to influence the process of urban runoff to achieve oxygen concentration suitable for fish. Oxygen fluctuates significantly on a diurnal basis. Furthermore the purpose is to establish a model by which we are able to use rain data to estimate extreme values for comparison with standards for minimum oxygen concentration. The experience is that the fluctuations cannot be explained adequately on a deterministic basis alone. Significant stochastic variation has to be accounted for.

Data from a small stream is used to identify a dynamic model of the oxygen level as a function of solar radiation, water depth, and rain. The model is formulated in continuous time as two coupled stochastic differential equations. The continuous time formulation makes it possible directly to interpret the parameters of the model. Hence the model is useful for monitoring the actual state of the stream.

In this paper a grey box modelling approach, which is a statistical method taking the known physical relations into account, is used. This approach is closely related to the continuous time formulation. The parameters of the model are estimated using discrete time data and a maximum likelihood method. In evaluating the likelihood function a Kalman filter is used.

The dynamic model makes it possible to assess the transient impact of the urban runoff due to rain events, as well as the effect due to solar radiation. The ultimative outcome of the analysis is to determine the required size and location of storage basins to be constructed in the sewer system, in order to decrease combined sewer outflows and extreme oxygen depletion during rain.

Keyword: Continuous time stochastic modelling; Grey-box models; Oxygen dynamics; Time series analysis;

2 Introduction

By long standing tradition, rivers and streams are recipients for diluted wastewater that may pollute the water. This would be of little importance if rivers and streams were not considered a quality recreational advantage, especially in urban areas. In countries, where surface water is abstracted for drinking water, it is even more of a problem. Therefore, modelling water quality is an important pursuit in order to understand and quantify the processes, which take place in the water. The oxygen level in a river (measured as dissolved oxygen) is one of the most universal environmental measures of water quality, which makes it one of the most interesting processes to model.

In urban areas, one of the phenomena that most often threaten water quality is the sudden transient impact from runoff due to heavy rain events, where the wastewater treatment plants cannot keep up with the huge amounts of water. Modelling the dynamics and seemingly random fluctuations cannot be done adequately on a deterministic basis. The stochastic variation must be included.

The interest in stochastic modelling of dissolved oxygen as a quality parameter in rivers and streams is not new. Substantial work was done more than a decade ago (Beck and Young (1975a), Whitehead and Young (1975), Thyssen (1980), Cosby (1984), Cosby et al. (1984), Erlandsen and Thyssen (1983)). Recently, new work is emerging Jacobsen and Voss (1994b), Jacobsen and Madsen (1996), Pedersen (1994) on the modelling and the modelling method, and Stigter and Beck (1994) on model structure identification. Once a system is adequately modelled, different control measures may be tried out as in Young and Beck (1974). The primary goal of this paper is to model the transient impact of rain events on the oxygen content of a small creek.

This is a new venue, since previous work has concentrated mainly on the relationship between dissolved oxygen (DO) and biological oxygen demand (BOD) alone. However, already Whitehead and Young (1975) noted: *It is the transient violations of water quality standards that cause most problems in the short term.* In their paper a deterministic stream flow model was combined with a black box stochastic rainfall-runoff model. The dynamics from precipitation were not directly included in the model and only daily data were available. Beck and Young (1975a) argued that further research on the diurnal variation was required. The data used in the present study are sampled on an hourly basis and thus support consideration of the distinct diurnal variation of the DO-BOD relationship.

In this paper a grey box modelling approach is used. This approach is closely related to a continuous time formulation, as it is formulated using two coupled stochastic differential equations. The continuous time formulation is well suited for taking into account known physical relations and enables a direct interpretation of the parameters of the model. The parameters are estimated using discrete time data and a maximum likelihood method.

3 Data description

Nearly two month of data are collected from a small urban stream, Harrestrup river, in a suburb of Copenhagen, Denmark. The measurements of the input and output variables for a five day period are depicted together, scaled for comparison, in Fig. 1. The figure shows that the oxygen level

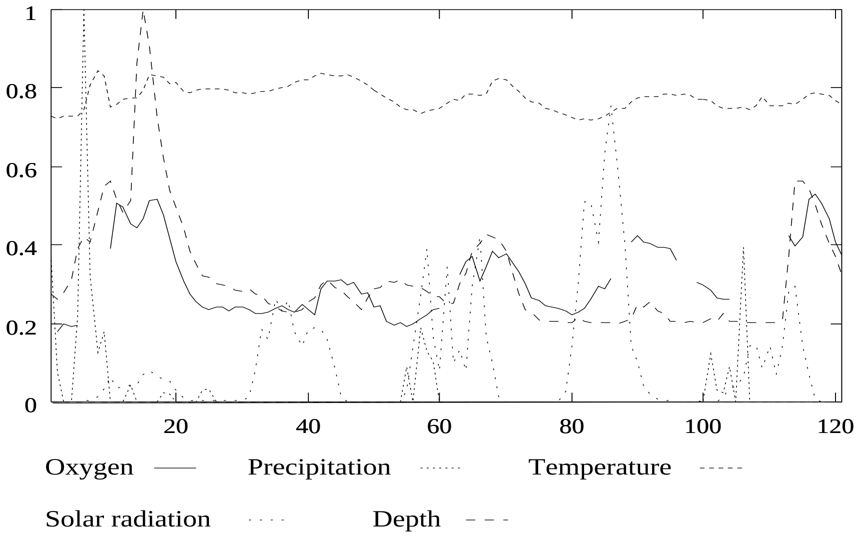


Figure 1: The scaled measurements of oxygen, solar radiation, water depth, temperature and precipitation. Note the missing values for the oxygen measurements.

is affected by variations in both solar radiation and precipitation. There seems to be a delayed response in relation to the input, specifically from a shift in the solar radiation. The response to precipitation input is not as distinct as the response to the solar radiation. The depth of the water in the stream is seen to be highly correlated with the precipitation, although delayed by the runoff process. Whenever the water level shows an increase, so does the oxygen level.

Collection and sampling of data, as well as a thorough description of each of the variables, can be found in Jacobsen and Voss (1994b) and Jacobsen and Madsen (1996). All input and output variables are listed in Table 1.

Temperature is measured in deg. Celsius.

TABLE 1: INPUT AND OUTPUT VARIABLES

Symbol	Parameter	Unit
C	Oxygen (DO)	$[M\ L^{-1}]$
L	Organic matter (BOD)	$[M\ L^{-1}]$
I	Solar radiation	$[E\ L^{-2}\ T^{-1}]$
P_r	Precipitation	$[L\ T^{-1}]$
h	Water level	$[L]$

Table 1: T: Time; M: Mass; L: Length; E: Energy

3.1 The model

The collected data is used to identify a dynamic model of the oxygen level as a function of solar radiation, temperature, water level, and precipitation.

Depth of the water is included in order to model the effects from the transient rain events. The water level varies quickly as a response to rain, when it exceeds a certain amount. Including this as an input variable facilitates taking the dynamics of these non-uniform effects into account. In Jacobsen and Madsen (1996), it is argued that the variation of oxygen and organic matter can be described by the following non-linear state space model:

$$\begin{aligned}
\begin{bmatrix} dC \\ dL \end{bmatrix} &= \begin{bmatrix} -\frac{K}{h\sqrt{h}} & -K_c \\ K_3 & -K_l \end{bmatrix} \begin{bmatrix} C \\ L \end{bmatrix} dt \\
&+ \begin{bmatrix} \beta & \frac{\sqrt{C}\,K_b}{h} \\ 0 & \kappa \end{bmatrix} \begin{bmatrix} I \\ P_r \end{bmatrix} dt \\
&+ \begin{bmatrix} \frac{K}{h\sqrt{h}} C_m(T) - R(T) \\ 0 \end{bmatrix} dt + \begin{bmatrix} dw_1 \\ dw_2 \end{bmatrix} \qquad (3.1) \\
\text{where} \quad C_m(T) &= 14.54 - 0.39T + 0.01T^2 \qquad [mg/l] \qquad (3.2)
\end{aligned}$$

$C_m(T)$ is the saturated oxygen as a function of temperature, T. This function was suggested by Beck and Young (1975a) in a study of the river Cam in England. The parameters of the model are briefly explained in Table 2.

TABLE 2: PARAMETERS AND THEIR UNITS

Symbol	Parameter	Unit
K	Adsorption coefficient	$[T^{-1}\,L]$
K_c	Oxygen consumption constant	$[T^{-1}]$
K_3	Organic matter production	$[T^{-1}]$
K_l	Degradation constant for organic matter	$[T^{-1}]$
β	Photosynthetic constant	$[E\,L^{-2}\,T^{-1}]$
R_T	Respiration at temperature, T	$[M\,L^{-1}\,T^{-1}]$
K_b	Delayed oxygen demand from the bottom	$[M\,T^{-1}]$

Table 2: T: Time; M: Mass; L: length; E: Energy

The processes considered for dissolved oxygen are reaeration, consumption from degradation of organic matter, net production from photo synthesis and respiration, and a delayed oxygen demand from the bottom related to rain. In order to include the depth of the river, reaeration is expressed via the adsorption coefficient, K, which is a function of the gravity constant and the slope of the river (see Harremoës and Malmgren-Hansen (1990)). The usual reaeration coefficient, K_2 equals $K/\sqrt{h}$. The delayed oxygen demand from the bottom, K_b is modelled as a half-order function (i.e. a square root function) of the oxygen concentration, as this is considered to describe the dynamics in a bacterial layer, such as the one assumed to be on the lining of the creek, Harremoës et al. (1989). This delayed oxygen demand is believed to be a consequence of rain. A time delay of four hours was found between the P_r and the oxygen concentration, i.e. $P_r(t-4)$ is in fact used in (3.2). It was found by trying out all reasonable candidates for the time delay (see Jacobsen and Voss (1994b)).

In the modelling, however, it was found that only by letting the oxygen concentration remain constant as the mean value, the estimation procedure would converge. For organic matter, expressed as BOD, there is a degradation and a production term.

The stochastic terms dw_1 and dw_2 are additive terms introduced to describe the model uncertainty. They may be representative of unknown terms, that are lacking in the model, due to for instance meteorological or other poorly understood processes. Beck and Young (1975a), include a term which models algae growth as a function of sustained sunlight effect over 24 hours as only daily data from the river Cam was available. Algae growth was not expected to be of primary interest in this specific stream.

Only the oxygen concentration is measured. Thus, the so-called observation equation, can be written as:

$$C_r(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} C(t) \\ L(t) \end{bmatrix} + e(t) \tag{3.3}$$

where $e(t)$ is the measurement error. BOD is not measured, but due to the state space formulation this is not necessary, since it is estimated using a Kalman filter (see Jacobsen and Voss (1994b) or Jacobsen and Madsen (1996)).

3.2 Modelling the photo synthesis

With hourly data for solar radiation there is a possibility to model photosynthesis directly. Cosby (1984) and Erlandsen and Thyssen (1983) investigated several empirical functions for photo synthesis, using data from Gryde river, a small second order stream in Jutland, Denmark. Here only the most straight forward linear model and one of the proposed two-parameter models are used.

$$P(I) = \beta\, I \tag{3.4}$$

$$P(I) = P_m\, E_0 \frac{I}{P_m + E_0\, I} \tag{3.5}$$

I is the solar radiation, P_m is the maximum photosynthetic rate and E_0 the slope of the $P - I$ curve as the light approaches zero. The two parameter model was ranked the best in the references previously mentioned.

Both expressions were tried, but the one-parameter expression performed equally well compared to the more complicated model. Therefore the final estimates are shown only for the linear model (3.4).

3.3 Temperature corrections

Correcting several of the parameters for temperature has been suggested by Dobbins (1964) (the K_2 coefficient) and Erlandsen and Thyssen (1983). The latter estimated the coefficient for the respiration in a study of Gryde river, in Denmark. Their analysis was carried out for night time observations only. The following corrections, found in the literature mentionened above, were introduced into the equations:

$$K_2 = \kappa\, \theta^{T-20} \qquad \theta = 1.015\text{-}1.047 \tag{3.6}$$

$$\beta = \pi\, \theta^{T-20} \qquad \theta = 1.0241 \tag{3.7}$$

$$R_T = \rho\, \theta^{T-20} \qquad \theta = 1.0713 \tag{3.8}$$

However, no improvement was found. Hence it was decided to leave out the temperature corrections.

In other studies, Beck and Young (1975a) of the river Cam and Whitehead and Young (1975) of the Bedford-Ouse, two measuring stations were used. This enabled the model to include the volume and to view the system as a continuous stirred tank reactor. Here only one measuring station was available at the present time. This is also the case in the modelling of Gryde river (Thyssen (1980), Cosby et al. (1984), Pedersen (1994)). With observations from only one measuring station the formulation in Eq. (3.2) and (3.3) implicitly assume uniform and steady flow conditions. This may not be fully correct in the case of sudden rain events. However, since the time constants (of the transient impact) are small compared to the ones involved in the DO-BOD exchange, it will suffice for the initial investigation of the dynamics of the oxygen system related to rain events in urban rivers.

3.4 The estimation method

To estimate the parameters of the model, a program (CTLSM) for modelling continuous time systems (Melgaard and Madsen (1993b)) was used. The methods used are described extensively in Madsen and Melgaard (1991). The estimation is performed by a maximum likelihood method, and in the evaluation of the likelihood function an ordinary Kalman filter is used. However, for non-linear systems, an extended Kalman filter (EKF) is used.

The stochastic terms $w_1(t)$ and $w_2(t)$ are assumed to be Wiener processes and the measurement error, $\boldsymbol{e}(t)$ is assumed to be a Gaussian white noise process. Furthermore, it is assumed that $w_1(t)$, $w_2(t)$ and $\boldsymbol{e}(t)$ are mutually independent.

Young (1974) warns that even if all theoretical assumptions were fully satisfied, there is no guarantee of convergence for the EKF, but with appropriate care in use, it should work well in practice. Cosby and Hornberger (1984) used the EKF to discriminate well among various models applied to synthetic data, but felt that the question of whether this would still apply to real systems was not adequately answered. In Cosby et al. (1984) the EKF was applied to real data and proved useful for the estimation. However, whether the correct model was indeed identified, is obviously not known.

Specifying a non-linear model and then using the option in CTLSM for the EKF resulted in convergence problems in the case of Harrestrup river data. Whether this is inherent in the EKF itself or due to an incorrect model is a matter still to be investigated.

4 Results and discussion

The parameters were estimated from the hourly data and converted to per day values, since this is the normally used units found in the literature. The estimates found and their standard deviations are shown in Table 3.

Estimating parameters in a continuous time model has the advantage that the parameter estimates may be directly evaluated, physically. A likelihood ratio test has shown that R_T is not significantly different from zero. All other parameters were significant. From Table 1 it is seen that the parameters are estimated with a smaller standard deviation when R_T is set equal to zero, instead of being estimated. The measured and the predicted timeseries for oxygen are seen in Fig. 2.

The estimated parameters are not directly comparable with estimates found in the literature because of the inclusion of water depth in the equations. If we take the inclusion of water depth into account, the estimates for reaeration and degradation, fall into ranges proposed in the literature (see Simonsen (1974), Harremoës et al. (1989), Harremoës and Malmgren-Hansen (1990), Beck and Young (1975a), Whitehead and Young (1975)). Especially reaeration have been studied intensively by Thyssen (1980) and Cosby et al. (1984). This is assuming that the units correspond to the sampling, such that the estimates for data measured on a daily basis are per day. However, the rivers, for which the estimations were obtained, are all different. The river Cam, for instance, have a mean temperature of 8 ^{0}C, while the mean temperature in Harrestrup is nearly 15 ^{0}C.

Respiration, has not been the subject of most of the reported investigations, but it is intuitively known that it should not be zero. In most modelling

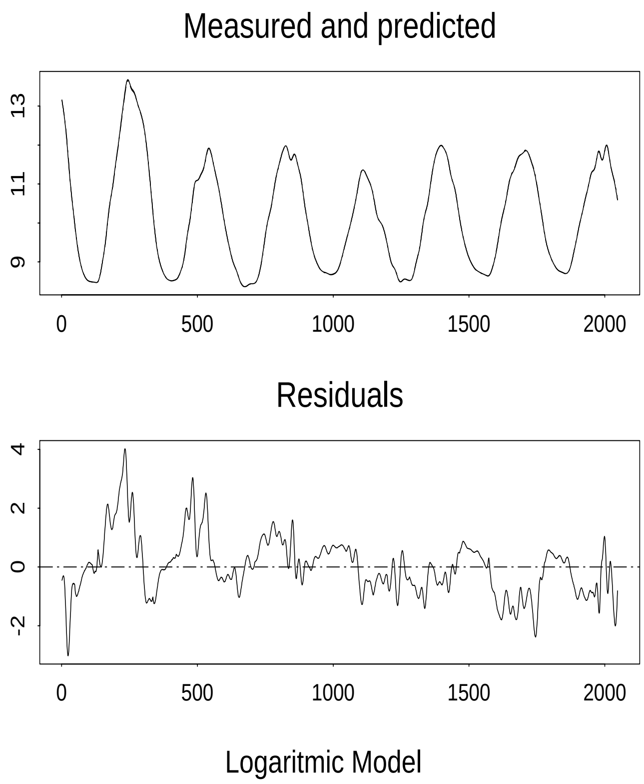


Figure 2: The prediction and the standardized residuals.

TABLE 3: MAXIMUM LIKELIHOOD ESTIMATES

Parameter	R_T est.	$R_T = 0$	Unit
K	0.159	0.159	d^{-1}
	(0.014)	(0.013)	
K_c	21.276	21.276	d^{-1}
	(1.634)	(1.591)	
K_3	2.068	2.068	d^{-1}
	(0.121)	(0.116)	
K_l	10.422	10.422	d^{-1}
	(1.041)	(1.020)	
R_T	$0.27\ 10^{-9}$	0.000	$mg\ l^{-1}\ d^{-1}$
	$(0.13\ 10^{-7})$	—	
β	$0.12\ 10^{-2}$	$0.12\ 10^{-2}$	$mg\ W^{-1}\ m^{-1}$
	$(0.73\ 10^{-4})$	$(0.72\ 10^{-4})$	
K_b	14.574	14.570	$mg\,l^{-1}$
	(4.309)	(4.607)	
κ	-110.420	-110.420	$mg\,l^{-1}$
	(28.795)	(30.337)	

Table 3: ML estimates per day, and their standard deviations in ().

efforts (deterministic or stochastic) this term has always been considered constant, which seem to be mostly for the sake of obtaining a linear model. Odum (1956) proposed that respiration may be dependent of organic matter, but this have not been tried specifically here. However, the parameters for degradation of organic matter, K_c and K_l, which causes depletion of oxygen and organic matter, may then be viewed as a combination of both degradation and respiration. This could explain the rather large constants. Also, the respiration of a concrete lined creek like Harrestrup river, with little actual plant-life, should not be expected to be of the same size as natural healthy country rivers. Thus the extremely small (or non-existing) reaeration may be characteristic of this kind of urban waterway.

The model is evaluated statistically by testing each of the parameters for a hypothesis that it should be equal to zero. The assumption for the stochastic terms is tested by testing the hypothesis that the residuals are white noise. This is tested using both the autocorrelation function (see Fig. 3) and the cumulative periodogram (see Fig. 4), Melgaard and Madsen (1993b).

Furthermore, the cross correlation functions between the residuals and various input signals are calculated.

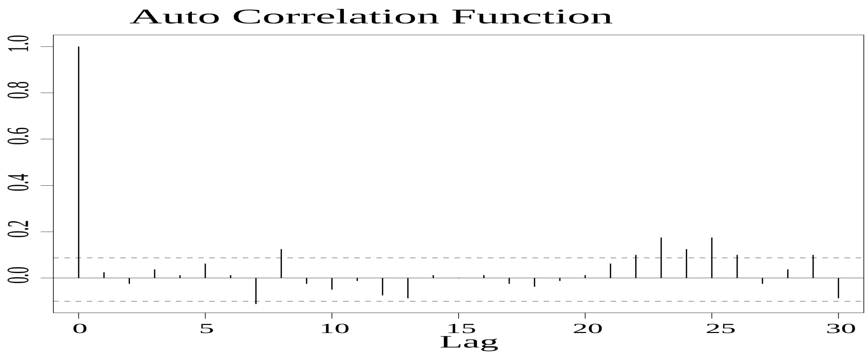


Figure 3: The autocorrelation (ACF) function for the residuals.

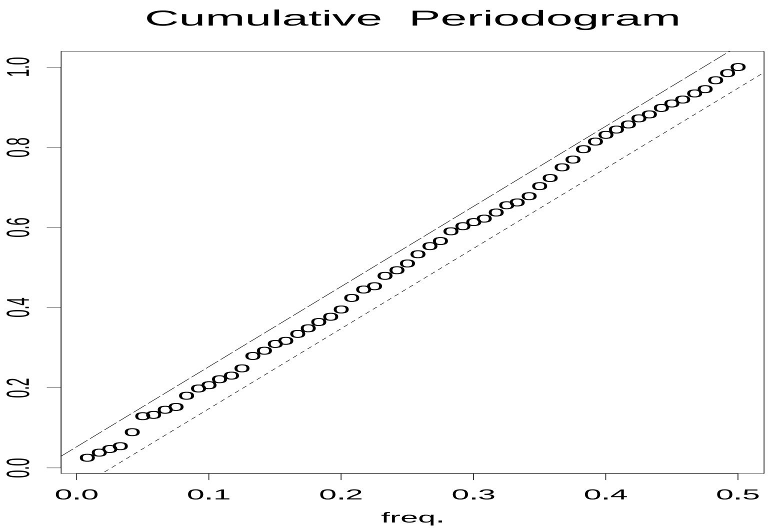


Figure 4: The cumulative periodogram for the residuals.

5 Conclusions

In this paper a continuous time model for the oxygen dynamics of a small stream describing the transient impact of rain has been formulated. This was based on known physical equations for the system and estimated using a maximum likelihood method. Thus, the model is based on the physical laws of the system, as well as the actual data. The method used accommodates missing observations (see Jacobsen and Madsen (1996)) and estimates the parameters of a rather complex dynamical system using the available physical knowledge combined with statistical modelling tools. The method also facilitates a direct physical interpretation of the estimated parameters.

All the estimated parameters were within physically expected bounds, except that the respiration was not significantly different from zero. This may not be as surprising after all, when the actual type of urban river is taken into account. There are few macrophytes in this creek, unlike in natural country rivers.

A model has been established for a creek, highly influenced by urban runoff, based on data for solar radiation, precipitation and the varying depth of the water. An attempt to use a non-linear model was not succesful. The average concentration of oxygen is used instead to model the half-order dynamics of the oxygen consumption from the bacterial layer on the slabs of concrete lining the creek bottom.

Several tests for white noise behavior of the residuals indicates that the model is able to describe the dynamics. In the literature, some models are simple one-dimensional models of the type proposed by Odum (1956). However, an attempt to reduce the model to a one-dimension model proved the need of a second state and hence a two-dimensional model. As the second state the organic matter seems to be the natural choice.

Paper C

Modelling the Dynamical Effect of Solar Radiation on the Oxygen Content of a Small Creek.

In press: EnvironMetrics (1997).

Abstract

Data from a small stream is used to explore various dynamical models of the oxygen level. Special attention is given to an analysis and a description of how the solar radiation affects the oxygen level. The models are formulated in continuous time using stochastic differential equations. The parameters of the models are estimated using discrete time data and a maximum likelihood method. The continuous time formulation makes it possible to directly interpret the parameters of the model, as the estimated parameters have specific physical, biological or chemical meaning. The advantages of using a low pass filter on the point measured solar radiation, as an input variable, is demonstrated. This investigation will eventually contribute to a comprehensive study of the dynamics that governs the oxygen content in a recipient.

KEY WORDS: Solar radiation; Oxygen dynamics; Time series analysis; Dynamical modelling; Low pass filter.

1 Introduction

The variation of the oxygen content in aquatic systems like rivers is highly dependent on the variation of short wave radiation. About 44 % of the total radiation lies in the photo synthetic interval. Photosynthesis is a chemical reaction, by which plants (in a river) convert light energy into chemical energy. Production by photosynthesis and consumption from respiration takes place in the water phase by phytoplankton, and by the macrofytes growing from the river bed.

The Gryde river is a small creek, located on the Danish Jutland peninsula. There is no biological oxygen demand exerted from organic matter in excess of natural levels and the agriculture surrounding the river is not expected to supply large amounts of oxygen consuming matter. Measuring stations were installed in the late 1970's and other data sets from this river have since been the subject of many studies (Thyssen et al. (1983), Erlandsen and Thyssen (1983), Jeppesen and Thyssen (1984), Cosby (1984), Cosby et al. (1984), Cosby and Hornberger (1984), Erlandsen and Sørensen (1984)).

The purpose of the present work is to investigate the photosynthesis as a function of solar radiation, as observed indirectly by the oxygen concentration. Eight different mathematical models for the photosynthesis-light relationship will be investigated.

The models have all been previously suggested (Thyssen (1980), Cosby (1984), Cosby et al. (1984)). Seven of the models have been analysed in Cosby et al. (1984), however the methods used in this paper are different. In their approach, a recursive estimation technique using an extended Kalman filter (EKF) (see Cosby and Hornberger (1984)) was used to estimate the parameters. As pointed out by the authors, and demonstrated clearly in Ljung (1979), this technique is critically dependent on a correct specification of some variance parameters used in the EKF. The recursive estimates are used to construct average values/estimates of the parameters. Yet another method was presented by El-Shaarawi et al. (1988), who used a non-linear regression analysis approach via a Taylor series expansion. Five models were investigated in this work. The disadvantage of this method lies in the approximation, the Taylor series expansion represent.

In the present paper, a maximum likelihood method, described in Melgaard and Madsen (1993a), is used to estimate the parameters of a continuous time, non-linear model. Compared to the methods used by Cosby et al. (1984) our method directly provides estimates of the parameters. Furthermore, a method is suggested for low pass filtering the solar radiation data. This filtering is suggested to compensate for the fact that the solar radiation is a very high frequency signal, which is probably measured some distance from the main body of the water. Therefore, only the low frequency part of the variation is expected to be representative. This is confirmed by the fact that the use of the filtered solar radiation as input leads to more precise parameter estimates and a considerably improved description of the variation of the oxygen content.

2 Description of data

The data in this study is measured several years after the previously mentioned studies were conducted. The aim is more concerned with the estimation of the physical parameters than with the performance of our estimation routine, as this have been tested on different kinds of physical models (see e.g. Madsen et al. (1990), Melgaard (1994), Melgaard et al. (1994), Jacobsen and Voss (1994b), Jacobsen and Madsen (1996), Madsen and Holst (1995)).

The data consists of measurements of oxygen, saturated oxygen, temperature and solar radiation[1], measured every 5 minutes during one week in June 1985 (see Figure 1). Water depth is also measured, but is not ex-

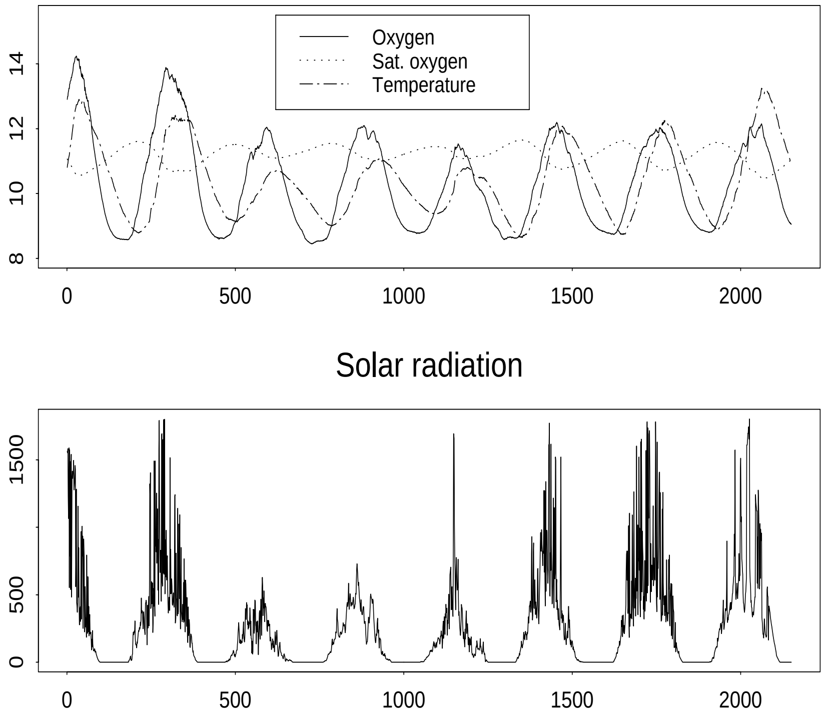


Figure 1: Measurements from Gryde river of the oxygen concentration and the saturated oxygen concentration in mg/l, the water temperature in degrees Celsius and the solar radiation in Langley/min. The X-axis shows the time in 5 minute intervals.

plicitly involved in this model, as the depth is almost constant (about 73 cm.) over the period. Also the cross section of the river is fairly uniform.

[1]Here solar radiation is measured in Langley, Ly, which is a non standard unit used in many earlier studies. Today solar radiation is most frequently measured in W/m^2 per hour: $Ly = cal/cm^2 = 41.868kJ/m^2$

See Thomsen and Thyssen (1979), Thyssen et al. (1983), Erlandsen and Thyssen (1983) for a description of the instrumentation and the experimental set-up.

3 Filtering of data

The solar radiation is measured as a point measurement, since it is obtained using one single solarimeter. This means that the variation of the solar radiation is dominated by high frequency variation, due to movements of clouds (Madsen et al. (1985)).

In this paper it will be demonstrated that an improved description of the variation of the oxygen content in the river, is obtained using a low pass filtering of the solar radiation data instead of the raw data itselves. There are several possible explanations for this: First of all, the photosynthesis may show some persistence, such that the irradiance to photosynthesis relationship must be a dynamic one. Secondly the measurement of solar radiation does not take place along the element of water of interest – it is most likely measured some distance away from the position of the main body of the river. Therefore, the high frequency part of the variation is of no – or minor – relevance to the actual variation of the oxygen.

Depending on water quality, and how the solar radiation penetrates the water surface, the river may act as a low pass filter in itself. If this is the case it may not be correct to use the raw and unfiltered solar radiation data in the estimation procedure. The signal to noise ratio could obscure important dynamical information, thus making it more difficult to correctly estimate the parameters of the model. Hence, a low pass filtration of the solar radiation data should be performed.

The effects of considering a low pass filtered series of solar radiation as input is investigated by choosing a simple low pass filter. In further studies of models for the oxygen content in aquatic systems, we will consider the possibility of estimating an optimal low pass filter for the solar radiation.

3.1 A low pass filter

An ideal low pass filter, is a filter which only allows frequencies smaller than a given frequency to pass undisturbed, while holding back the higher frequencies. In practice, however, it is impossible to obtain the ideal low pass filter (Bloomfield (1976)).

Kaiser and Reed (1977) suggests four different filter types that are all fairly simple and efficient. Of these, the so-called Nearly Equal Ripple approximation, NER, which is an approximation to the ideal low pass filter, is close to being the most efficient. It is characterised by nearly equal pass- and stop-band magnitude errors.

For a discrete time series the filtered output data, y_t, is calculated by folding the original input data, x_t, with the impulse response function, h, for the filter, i.e.

$$y_t \;=\; \sum_{k=-\infty}^{\infty} h_k \, x_{t-k} \tag{3.1}$$

This filter, however, requires an infinite input signal, which is not available in practice. Therefore only a span, N_p, of the filter weights is used at the expense of a less ideal filter. The folding equation thus reduces to:

$$y_t \;=\; \sum_{k=-N_p}^{N_p} h_k \, x_{t-k} \tag{3.2}$$

where N_p is the number of terms after which the series is truncated. Kaiser and Reed (1977) suggests an odd number of terms to be used in the filter, chosen by $N = 2N_p + 1$, to avoid a half-interval time delay.

In Figure 2 different types of filters are shown. By increasing the number of terms in the filter, N_p, a steeper slope is obtained, and thereby a more ideal low pass filter. Simple averaging is frequently considered as a low pass filter. As indicated in Figure 2 averaging is far from the ideal low pass filter.

The NER (nearly equal ripple) filter uses the adjustable weighting function, h_k, the effect of which is to reduce the error in size to $\pm\epsilon$ at the expense of a slightly broadened transition band δ. To design it, certain parameters have to be selected. In addition to the ones mentioned above (ϵ and δ) is typically the band edge frequency, β (see Figure 3).

The effect of using an adjustable weighting function is to reduce the error in size to $\pm\,\epsilon$ at the expense of a slightly broadened transition band, δ. This is accomplished by spreading the large approximation error at the band edge (resulting from series truncation) over the whole of the pass and stop

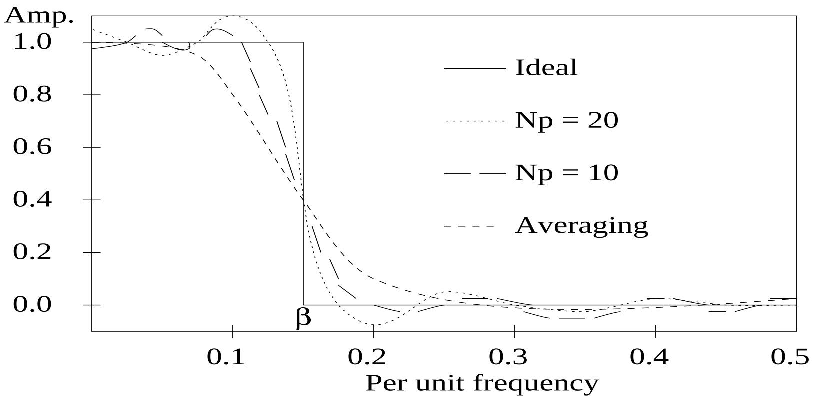


Figure 2: Characteristics for different type filters (Amp. is short for amplitude).

bands. A reduction in passband ripple or the approximation error is paid for by a wider transition band.

The approximation error, ϵ, the transition band, δ, and the number of coefficients, N_p, are related by the design formula (Kaiser and Reed (1977)):

$$N_p \quad \approx \quad (-20 \log_{10} \epsilon - 7.95)/28.72\,\delta \tag{3.3}$$

The filter is described in more detail in Kaiser and Reed (1977), Jacobsen and Voss (1994a). In the present work the NER-filter is used to filter the high frequency variations from the solar radiation data with the parameters: $N_p = 26$, $\epsilon = 0.02$, $\delta = 0.035$ and $\beta = 1/12$ (according to the Shannon theorem). Figure 4 shows the time series of the solar radiation data after being filtered with the NER filter. A common use of the NER filter is to avoid aliasing when sub-sampling from a faster sampling interval. By using the filter one avoids folding the high frequencies into the sampled Nyquist interval (Bloomfield (1976)).

4 Model equations

Most often the simple first order model, formulated by Odum (1956), is used for a description of the dynamics of the oxygen concentration in rivers:

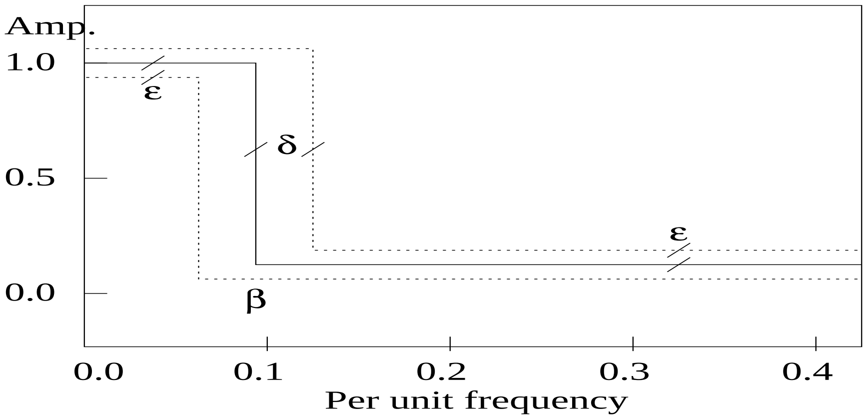


Figure 3: Design definitions for the parameters of the NER-filter, where ϵ is the approximation error, δ is the width of the transition band and β the band edge frequency.

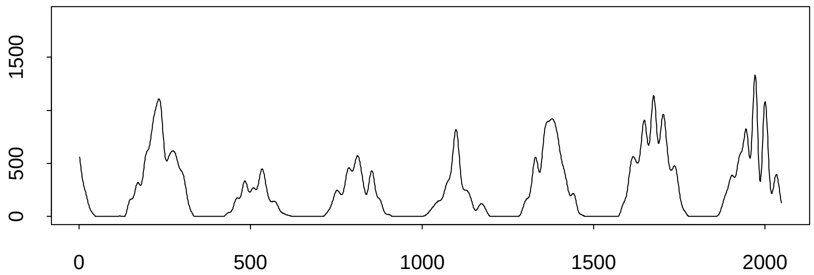


Figure 4: Solar radiation (in Langley/min.) data after using the NER filter with $\epsilon = 0.02$, $\delta = 0.035$ and thus $N_p = 26$. The X-axis shows time in 5 minute intervals.

$$\frac{dC}{dt} = K(C_m - C) + P(I) - R \tag{4.4}$$

where C is the oxygen concentration, C_m the saturated oxygen concentration (both in mg/l), K the reaeration constant (in hr^{-1}), and R (mg/l), the respiration, while $P(I)$ (mg/l) is the photosynthesis described as a function of solar radiation, I (in Ly/min.).

This model includes only the state of oxygen. The respiration was considered constant, while the photosynthesis was expected to vary in a diurnal sine-like fashion, (solar radiation was not measured). Often solar radiation is not measured and therefore the photosynthesis is modelled by some empirical function e.g. a sustained sunlight effect (see Young and Beck (1974), Beck and Young (1975a), Whitehead and Young (1975), Whitehead (1980)). Later studies (Thyssen et al. (1983), Erlandsen and Thyssen (1983), Jeppesen and Thyssen (1984), Cosby (1984), Cosby et al. (1984), Cosby and Hornberger (1984), Erlandsen and Sørensen (1984), El-Shaarawi et al. (1988)) modelled the photosynthesis directly as a function of solar radiation.

In the models for photosynthesis, some physical parameters are introduced. Regardless of the exact function for the photosynthesis-light relationship, the following main parameters are introduced to characterise the relationship; the slope of the initial photo-synthetic curve as the light intensity approaches zero, E_o, and the maximum photo-synthetic rate, P_m. The ratio E_o/P_m is the light intensity at which, the line tangent to the initial slope intersects the line tangent to the maximum photo-synthetic rate. The equations for 8 different photo-synthetic models (as a function of the measured solar radiation) are listed below and depicted in Figure 5.

4.1 Temperature dependency

Some of the parameters in (4.4) are temperature dependent, namely the reaeration coefficient, K, the respiration rate, R and the photosynthesis light relationship $P(I)$. Typically they are modelled by exponential functions of the Arrhenius type:

$$\begin{aligned} K(T) &= \kappa_{20}\, 1.0241^{T-20} && [hr^{-1}] && (4.5) \\ R(T) &= \rho_{20}\, 1.07^{T-20} && [mg/l] && (4.6) \\ P(T,I) &= 1.035^{T-20}\, P(I) && [mg/l] && (4.7) \end{aligned}$$

$$
\begin{aligned}
1: \quad P(I) &= E_o I \\
2: \quad P(I) &= E_o I^{\alpha} \\
3: \quad P(I) &= P_m \frac{1}{P_m/E_o I + 1} \\
4: \quad P(I) &= P_m \frac{1}{\sqrt{(P_m/E_o I)^2 + 1}} \\
5: \quad P(I) &= P_m(1 - \exp(-E_o I/P_m)) \\
6: \quad P(I) &= P_m \tanh(E_o I/P_m) \\
7: \quad P(I) &= E_o I(1 - E_o I/P_m) \\
8: \quad P(I) &= P_m \ln(1 + E_o I/P_m)
\end{aligned}
$$

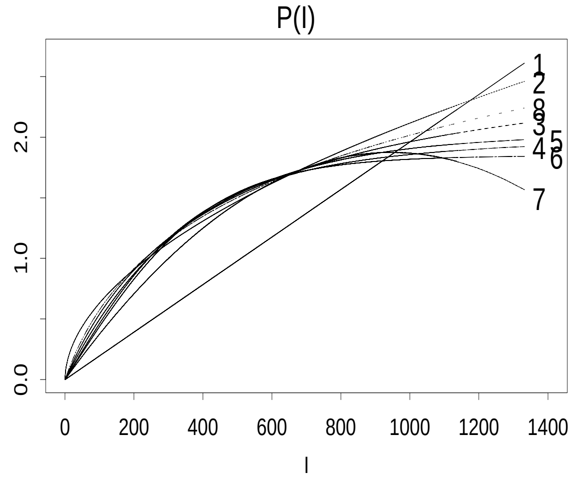


Figure 5: Equations for the eight photo-synthetic models, P(I) as a function of the solar radiation. The unit for P(I) is mg/l. The functions are shown here with the parameter values estimated from the filtered solar radiation data.

where the water temperature, T is measured in degrees Celsius. The correction parameter, 1.07 for respiration are suggested by Simonsen (1974) and Erlandsen and Thyssen (1983), whereas the correction parameter, 1.0241 for reaeration was found in Pedersen (1994) and the temperature correction for photosynthesis is 1.035 according to Thyssen (1980). Respiration from plants is specific to the type of plant and its habitat.

The differential equation (4.4) may then be rewritten as

$$\frac{dC}{dt} = AC + a \qquad [mg/l] \tag{4.8}$$

where the terms (4.5), (4.6) and (4.7) are included, i.e.

$$\begin{aligned} A &= -\kappa_{20} 1.0241^{T-20} && (4.9) \\ a &= \kappa_{20} 1.0241^{T-20} C_m + 1.035^{T-20} P(I) - \rho_{20} 1.07^{T-20} && (4.10) \end{aligned}$$

Now the model is easily formulated as a state space model. The state space model furthermore contains a description of how the measurements are related to the oxygen concentration – see Jacobsen and Madsen (1996) for details.

The parameters of the model are estimated using a program called CTLSM (see Madsen et al. (1990), Melgaard and Madsen (1993a), Madsen and Holst (1995)). This program is able to estimate embedded parameters in stochastic linear and non-linear differential equations. The program uses a maximum likelihood method where an extended Kalman filter is used in evaluating the Maximum Likelihood function (see Melgaard and Madsen (1993a), Jacobsen and Madsen (1996)).

Our implementation of the extended Kalman filter is quite different from that of Cosby and Hornberger (1984). In their approach the extended Kalman filter is used to obtain recursive estimates of the parameters by including the parameters in a state vector. Final parameter estimates are obtained by averaging the recursive estimates. Their approach requires a reasonable prior specification of noise covariances for the state variables. This is a rather difficult task and furthermore it is well-known from Ljung (1979), that divergence of the parameter estimates can be traced to miss-specification of the state noise covariance. Also, the method suggested by Cosby and Hornberger (1984) does not incorporate any temperature dependence of the parameters.

In our work we estimate the parameters directly and we avoid extending the state vector and the intermedium recursive estimates. Also the noise covariance is estimated along with the parameter estimates, hence the potential of mis-specification of noise is avoided. The extended Kalman filter is solely used to obtain the one-step predictions, which enters the likelihood function (Jacobsen and Madsen (1996)). Finally, the method is rather simple to use.

5 Results and discussion

Tables 1 and 2 show the estimated parameters, and their standard variation, for all the models and for unfiltered and filtered solar radiation, respectively. The variance of the simulation error is denoted by σ^2_{sim}.

Par.	E_o	P_m	ρ_{20}	κ_{20}	σ^2_{sim}	R^2
Unit	$\frac{mgO_2/l\,hr}{Ly/min}$	$mgO_2/l\,hr$	$mgO_2/l\,hr$	hr^{-1}		
1	$0.138\ 10^{-2}$	–	0.802	0.204	0.552	0.698
	$(0.403\ 10^{-4})$	–	(0.026)	$(0.758\ 10^{-2})$		
2	$0.761\ 10^{-1}$	–	1.450	0.297	0.207	0.887
	$(0.688\ 10^{-2})$	–	(0.035)	$(0.785\ 10^{-2})$		
3	$0.675\ 10^{-2}$	2.678	1.431	0.305	0.199	0.892
	$(0.329\ 10^{-3})$	(0.076)	(0.030)	$(0.695\ 10^{-2})$		
4	$0.458\ 10^{-2}$	2.045	1.393	0.303	0.206	0.888
	$(0.173\ 10^{-3})$	(0.049)	(0.031)	$(0.744\ 10^{-2})$		
5	$0.549\ 10^{-2}$	2.050	1.405	0.304	0.204	0.888
	$(0.209\ 10^{-3})$	(0.051)	(0.030)	$(0.726)\ 10^{-2})$		
6	$0.425\ 10^{-2}$	1.895	1.372	0.300	0.212	0.884
	$(0.155\ 10^{-3})$	(0.046)	(0.031)	$(0.783\ 10^{-2})$		
7	$0.339\ 10^{-2}$	8.314	1.244	0.283	0.256	0.348
	$(0.918\ 10^{-4})$	(0.183)	(0.030)	$(0.770\ 10^{-2})$		
8	$0.883\ 10^{-2}$	0.761	1.439	0.304	0.198	0.892
	$(0.580\ 10^{-3})$	(0.033)	(0.031)	$(0.736\ 10^{-2})$		

Table 1: Parameter estimates for unfiltered data. Standard deviations are shown in parentheses (). The exponent parameter α for model 2 was estimated 0.468 (0.013).

Cosby et al. (1984) found a reaeration coefficients in the range of 0.25 – 0.45 hr^{-1} This was for data from 1978, from five measuring stations with 12 weekly datasets from February to October. El-Shaarawi et al. (1988) found reaeration coefficients in the range of 0.25 – 0.33 for a Canadian river. This compares very well with our estimates.

Par.	E_o	P_m	ρ_{20}	κ_{20}	σ^2_{sim}	R^2
Unit	$\frac{mgO_2/l\,hr}{Ly/min}$	$mgO_2/l\,hr$		$mgO_2/l\,hr$	hr^{-1}	
1	$0.196\ 10^{-2}$	–	1.084	0.272	0.396	0.783
	$(0.310\ 10^{-4})$	–	(0.018)	$(0.526\ 10^{-2})$		
2	$0.563\ 10^{-1}$	–	1.560	0.332	0.189	0.897
	$(0.319\ 10^{-2})$	–	(0.019)	$(0.415\ 10^{-2})$		
3	$0.671\ 10^{-2}$	2.772	1.532	0.335	0.177	0.903
	$(0.181\ 10^{-3})$	(0.049)	(0.014)	$(0.526\ 10^{-2})$		
4	$0.475\ 10^{-2}$	2.016	1.500	0.336	0.179	0.902
	$(0.100\ 10^{-3})$	(0.027)	(0.017)	$(0.403\ 10^{-2})$		
5	$0.570\ 10^{-2}$	2.027	1.514	0.336	0.180	0.902
	$(0.130\ 10^{-3})$	(0.029)	(0.018)	$(0.404\ 10^{-2})$		
6	$0.446\ 10^{-2}$	1.847	1.486	0.335	0.183	0.900
	$(0.951\ 10^{-4})$	(0.024)	(0.018)	$(0.426\ 10^{-2})$		
7	$0.396\ 10^{-2}$	7.500	1.413	0.325	0.217	0.881
	$(0.622\ 10^{-4})$	(0.089)	(0.017)	$(0.424\ 10^{-2})$		
8	$0.822\ 10^{-2}$	0.852	1.541	0.336	0.178	0.902
	$(0.301\ 10^{-3})$	(0.024)	(0.018)	$(0.408\ 10^{-2})$		

Table 2: Parameter estimates for filtered data. Standard deviations are shown in parentheses (). The exponent parameter α for model 2 was estimated 0.525 (0.083).

In order to compare the models, we introduce the R^2 value as the ratio to the difference of the variance and the simulation error, and the variance

$$R^2 = \frac{V[Y] - [V[Y] - Y_{sim}]}{V[Y]} \tag{5.11}$$

where for the actual data, the variance $V[Y] = 1.8218$. It is seen in Tables 1 and 2 that the simulation performance for most of the models are nearly the same, with models 3, 4, 5 and 8 almost identical. Of these, model 8 is selected to show simulations and residuals.

Considering the individual parameters it must be recognised that E_0 and P_m have a different interpretation for the different models, but the values of ρ_{20} and κ_{20} can be compared for all models.

It is seen from Table 1 that, apart from the linear model 1 and to some extent model 7, the values of ρ_{20} and κ_{20} are the same if we take into account the uncertainty of the estimates. For the filtered case it is seen that the values of ρ_{20} and κ_{20} are slightly higher and the precision of the parameter estimates is much better. We contribute the difference in all the parameter estimates to the larger uncertainty, attributed to the unfiltered data.

In Figure 5 it is seen that several of the functions are almost indistinguishable. For both the filtered and unfiltered case the best (simulation) fit was observed for models 3 and 8. Model 3 is a Michaelis-Menten type model advocated by for instance Erlandsen and Thyssen (1983), Erlandsen and Sørensen (1984). The latter, however, then went on to simplify further to a linear model, which is seen here to give a significantly less adequate fit. Erlandsen and Thyssen (1983) finds the logarithmic function (model 8) to give the best overall fit, although closely followed by model 3. A Michaelis-Menten reaction (model 3) is an enzymatically controlled function, that assumes a constant enzymatic activity. As an approximation it captures the essence of a linear growth towards a saturation point.

A comparison of the filtered and unfiltered parameter estimates for model 3, along with the estimates from Pedersen (1994) is shown in Table 3. Standard deviations for their estimates are not available. Erlandsen and Thyssen (1983) do not report their parameter estimates, as that paper concentrated more on estimating the temperature dependency parameter from the respiration. Therefore the comparison is done for model 3 only. No comparison is done with the results from El-Shaarawi et al. (1988), as their study did not include model 3.

Parameter	E_o	P_m	ρ_{20}	κ_{20}
Unit	$\frac{mgO_2/l\,hr}{Ly/min}$	$mgO_2/l\,hr$	$mgO_2/l\,hr$	hr^{-1}
Unfiltered	$0.675\ 10^{-2}$	2.676	1.428	0.300
	$(0.330\ 10^{-3})$	(0.083)	(0.032)	$(0.758\ 10^{-2})$
Filtered	$0.671\ 10^{-2}$	2.016	1.536	0.336
	$(0.181\ 10^{-3})$	(0.052)	(0.018)	$(0.526\ 10^{-2})$
Pedersen	$0.539\ 10^{-2}$	3.606	1.464	0.320
(1994)	–	–	–	–

Table 3: Comparisons of model 3. Standard deviations are shown in parentheses (), but are not available for the prior estimated parameters in the photo-synthetic function.

Considering the estimated models in Figure 5 and the simulation performance for the various models, it is concluded that all the monotone and concave functions give a reasonable description of the photosynthesis-light relation.

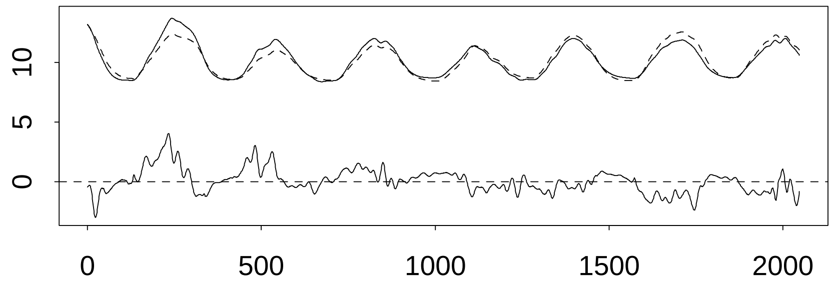


Figure 6: Time series of measured and simulated oxygen (in mg/l) plus prediction errors for model 8 (the logarithmic model).

5.1 Suggestions for improvements

The models investigated here, were very simple models, that did not take phenomena into account other than reaeration, photosynthesis and respiration. There may be other physical processes related to rivers that influences the oxygen content.

Benthic organisms may have a significant influence, as shown by Thyssen (1982), Whitehead (1980), Boyle and Scott (1984). The latter two papers reported an investigation from a small river in southern England, quite similar to Gryde river.

While light intensity may be the most important variable for the photosynthesis, there is also a dependency on carbon dioxide (CO_2), the water level and possibly the oxygen concentration. The CO_2 originates from the carbon system and is related to pH.

With respect to plants, there are two theories. Endogenous reactions in the plants involves consumption of CO_2. When the pools are full, there is, loosely speaking, no need to keep up the same rate. This results in a lower rate of photosynthesis in the afternoon. The other theory encompasses the fact that pH is higher in the afternoon as the temperature rises. This would also cause changes in the CO_2 levels, thereby changing the rate of photosynthesis in the afternoon. Documentation for these theories is scarce outside the laboratory and the data available for the present study does not make it possible to identify these differences.

It is clear, that in order to obtain a series free of disturbing high frequency noise, without destroying important information in the data, a low pass filter should be applied. Here an available low-pass filter was used. In further work the possibility of estimating an optimal low pass filter for the solar radiation will be considered.

Only a small set of new data were available for the present study. In order to obtain more prominent conclusions, more data sets from the same, and other, rivers should be investigated. For the present set of data, it is for instance not known, what the reason for the large oscillations during the first two days might be. A respiration function solely dependent on temperature may ignore the effect from the oxygen-concentration itself, inhibiting matter, plant types and light intensity (Simonsen (1974)).

6 Conclusion

Several photosynthetic models as a function of solar radiation data were compared. The investigation was performed by estimating the parameters in stochastic differential equations. Both unfiltered and low pass filtered solar radiation data were used as input, and it is concluded that

- The monotone and concave functions gave the best results and led to nearly the same estimates of the reaeration- and respiration temperature coefficients.
- A low pass filtering of the solar radiation data gave more precise parameter estimates.
- Investigations of more data sets is necessary to obtain a better understanding of the photo-synthetic relationship and whether other phenomena should be included in models for aquatic systems.

Acknowledgements

We wish to thank Assoc. Prof. M. Erlandsen of Institute for Biostatistics, Århus University, Denmark and Biologist T.M. Iversen, director of the National Environmental Research Institute (NERI), Silkeborg, Denmark, for providing the data and for advice and fruitful discussions. Also Prof. K. Sand of University of Copenhagen, Lab. for Fresh Water Biology, Hillerød, Denmark, is thanked for willingly sharing his knowledge of the biological system in small rivers.

Paper D

Estimation of a Low Pass Filter for Solar Radiation Data.

In press: Proceedings of the 9.th Conference of the European Consortium for Mathematics in Industry, ECMI (1997).

Abstract

In the paper a method is suggested for a simultaneous estimation of parameters in some differential equations, describing the variation of oxygen in a small creek, and the parameter in a low pass filtering of the solar radiation input. The low pass filtering is obtained by using a simple first order filter.

The oxygen dynamics are modelled by using linear and non-linear stochastic differential equations. The modelling involves a description of the influence from variations in water temperature and solar radiation. The latter affects the oxygen level via the photo synthesis process.

It is shown that the simultaneous low pass filtering of the solar radiation data provides more precise estimates. The low pass filter compensates for the fact that the solar radiation is a high frequency signal and a point measurement, which probably is measured some distance away from the main body of the river. The estimated filter enables an on-line monitoring of water quality. Thus, immediate assessment of receiving waters, as well as of data quality, is possible.

Key Words: Data filtration; Water quality; On-line monitoring; Parameter estimates.

2 Introduction

To describe the oxygen variation in an aquatic river system, like a river, the photosynthesis must be taken into account. This process is highly dependent on the solar radiation (for a thorough description of the oxygen dynamics in rivers (see Simonsen (1974), Thyssen (1982), Jacobsen and Voss (1994b), Jacobsen and Madsen (1996)). However, the solar radiation measurements are most frequently obtained as point measurements and often show high frequency variation.

In this paper a method is suggested for estimating the parameter of low pass filter along with the estimation of the parameters of a model for the oxygen dynamics. The simultaneous estimation is carried out by setting up a state space model for a set of stochastic differential equations describing both the oxygen dynamics and the low pass filter. For the oxygen dynamics, two models, a linear and a non-linear model, are investigated. Estimating the filter directly in a state space model enables the filter to be used on-line, which may be of use in monitoring river water quality.

3 The model and the filter

For a description of the formulation of the differential equation for the oxygen dynamics see Jacobsen and Voss (1994b), Jacobsen and Madsen (1996). The data used in this investigation is from a small creek in Denmark. Oxygen concentration, solar radiation, temperature, and the water level are sampled at five minute intervals for a period of one week. The collection of data is thoroughly described in Jacobsen et al. (1998a). and similar data from the river, is described in Thomsen and Thyssen (1979), Thyssen (1982) and Cosby et al. (1984).

3.1 The model for the oxygen dynamics

The differential equation for the oxygen dynamics is:

$$\frac{dC}{dt} = K(C_m - C) + P(I) - R \tag{3.1}$$

where C is the oxygen concentration, C_m is the saturated oxygen concentration, K the reaeration coefficient, and R, the respiration. The term $P(I)$ describes how the solar radiation, I, affects the oxygen concentration. The quantities K, R and the function $P(I)$ are temperature dependent as described in Jacobsen et al. (1998a).

Very often a simple linear relation is used for $P(I)$, i.e. $P(I) = E_o I$. However, in Jacobsen et al. (1998a) several possibilities for the function $P(I)$ were investigated and it was shown that monotone and concave functions gave the best results. Among those, the Michaelis-Menten type model yield good results:

$$P(I) = P_m \frac{1}{(P_m/E_o I) + 1} \tag{3.2}$$

which clearly is a non-linear model. The function describes the photosynthesis as a function of solar radiation to a saturation level.

It is well known that spectrum analysis can be used to describe how the variations of a signal is distributed on frequencies. An analysis has shown that for the considered sampling time the variations of the solar radiation contains a rather large amount of high frequency variation, and it is shown that using a low pass filtered solar radiation as input leads to a considerable improvement of models describing the dynamics of the oxygen concentration (Jacobsen et al. (1998a)). only the low frequency part of the variation correlates well with the variation of oxygen (via the photosynthetic process). The low pass filtering is possibly acting as a simple description of the persistence of the photosynthetic process of the plants.

The fact, that the measurements of the solar radiation are very often point measurements and measured some distance away from the main body in the river, might partly explain why the use of a low pass filtered solar radiation caused a considerable improvement in the description of the dynamics of the oxygen concentration.

The method used in Jacobsen et al. (1998a) solar radiation in the function $P(I)$ in eq. (3.2). Since a traditional low pass filter requires also future solar radiation measurements, the method can only be used in off-line estimates; but for monitoring purposes an on-line estimation is required. In the following we suggest a simple low pass filter which can be used for online estimation of the parameters, and hence for monitoring the states of the river.

3.2 The low pass filter

In order to enable a simultaneous estimation of the parameters, associated with the differential eq. (3.1) and the filter, the filter is formulated in continuous time by the following differential equation:

$$\frac{dI_f}{dt} = aI_f + bI \tag{3.3}$$

where I_f is the filtered solar radiation.

Some requirements for the coefficients (a,b) in eq. (3.3) are readily found by considering the transfer function from the measured solar radiation, I, to the filtered solar radiation, I_f. The demand of a stationary gain equal to one, specifies $b = -a$. Furthermore, low pass filtering, or stability of the system (3.3), is obtainable only for $a < 0$.

The observations are given in discrete time and the sampling time is $\tau = 5$ min. Using the requirement $b = -a$, the discrete time version of eq. (3.3) is obtained by integration through the sample interval $[t, t+\tau]$:

$$I_f(t+\tau) = \phi(\tau)\ I_f(t) + (1-\phi(\tau))I(t) \tag{3.4}$$

where $\phi(\tau) = e^{a\tau}$. Since $a < 0$ (and $\tau > 0$) we have that $0 < \phi(\tau) < 1$.

No filtering is obtained for $a \to -\infty$, and in this case $I_f(t+\tau) = I(t)$. Hence, in order to avoid a time lag in the differential equation for the oxygen equation, $I_f(t+\tau)$ is used as input variable. This means that the "on-line" calculations actually are $\tau = 5$ min. behind the actual measurements.

3.3 The combined model and filter

The parameter, a, of the filter is estimated by including a state variable, that describes the filtered solar radiation. Finally, stochastic terms are added to the deterministic differential equations. For the linear model, $P(I_f) = E_o I_f$ the stochastic differential equations then can be written as the state space model:

$$\begin{aligned} \begin{bmatrix} dC \\ dI_f \end{bmatrix} &= \begin{bmatrix} -K(T) & E_0(T) \\ 0 & a \end{bmatrix} \begin{bmatrix} C \\ I_f \end{bmatrix} dt \\ &+ \begin{bmatrix} -R(T) & K(T) & 0 \\ 0 & 0 & -a \end{bmatrix} \begin{bmatrix} 1 \\ C_m \\ I \end{bmatrix} dt + \begin{bmatrix} dw \\ 0 \end{bmatrix} \end{aligned} \tag{3.5}$$

where a < 0 and dw is a stochastic term describing the deviation between the true oxygen concentration and the oxygen concentration as described by the model (3.1).

The observed oxygen concentration is:

$$Y(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} C(t) \\ I_f(t) \end{bmatrix} + e(t) \tag{3.6}$$

.where $e(t)$ is the measurement error. The state space model may be written in matrix notation:

$$d\boldsymbol{X} = \boldsymbol{A}\boldsymbol{X}dt + \boldsymbol{B}\boldsymbol{U}dt + d\boldsymbol{w}(t) \tag{3.7}$$

$$\boldsymbol{Y}(t) = \boldsymbol{C}\boldsymbol{X}(t) + \boldsymbol{e}(t) \tag{3.8}$$

where $\boldsymbol{X}$ is the state vector containing the state of oxygen and the filter. The input vector $\boldsymbol{U}$ contains the measured solar radiation and the other inputs. The matrix $\boldsymbol{A}$ characterises the dynamical behaviour of the system and $\boldsymbol{B}$ is a matrix which specify how the input signals enter the system. In order to be able to find maximum likelihood estimates of the parameters based on Gaussian distributions, the process $\boldsymbol{w}(t)$ is restricted to be a Wiener-process. The measurement error $\boldsymbol{e}(t)$ is assumed to be Gaussian distributed white noise with zero mean. Furthermore, it is assumed that $\boldsymbol{w}(t)$ and $\boldsymbol{e}(t)$ are mutually independent.

A better (closer to an ideal low pass) filter is expected if we consider a second or higher order filter. This would involve three or more coupled differential equations as this is comprised of two or more first order filters in series.

In the next section it is briefly outlined how the parameters of the model (state space model (3.5)), including the parameter in the low pass filter, are estimated by a maximum likelihood method.

4 Maximum likelihood estimates

All observations are equidistantly spaced at five minute intervals. However, in order to simplify the notation, we shall assume that the time index t belongs to the set $\{0, 1, 2, ..., N\}$, where N is the number of observations. Introducing

$$\mathcal{Y}(t) = [\boldsymbol{Y}(t), \boldsymbol{Y}(t-1), \ldots, \boldsymbol{Y}(1), \boldsymbol{Y}(0)]' \tag{4.9}$$

i.e. $\mathcal{Y}(t)$ is a vector containing all the observations up to and including time t.

All the unknown parameters, denoted by the vector $\boldsymbol{\theta}$, are embedded in the continuous time state space model (eq.s (3.7) and (3.8)). The observations are, however, given in discrete time. Hence, the state space model has to be sampled in order to calculate the likelihood function. Under the assumption that the input variables are constant through the sampling time, the discrete time state space model can be written (Madsen and Melgaard (1991), Melgaard and Madsen (1993a), Jacobsen and Madsen (1996)):

$$\begin{aligned} \boldsymbol{X}(t+1) &= \boldsymbol{\Phi}\, \boldsymbol{X}(t) + \boldsymbol{\Gamma}\, \boldsymbol{U}(t) + \boldsymbol{v}(t) && (4.10) \\ \boldsymbol{Y}(t) &= \boldsymbol{C}\, \boldsymbol{X}(t) + \boldsymbol{e}(t) && (4.11) \end{aligned}$$

where $\boldsymbol{\Phi} = e^{\boldsymbol{A}\tau}, \boldsymbol{\Gamma} = \int_0^\tau e^{\boldsymbol{A}s}\boldsymbol{B}ds$ and $\boldsymbol{v}(t)$ is a white noise sequence.

The likelihood function is the joint probability density of all the observations assuming that the parameters are known, i.e.

$$\begin{aligned} L'(\boldsymbol{\theta}; \mathcal{Y}(N)) &= p(\mathcal{Y}(N)|\boldsymbol{\theta}) \\ &= p(\boldsymbol{Y}(N)|\mathcal{Y}(N-1), \boldsymbol{\theta})p(\mathcal{Y}(N-1)|\boldsymbol{\theta}) \\ &= \left(\prod_{t=1}^{N} p(\boldsymbol{Y}(t)|\mathcal{Y}(t-1), \boldsymbol{\theta})\right)p(\boldsymbol{Y}(0)|\boldsymbol{\theta}) && (4.12) \end{aligned}$$

where successive applications of the rule $P(A \cap B) = P(A|B)P(B)$ is used to express the likelihood function as a product of conditional densities.

Since both the model error and the measurement error are normally distributed the conditional density is also normal. The normal distribution is completely characterized by the mean and the variance. Hence, in order to parameterize the conditional distribution, we introduce the conditional mean and the conditional variance as

$$\hat{\boldsymbol{Y}}(t|t-1) = E[\boldsymbol{Y}(t)|\mathcal{Y}(t-1), \boldsymbol{\theta}] \tag{4.13}$$

$$\boldsymbol{R}(t|t-1) = V[\boldsymbol{Y}(t)|\mathcal{Y}(t-1), \boldsymbol{\theta}] \tag{4.14}$$

respectively. It may be noticed that these corresponds to the one-step prediction and the associated variance, respectively. Furthermore, it is convenient to introduce the one-step prediction error (or innovation)

$$\boldsymbol{\epsilon}(t) = \boldsymbol{Y}(t) - \hat{\boldsymbol{Y}}(t|t-1) \tag{4.15}$$

Using (4.12) – (4.15) the conditional likelihood function (conditioned on $\boldsymbol{Y}(0)$) becomes

$$\begin{aligned} L(\boldsymbol{\theta}; \mathcal{Y}(N)) = \prod_{t=1}^{N} \Big(&(2\pi)^{-m/2} \det \boldsymbol{R}(t|t-1)^{-1/2} \\ &\exp(-\tfrac{1}{2}\boldsymbol{\epsilon}(t)' \boldsymbol{R}(t|t-1)^{-1} \boldsymbol{\epsilon}(t)) \Big) \end{aligned} \tag{4.16}$$

where m is the dimension of the $\boldsymbol{Y}$ vector (in the present case m= 1). Traditionally the logarithm of the conditional likelihood function is considered

$$\begin{aligned} \log L(\boldsymbol{\theta}; \mathcal{Y}(N)) = &-\tfrac{1}{2} \sum_{t=1}^{N} \big(\log \det \boldsymbol{R}(t|t-1) + \boldsymbol{\epsilon}(t)' \boldsymbol{R}(t|t-1)^{-1} \boldsymbol{\epsilon}(t) \big) \\ &+ \text{const} \end{aligned} \tag{4.17}$$

In the linear case, the conditional mean and variance eq. (4.14) can be calculated recursively by using a Kalman filter. In the non-linear case, an extended Kalman filter is used for calculating the conditional mean and variance. The numerical details of these algorithms can be found in Madsen and Melgaard (1991).

The maximum likelihood estimates are found by maximization of the log likelihood function (4.17), and the uncertainties of the parameter estimates are found using the observed curvature of the log likelihood function, evaluated at the final estimates (Madsen and Melgaard (1991), Melgaard and Madsen (1993a)).

5 Results and discussion

Tables 1 and 2 show the obtained maximum likelihood estimates for the parameters, and their standard variations, for the linear and the non-linear case, respectively. The un- and pre-filtered estimates were obtained previously (see Jacobsen et al. (1998a)) low pass filter.

Linear model:

Par.	Unit	Un-filtered	Pre-filtered	On-line filter
E_o	$\frac{mgO_2/l\,hr}{Ly/min}$	$0.179\ 10^{-2}$ $(0.363\ 10^{-4})$	$0.196\ 10^{-2}$ $(0.305\ 10^{-4})$	$0.178\ 10^{-2}$ $(0.356\ 10^{-4})$
R	$mgO_2/l\,hr$	1.000 (0.022)	1.078 (0.018)	1.050 (0.024)
K	hr^{-1}	0.248 $(0.670\ 10^{-2})$	0.269 $(0.495\ 10^{-2})$	0.259 $(0.682\ 10^{-2})$
a	hr^{-1}	–	–	-16.208 (1.828)
σ^2	–	$0.700\ 10^{-3}$	$0.306\ 10^{-3}$	$0.656\ 10^{-3}$

Table 1: Estimated parameters for the linear model. Standard deviations are shown in parentheses and σ^2 is the variance of the prediction error.

Non-linear model:

Par.	Unit	Un-filtered	Pre-filtered	On-line filter
E_o	$\frac{mgO_2/l\,hr}{Ly/min}$	$0.467\ 10^{-2}$ $(0.167\ 10^{-3})$	$0.667\ 10^{-2}$ $(0.165\ 10^{-3})$	$0.545\ 10^{-2}$ $(0.232\ 10^{-3})$
P_m	$mgO_2/l\,hr$	3.503 (0.128)	2.727 (0.042)	3.173 (0.096)
R	$mgO_2/l\,hr$	1.377 (0.025)	1.509 (0.016)	1.451 (0.027)
K	hr^{-1}	0.307 $(0.615\ 10^{-2})$	0.328 $(0.384\ 10^{-2})$	0.319 $(0.611\ 10^{-2})$
a	hr^{-1}	–	–	-39.088 (4.096)
σ^2	–	$0.514\ 10^{-3}$	$0.155\ 10^{-3}$	$0.501\ 10^{-3}$

Table 2: Estimated parameters for the non-linear model. Standard deviations are shown in parentheses and σ^2 is the variance of the prediction error.

First of all, it is noted that all parameter estimates for the on-line filter lie between the values obtained by using the raw solar radiation data and values obtained by pre-filtering the solar radiation data. This means that the very simple on-line filter clearly gives better estimates than those obtained by using the raw (un-filtered) solar radiation data. In both cases the variance of the prediction error is only slightly smaller than those found when not filtering at all.

As previously pointed out in the litterature (Jacobsen et al. (1998a)), seen between the estimated parameters using the linear and the non-linear functions to describe the influence of the solar radiation. Considering the variance of the one-step prediction error, it is clearly seen that the non-linear model gives a much better description.

As pointed out in Jacobsen et al. (1998a) is a very simple model, that only takes the very basic physical and biological processes into account. Including other physical and biological processes may improve the total model (Jacobsen et al. (1996)).

6 Conclusion

A method for simultaneously estimating the parameters of a simple model for the oxygen dynamics and the parameter of a first order low pass filter of the solar radiation input is suggested. The approach uses a stochastic state space model formulation of the combined model and filter. A maximum likelihood method for estimating the parameters of the combined model is used. In the linear case, an ordinary Kalman filter, and in the non-linear case an extended Kalman filter, was used to evaluate the likelihood function.

The estimated filter was not as effective as pre-filtering the solar radiation data. However, pre-filtering can only be performed off-line, whereas the method proposed in this paper provides the possibility for on-line monitoring.

The possibility for on-line monitoring is a great advantage at a time where industry is being asked to monitor their waste water and the effect it has on the water quality in the receiving waters. Also, it is very useful to monitor data collection for these purposes. Enormous investments are placed in monitoring programs and the measuring equipment is more and more sophisticated. It is prudent that these investments are looked after and protected.

On-line measurements are often measured with small sampling intervals, which then provides very high frequency data. A part of the high frequency variation might be a result of stochastic fluctuations, not necessarily representative of the dynamics in the system. Thus, filtering the data directly may result in a better monitoring program, as demonstrated in this paper.

Paper E

A Method for Automatic Calibration of Parameters for an Integrated Model.

Submitted to: Water Resources Research (1997).

Abstract

A method for automatic calibration of parameters of an integrated deterministic model for urban water quality is suggested for a synthetic catchment and drainage system. A perfect structure of the deterministic model is assumed. This has the advantage that the model structure does not contribute to possible deviations in the tests as to whether the calibration procedure does indeed find the correct parameters.

A traditional statistic method, the so-called output-error method, is used to calibrate system parameters for both runoff and receiving water quality. The output error is the difference between a measured time series and a simulated output. By minimizing some strictly convex function of this difference it is possible to find the parameters, which gives the best fit of the model to the measured data. This is provided the model structure is well behaved, i.e. that the parameters are structurally identifiable. A Monte Carlo validation of the calibration routine is carried out. It is shown that even with data contaminated by a quite high level of noise, the calibration method yield results, that are not significantly different from the correct values.

Key Words: Automatic calibration; Integrated urban river modelling; Output-error method; Water quality.

1 Introduction

Working with deterministic models always includes a long and often tedious phase of calibrating the parameters of the model. An experienced operator can reduce the time spent on this task considerably, but still the time allocated to this problem is significant. The calibration task is usually solved by heuristic procedures, but it is essentially an optimization problem. In this paper we suggest an automatic calibration method based on the so-called output error method for estimating parameters in dynamic systems.

Water quality in a river is most often evaluated by the concentration of dissolved oxygen. In fact, it is the most used universal environmental factor, as this quantity determines the trophic level a river is able to support. In Denmark, the Danish Water Pollution Control Committee has set up some water quality criteria, based on the return period of dissolved oxygen concentrations. Traditionally, heavy rain events will cause discharge (from overflow structures) of oxygen depleting contaminants (such as organic matter, expressed as COD) into urban rivers from combined sewer systems (Jacobsen et al. (1996), Rauch and Harremoës (1996)).

The output-error method is a well known and very simple principle for estimating parameters in dynamic systems (Söderström and Stoica (1989), Söderström (1994)). Applied to a deterministic model, several parameters may be calibrated automatically using a measured time series. However, if too many parameters are included, or if the information content in the data is not sufficient, the system may not be structurally identifiable (Ljung and Glad, 1994).

1.1 The integrated model

The integrated model used in this investigation, is a combination of a simplified model for urban runoff with pollution, and a model for oxygen depletion in rivers. In both models the traditional hydrograph method (Rauch and Harremoës, 1996) of routing the water volume from element to element is used.

The oxygen dynamics in the river is modelled by empirical functions for the various sources and sinks of oxygen. These are oxygen production by photosynthesis, depletion by respiration, and reaeration from the exchange of oxygen between the air and water. Both immediate and delayed oxygen consumption are taken into account. This model is written as (see e.g. Odum (1956), Dobbins (1964)):

$$\frac{dC}{dt} = K_2(C_m - C) + P(I) - R + S \tag{1.1}$$

where C is the oxygen concentration, K_2 the reaeration coefficient, C_m the saturated oxygen concentration and R is the respiration. The term S, denotes the remaining sources and sinks from degradation of organic matter, ammonia oxidation, sedimentation, etc. P(I) is a function describing how the photosynthesis depends on solar radiation, I. Various models for P(I) are investigated in Jacobsen et al. (1998b) and most of these models use the maximum photosynthetic rate, P_{max}, as a parameter.

More details of this model may be found in Rauch and Harremoës (1996) and Jacobsen et al. (1996). Annual and diurnal fluctuations in temperature, wastewater and streamflow, which are typical for Denmark, are built into the model. A model for the wastewater treatment plant could be incorporated if a complete integrated model of the urban waste water systems is wanted (see Rauch and Harremoës (1995), Rauch and Harremoës (1996)).

This integrated model thus consist of a module for the catchment runoff and a module for the river. For the runoff model, measured overflow from a combined sewer overflow structure (CSO) can be used to calibrate the runoff coefficient, and other pertinent parameters from the catchment. The automatic calibration procedure proposed in this paper has been used to calibrate the loss from infiltration, along with the runoff coefficient. For the model (1.1) describing the variation of oxygen in the river, the reaeration coefficient, K_2, the maximum photosynthetic rate, P_{max}, and the respiration, R, may be calibrated. K_2 and R are temperature dependent and a parameterization is chosen which leads to calibrated values which are valid at 20 ^oC.

The above mentioned combination of models is just an example. The calibration procedure, can use any model or any combination of models.

1.2 The system

The system, upon which this calibration procedure has been tested, is a small part of a much larger system involving an urban river in the greater Copenhagen area. It consists of three catchments, each with an overflow structure. These are connected by overflows to the other part of the system, i.e. the river. Each catchment (SP, BR and DM) is characterised by its size, number of people and waste water flow. A catchment for the river represents the upstream conditions. The total system is shown in Figure 1.

The river is divided into sections, into some of which waste water is discharged. In this study the catchment runoff and the river water quality is the primary concern. What happens to the water upon leaving the system is not considered.

The whole system is set up as individual elements, which are connected so that the calculations proceed from element to element using the standard hydrograph method of moving masses of water, which contain oxygen and

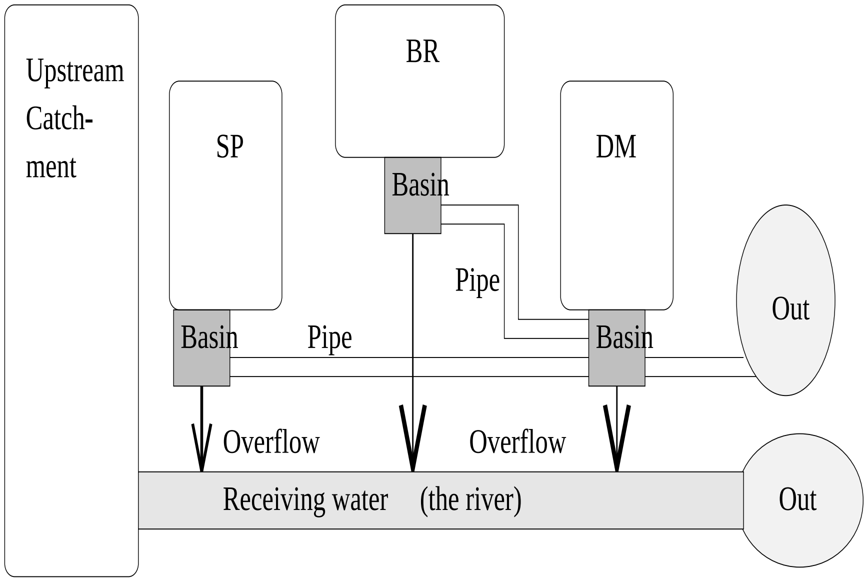


Figure 1: A schematical figure of the considered system. The arrows show overflow from the basins.

contaminants. This includes dissolved and particulate COD, ammonia and nitrate. The pollutant concentrations used in dry weather flow and runoff during rain are seen in Table 1.

For the simulation three rain events over a two day period are used. In this specific system these rain events cause overflow from the DM catchment. Thus depletion of oxygen from COD (from that overflow) occurs downstream from the overflow point.

2 The calibration method

The calibration method is based on the so-called *output error method* (OEM) for estimating parameters in dynamic systems (Söderström and Stoica, 1989). In the single-input-single-output (SISO) case the OEM is related to a model of the form

$$y(t) = f(u(t), u(t-1), \ldots; \theta) \tag{2.2}$$

Parameter	Dry weather	Rain	Unit
COD sol.	90.0	25.0	[mg/l]
COD par.	250.0	70.0	[mg/l]
N-total	45.0	10.0	[mg/l]
$NH_4 + NH_3$	25.0	4.4	[mg/l]

Table 1: Mean pollution concentrations in the integrated model. Dry denotes the concentrations in dry weather flow (only waste water undiluted by rain). COD is separated into an easily biodegradable soluable part and a slower biodegradable particular fraction.

where θ is the unknown parameters and f is a known function of the actual and previous input values. That is, the output $y(t)$ of the model is described as a sum of a function of historical input values and a noise (or error) term $v(t)$. It is noticed, that the model contains no description of the noise term. The above model can, off course, be extended to the multiple input and multiple output case.

Most statistical methods for estimating parameters in dynamic models, such as the Maximum Likelihood method (ML), and the Prediction Error Method (PEM) (Ljung, 1987), require a description – or parameterization – of the noise term; but this is not needed for the output error method.

The most common choice of criterion for defining the best set of parameters is the sample covariance of the simulation error, i.e.

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}(y(t) - f(u(t), u(t-1), \ldots; \theta))^2 \tag{2.3}$$

where N is the length of the measured time series.

In the multi-variable case some scalar-valued function of the vector simulation error is considered. For example

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}||y(t) - f(u(t), u(t-1), \ldots; \theta)||^2 \tag{2.4}$$

but other alternatives exists as well.

The output-error estimate of θ is now defined as

$$\hat{\theta} = \arg\min_{\theta} V_N(\theta) \tag{2.5}$$

It is assumed that the model is structural identifiable (Ljung and Glad, 1994), i.e. the structure of the model allows for an estimation of the unknown parameters given measured and persistently excited time series of the input and output.

For conventional statistical methods for parameter estimation, like ML and PEM, it is possible to use statistical tests to validate the model and it is rather easy to obtain an estimate of the covariance matrix of the parameter estimates. For the output error method it is in general not possible to use statistical methods for testing the model, and the covariance of the parameter estimates takes in general a quite complicated form. This makes it hard or impossible to derive confidence intervals based on statistical theory.

For large data sets the covariance of the output error estimates takes the form (Söderström, 1994):

$$V[\hat{\theta}] = \frac{1}{N} V''(\theta_0)^{-1} R V''(\theta_0)^{-1} \tag{2.6}$$

where $V''(\theta_0)$ is the Hessian matrix of the criterion. The matrix R depends in a complicated way on the dynamics and on the properties of the disturbances. Only in the case when the disturbances, expressed as output errors, are uncorrelated in time the expression will be practical attractable; but this is normally not the case.

The criterion ((2.3) or (2.4)) is in general a rather complicated function of the unknown parameters, and in the present case, where the model is an integrated model for urban water quality, this is indeed the case. Therefor the criterion has to be minimized using numerical methods. In the present case a quasi-Newton method has been used (see Madsen and Tingleff (1990), Frandsen et al. (1995)).

3 Case results

For the purpose of testing the automated calibration procedure, a series was generated using the previously described model. The parameters are shown in Table 2. Thus the parameters and the model structure are known and it could be investigated whether the procedure is able to find the correct parameter values.

Parameter	Loss	Run_{co}	K_2	Pmax	R
value	2.00	0.75	3.0	7.5	6.0
unit	[mm/d]	[–]	[d^{-1}]	[mg/ld]	[mg/ld]

Table 2: Parameters used to generate the time series. Loss refer to permanent losses from infiltration per day, while Run_{co} is the runoff coefficient. The reaeration is denoted K_2, the maximum photosynthetic rate Pmax, and R is the respiration.

The calibration was done in two steps, since it is advantageous to use the causality between the two systems. First the runoff from the watershed was calibrated to get the runoff hydraulics correct. Afterwards, the receiving water model was calibrated based on the overflow from the calibrated catchment. For the system used here, it was assumed that only two measured time series were available; namely the overflow from the DM catchment and the oxygen content in the river, somewhere further downstream from that overflow.

Since only one time series (from one catchment) was known, it was assumed that the catchments were alike in respect to the runoff coefficient and the loss due to infiltration. These two parameters are those calibrated for the runoff part of the model. For the river part of the model, three parameters were calibrated, namely the reaeration coefficient, K_2, the maximum photosynthetic rate, P_{max}, and the respiration, R.

Several combinations of initial values were tested. In all cases the method gives the correct values. Tables 3 and 4 show the values for each run and the number of iterations and gradient calls in each case.

Parameter	Unit	Initial values				true value
Loss	[mm]	1.25	1.25	4.25	4.25	2.00
Run_{co}	[–]	0.55	0.95	0.95	0.55	0.75
Its/grds	[–]	8/12	12/15	12/21	9/14	–

Table 3: Initial values and the number of iterations (Its) and gradient calls (grds) for calibration of the *runoff part* of the model.

For the first and third combination of the initial values, Figure 2 shows each step towards the final result for the two parameters calibrated for the runoff part of the model. It is seen that the runoff coefficient is quickly found, and hereafter the search is concentrating on the coefficient for the

Parameter	Unit	Initial values				true value
K_2	$[d^{-1}]$	1.50	5.75	1.50	0.70	3.00
Pmax	[mg/ld]	5.50	9.85	12.50	14.50	7.50
R	[mg/ld]	3.00	8.25	4.30	12.30	6.00
Its/grds	[–]	13/19	12/18	17/27	22/28	–

Table 4: Initial values and the number of iterations (Its) and gradient calls (grds) for calibration of the *river part* of the model.

loss from infiltration. Although this is automatically done by the software, this complies with general practice, where the runoff coefficient traditionally is estimated first as this parameter determines the amount of water in the hydrograph. Then other parameters, describing the shape of the hydrograph, are calibrated. A similar plot is shown for the three parameters of the river model in Figure 3.

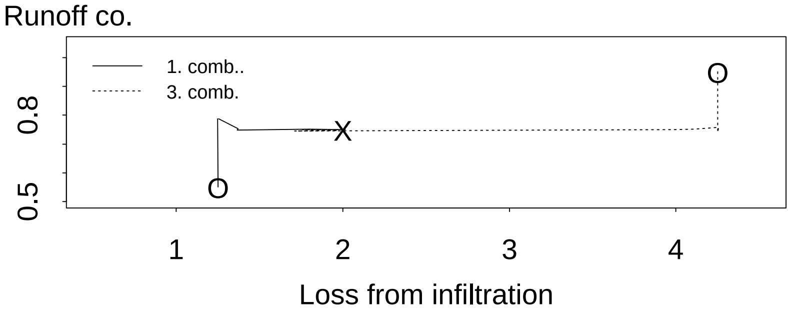


Figure 2: The iteration steps of the automatic calibration method towards the correct solution for the *runoff model.* Capital O shows the initial values of the two combinations, (1.25,0.55), and (4.25,0.95), while X shows the final estimate, which is (2.00,0.75) in both cases.

Figure 4 shows the measured and simulated time series for both the runoff and the oxygen concentration. The simulation was done after the calibration was performed with initial values quite far from the known values. The figure also shows the difference and it is noticed that the remaining difference is very small. This difference is due to numerical dispersion.

In the deterministic model for oxygen depletion a forcing function was used to predict the oxygen production from photosynthesis. If a measured series of solar radiation data were available, the description of the oxygen dynamics could have been made more realistic by using a model based

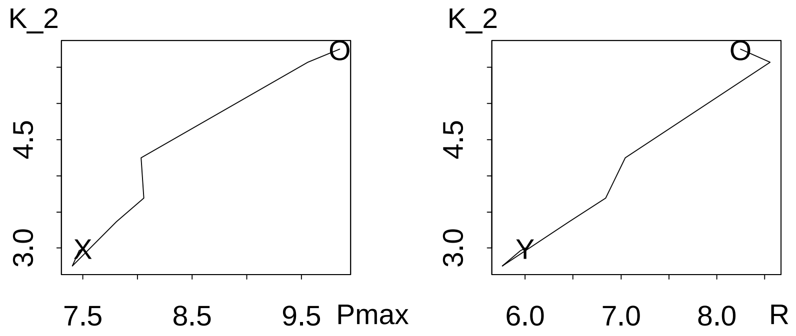


Figure 3: The iteration steps of the automatic calibration method towards the correct solution for the *river model*. Capital O shows the initial values of one combination, (9.85,5.75) and (8.25,5.75), while X and Y show the final estimates (7.50,3.00) and (6.00,3.00), respectively.

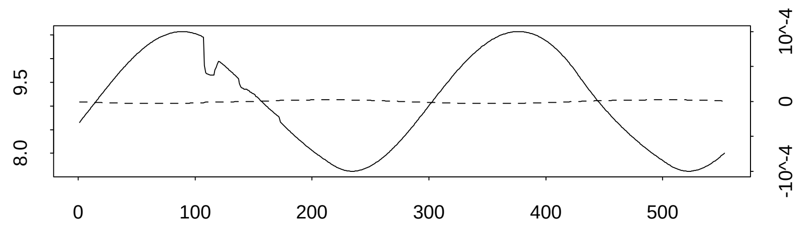


Figure 4: The measured and simulated time series of the oxygen concentration $[mg/l]$, (solid line) four to six km. downstream of the overflow. Note that it is not possible to distinguish between the two curves. The dotted line corresponding to the right hand y-axis show the difference between the measured and the simulated time series. The x-axis is 5 minute timesteps.

on the relationship between solar radiation and photosynthesis. This is described in detail in Jacobsen et al. (1998b) and Jacobsen et al. (1997a).

3.1 Validation

In practice measurements are encumbered with measurement noise. In order to validate the calibration method in the case of noisy measurements, a Monte Carlo approach is used. First, 50 stochastic independent realizations (noisy time series) were formed by using the model with known parameters and adding normally distributed noise with zero mean and a standard deviation of 0.5 mg/l. Then the parameter values were identified from each of the noisy time series by using the suggested calibration routine with otherwise identical initial conditions.

Table 5 shows the results from this Monte Carlo approach. For each of the parameters, the average and the variance ($\bar{\hat{\theta}}_i$ and $s^2_{\hat{\theta}_i}$) of the 50 calibrated parameters using the suggested procedure to the 50 noisy time series, are given. These can then be compared to the true values, θ_i, and in particular a test quantity, $|Z_i|$ (see Eq. (3.7)) can be used to test whether the automatic calibration procedure gives correct values in the case of noisy measurements.

Parameter	Unit	θ_i	$\bar{\hat{\theta}}_i$	$s^2_{\hat{\theta}_i}$	$\|Z_i\|$
K_2	$[d^{-1}]$	3.00	3.013	$2.75 \cdot 10^{-2}$	0.554
P_{max}	[mg/ld]	7.50	7.460	$2.62 \cdot 10^{-2}$	1.747
R	[mg/ld]	6.00	5.974	$2.36 \cdot 10^{-2}$	1.197

Table 5: Results from calibrations using 50 noisy time series. θ_i denotes the true values, $\bar{\hat{\theta}}_i$ the average of the 50 calibration results, while $s^2_{\hat{\theta}_i}$ is the variance of the results.

The stochastic independent simulations makes it possible to use statistical methods to verify that the calibrated parameters are not significantly different from the true parameters. The following hypothesis is tested for all parameters:

$$H_0: \quad \bar{\hat{\theta}}_i = \theta_i \qquad \text{vs.} \qquad H_1: \quad \bar{\hat{\theta}}_i \neq \theta_i \tag{3.7}$$

The test statistics for the hypothesis (3.7) is:

$$Z_i = \frac{\bar{\hat{\theta}}_i - \theta_i}{S_{\bar{\hat{\theta}}_i} / \sqrt{n}} \tag{3.8}$$

where $n = 50$. Under H_0 the test statistics is student-t distributed, with $n - 1$ degrees of freedom. When testing the hypothesis on the significance level $\alpha = 5$ % the critical area is $\{|Z_i| > 2.00\}$. As none of the test-values Z_i exceeds this value (see Table 5) we cannot reject H_0 for any of the parameters on the chosen level. Hence, it is concluded that the method is able to provide reasonable calibrations also in the case of noisy measurements.

4 Conclusions

A method for automatic calibration of parameters of a simulation model is suggested. The model considered is an integrated model for urban water quality, which consists of two parts describing the urban catchment runoff and the depletion of oxygen in rivers, respectively. The calibration is based on measurements of the input variables, i.e. the overflow from one of the overflow structures, and the output variable, namely the oxygen in the river.

The integrated model contains only five unknown parameters, which were calibrated automatically by the method. The method can, however, be used for more complicated systems were more traditional and heuristic procedures for calibration are difficult or infeasible.

The suggested method for automatic calibration of the parameters is based on the output-error method (OEM). In this method the parameters are found as those minimizing the sum of squared differences between the measured output values and the corresponding simulations. The advantage of this method is that it can be used as soon as it is possible to simulate the output for given input variables, provided that the model is structural identifiable. The most serious drawback of the method, compared to ordinary statistical methods for parameter estimation, is that the covariance of the calibrated parameter values takes in general a quite complicated form, which makes it hard or impossible to derive confidence intervals and to use tests based on statistical theory.

In order to validate the method a simple deterministic system with known parameters is considered, and it is shown that the automatic calibration method finds the true parameter values. Furthermore fifty stochastic independent series of noise contaminated measurements of oxygen are used for validation, and it is shown that also in the case of noisy measurements the method provides reasonable calibration values.

Paper F

A Stochastic Model for Two-station Hydraulics Exibiting Transient Impact.

Submitted to: The Watermatex conference (1997).

Abstract

The objective of the paper is to interpret data on water level variation in a river affected by overflow from a sewer system during rain. The simplest possible, hydraulic description is combined with stochastic methods for data analysis and model parameter estimation. This combination of deterministic and stochastic interpretation is called grey box modelling.

As deterministic description the linear reservoir approximation is used. A series of linear reservoirs in sufficient number will approximate a plug flow reactor. The choice of number is an empirical description of the longitudinal dispersion in the river. This approximation is expected to be a sufficiently good approximation as a tool for the ultimate aim: the description of pollutants transport in the river.

The grey box modelling is a statistical tool for estimation of the parameters in the deterministic model. The advantage is that the parameters have physical meaning, as opposed to many other statistically estimated, empirical parameters. The identifiability of each parameter, the uncertainty of the parameter estimation and the overall uncertainty of the simulation is determined.

Keywords: Grey box modelling; Maximum likelihood estimation; Stochastic hydraulic model.

2 Introduction

It is well known from models of reactor hydraulics, usually pertaining to waste water treatment plants, that modelling the reactor as a series of totally mixed reactors will mathematically merge towards a description for plug flow when the number of reactors in the series goes towards infinity. In hydrology this is called a linear reservoir model (see e.g. Chow et al. (1988)).

This description is proposed for a simple hydraulic model of an urban river, influenced by transient impact during rain events. Whitehead and Young (1975) used a similar concept in a deterministic streamflow forecasting model, where the variation in flow was modelled by a set of first order differential equations and combined with a stochastic black box model for rainfall runoff. This essentially comprises what we call a grey box model.

Due to uncertainty in the model, as well as uncertainty in the measurements, the stochastic variation must be included as an integrated part of the model as proposed in this paper. Furthermore Young and Beck (1974), Beck and Young (1975a) also used a linear reservoir type model with one reservoir and a transportation delay, while for the approach considered in this paper a series of reservoirs are used.

In this paper a grey box modelling approach of a two-station hydraulics is outlined. For simplicity a very simple system with only three reservoirs is considered. However, this simple model turns out to give a rather good description of the dynamics of the hydraulics under the influence from rain events. Furthermore, the number of reservoirs is supported by statistical analyses.

In the last section it is demonstrated that the modelling approach makes it very easy to suggest extensions of the model; for instance to cope with rivers, where the characteristics of the hydraulics change along the river stretch.

2.1 A simple two-station hydraulic model

For practical purposes an infinite number of reservoirs is not feasible. An adequate number must be determined by testing the model on real river data. The data used in this investigation has been sampled for about a week with a resolution of fifteen minute intervals from an urban river, which is influenced by transient impact from rain events. Water levels have been measured at two measuring stations about a kilometer and a half apart and rain has been measured a short distance upstream from these.

For each reservoir, the linearized change in volume over time is:

$$\frac{dV}{dt} = Q_{in} - Q_{out} + Q_{add} \tag{2.1}$$

where Q_{in} and Q_{out} is the input and output water flow to the reservoir, respectively. The state variable V is the volume and t is the time. Q_{add}, is a term which describes the transient impact from rain events. For simplicity, we assume that $Q_{add} = A_c \; \Phi \; P_r$, where A_c is the catchment area, Φ the runoff coefficient and P_r is the rain used as a proportional factor. It is assumed that the additional water flow is proportional to the intensity of the rain. A system with three reservoirs is shown in Figure 1.

The description of the series of linear reservoirs, depicted in Figure 1 involves a resistance, R_h, which is a function of the specific characteristics of the river, such as slope, roughness and the geometry. The term Q_{add} includes direct overflow from the sewer system, when the sewers and the storage basins cannot keep up with the combined amount of water from runoff and regular sewage, as well as increased masses of water being discharged to the recipient from a waste water treatment plant. Thus, rain data are used as a proportionality factor. Assuming that

$$Q_{in} \;=\; \frac{A\,h_{i-1}}{R_h} \qquad \text{and} \qquad Q_{out} \;=\; \frac{A\,h_i}{R_h} \tag{2.2}$$

We obtain the following equation for one reservoir:

$$\frac{dh_i}{dt} \;=\; \frac{h_{i-1}-h_i}{R_h} + \frac{A_c\;\Phi}{A}P_r \tag{2.3}$$

R_h is the resistance against water flow between the reservoirs, A is the surface area of the stretch of river, and P_r is the proportional factor from the intensity of rain.

For simplicity the resistances are assumed equal in the following. This reflects the belief that the resistance, as well as the river in question is homogeneous. For the proposed method in general, it is however, possible to let the resistance vary along the river. This would reflect stretches of rivers with different riverbed characteristic (e.g. different plants) or varying physical shape.

3 The grey box model in state space form

A combination of physical, biological and chemical knowledge from e.g. existing and well-known deterministic models with the information provided by data is a modelling method which uses all available information. This modelling principle was first labeled "grey box" modelling by Bohlin (1984), Bohlin (1994). Since traditional time series models based on data only, are called black box modelling and well-known models based on physical, biological and chemical information are sometimes called white box models, we suggest the label: Grey box modelling for the combined principle as is also done in Tulleken (1993). Young and Whitehead (1975) uses the term "internally descriptive" and "databased mechanistic" for essentially the same type of model. A similar linear reservoir model, as the one

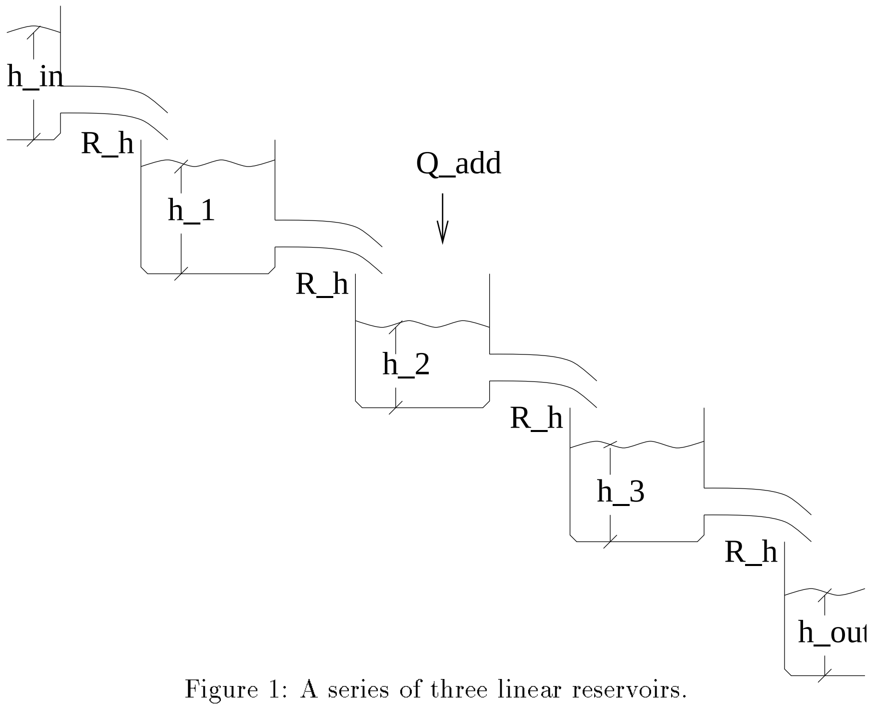


Figure 1: A series of three linear reservoirs.

proposed here, is reported in Whitehead et al. (1979). In both cases, however, they consider models in discrete time, whereas this paper considers continuous time models.

The suggested model is expected to consist of a number of reservoirs and the waterlevels, h_i in reservoir no i, are unmeasured state variables of the model. The known levels are h_{in} and h_{out} at the two measuring stations, respectively.

For the actual river the influx from rain is not considered to be diffuse along the whole stretch of the river section in question, but the influx is rather a point attribution. For the illustrated case the input from rain is only added for one reservoir.

For a model with three reservoirs, the differential equations become:

$$\frac{dh_{out}}{dt} = \frac{h_3 - h_{out}}{R_h} \tag{3.4}$$

$$\begin{aligned}
\frac{dh_3}{dt} &= \frac{h_2 - h_3}{R_h} \\
\frac{dh_2}{dt} &= \frac{h_1 - h_2}{R_h} + \frac{A_c\,\Phi}{A} P_r \\
\frac{dh_1}{dt} &= \frac{h_{in} - h_1}{R_h}
\end{aligned}$$

The water levels (h_1, h_2 and h_3) are not measured, but due to the state space formulation this is not necessary, since they can be estimated using a Kalman filter (see Jacobsen and Madsen (1996) or Jacobsen et al. (1996)).

For n reservoirs, the differential equations can be written as the state space model:

$$\begin{bmatrix} \frac{dh_{out}}{dt} \\ \frac{dh_n}{dt} \\ . \\ . \\ \frac{dh_1}{dt} \end{bmatrix} = \begin{bmatrix} \frac{-1}{R_h} & \frac{1}{R_h} & 0 & . & 0 \\ 0 & \frac{-1}{R_h} & \frac{1}{R_h} & . & 0 \\ 0 & . & . & . & 0 \\ . & . & . & . & . \\ 0 & . & . & . & \frac{-1}{R_h} \end{bmatrix} \begin{bmatrix} h_{out} \\ h_n \\ . \\ . \\ h_1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ . & . \\ 0 & \frac{A_c\,\Phi}{A} \\ \frac{1}{R_h} & 0 \end{bmatrix} \begin{bmatrix} h_{in} \\ P_r \end{bmatrix} \tag{3.5}$$

For simplicity, it is assumed that the contribution from rain is affecting only the second last reservoir. The observed water level, at the downstream measuring station, is:

$$\boldsymbol{Y} = \begin{bmatrix} 1 & 0 & . & . & 0 \end{bmatrix} \begin{bmatrix} h_{out} \\ h_n \\ . \\ . \\ h_1 \end{bmatrix} \tag{3.6}$$

To include the stochastic variation, an additive noise term, which represents the uncertainty in the input plus the uncertainty in the model, is added to Eq. (1.5). This is assumed to be a Wiener process (also called a Brownian motion). Also a term for describing the measurement error is introduced.

Thus, the stochastic model, when written in matrix notation, becomes:

$$d\boldsymbol{X} = \boldsymbol{A}\boldsymbol{X}\,dt + \boldsymbol{B}\boldsymbol{U}\,dt + d\boldsymbol{w}(t) \tag{3.7}$$
$$\boldsymbol{Y}(t) = \boldsymbol{C}\boldsymbol{X}(t) + \boldsymbol{e}(t) \tag{3.8}$$

where $\boldsymbol{X}$ is the state vector containing the states of water depths, $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices which characterise the dynamical behaviour of the system and specify how the input signals enter the system, respectively. The input vector $\boldsymbol{U}$ contains the measured input water level and the rain.

In order to be able to find maximum likelihood estimates of the parameters based on Gaussian distributions, the process $\boldsymbol{w}(t)$ is assumed to be a Wiener-process. The measurement error $\boldsymbol{e}(t)$ is assumed to be Gaussian distributed white noise with zero mean. Furthermore, it is assumed that $\boldsymbol{w}(t)$ and $\boldsymbol{e}(t)$ are mutually independent.

In the next section it is briefly outlined how the parameters of the model (state space model (1.5)), are estimated by a maximum likelihood method. The method accommodates missing observations (for a further description of this feature, see Jacobsen and Madsen (1996) and Jacobsen et al. (1996)).

4 Maximum likelihood estimates

All observations are equidistantly spaced at fifteen minute intervals. However, in order to simplify the notation, we shall assume that the time index t belongs to the set $\{0, 1, 2, ..., N\}$, where N is the number of observations. Introducing

$$\mathcal{Y}(t) = [\boldsymbol{Y}(t), \boldsymbol{Y}(t-1), \ldots, \boldsymbol{Y}(1), \boldsymbol{Y}(0)]' \tag{4.9}$$

i.e. $\mathcal{Y}(t)$ is a vector containing all the observations up to and including time t.

All the unknown parameters, denoted by the vector $\boldsymbol{\theta}$, are embedded in the continuous time state space model (eq.s (3.7) and (3.8)). The observations are, however, given in discrete time. Hence, the state space model has to be sampled in order to calculate the likelihood function. Under the assumption that the input variables are constant through the sampling time, the discrete-time state-space model can be written (Madsen and Melgaard

(1991), Melgaard and Madsen (1993a), Jacobsen and Madsen (1996)):

$$\begin{aligned} \boldsymbol{X}(t+1) &= \boldsymbol{\Phi}\,\boldsymbol{X}(t) + \boldsymbol{\Gamma}\,\boldsymbol{U}(t) + \boldsymbol{v}(t) && (4.10)\\ \boldsymbol{Y}(t) &= \boldsymbol{C}\,\boldsymbol{X}(t) + \boldsymbol{e}(t) && (4.11)\end{aligned}$$

where $\boldsymbol{\Phi} = e^{\boldsymbol{A}\tau}, \boldsymbol{\Gamma} = \int_0^\tau e^{\boldsymbol{A}s}\boldsymbol{B}ds$ and $\boldsymbol{v}(t)$ is a white noise sequence.

The likelihood function is the joint probability density of all the observations assuming that the parameters are known, i.e.

$$\begin{aligned} L'(\boldsymbol{\theta};\mathcal{Y}(N)) &= p(\mathcal{Y}(N)|\boldsymbol{\theta})\\ &= p(\boldsymbol{Y}(N)|\mathcal{Y}(N-1),\boldsymbol{\theta})p(\mathcal{Y}(N-1)|\boldsymbol{\theta})\\ &= \left(\prod_{t=1}^{N} p(\boldsymbol{Y}(t)|\mathcal{Y}(t-1),\boldsymbol{\theta})\right)p(\boldsymbol{Y}(0)|\boldsymbol{\theta}) && (4.12)\end{aligned}$$

where successive applications of the rule $P(A \cap B) = P(A|B)P(B)$ is used to express the likelihood function as a product of conditional densities.

Since both $\boldsymbol{v}(t)$ and $\boldsymbol{e}(t)$ are normally distributed the conditional density is also normal. The normal distribution is completely characterized by the mean and the variance. Hence, in order to parameterize the conditional distribution, we introduce the conditional mean and the conditional variance as

$$\begin{aligned} \hat{\boldsymbol{Y}}(t|t-1) &= E[\boldsymbol{Y}(t)|\mathcal{Y}(t-1),\boldsymbol{\theta}] && (4.13)\\ \boldsymbol{R}(t|t-1) &= V[\boldsymbol{Y}(t)|\mathcal{Y}(t-1),\boldsymbol{\theta}] && (4.14)\end{aligned}$$

respectively. It may be noticed that these correspond to the one-step prediction and the associated variance, respectively. Furthermore, it is convenient to introduce the one-step prediction error (or innovation)

$$\boldsymbol{\epsilon}(t) = \boldsymbol{Y}(t) - \hat{\boldsymbol{Y}}(t|t-1) \qquad (4.15)$$

Using (4.12) – (4.15) the conditional likelihood function (conditioned on $\boldsymbol{Y}(0)$) becomes

$$\begin{aligned} L(\boldsymbol{\theta};\mathcal{Y}(N)) &= \prod_{t=1}^{N}\Big((2\pi)^{-m/2}\det\boldsymbol{R}(t|t-1)^{-1/2}\\ &\qquad \exp(-\tfrac{1}{2}\boldsymbol{\epsilon}(t)'\boldsymbol{R}(t|t-1)^{-1}\boldsymbol{\epsilon}(t))\Big) && (4.16)\end{aligned}$$

where m is the dimension of the $\boldsymbol{Y}$ vector (in the present case m= 1). Traditionally the logarithm of the conditional likelihood function is considered

$$\begin{aligned}\log L(\boldsymbol{\theta};\mathcal{Y}(N)) &= -\tfrac{1}{2}\sum_{t=1}^{N}\left(\log\,\det \boldsymbol{R}(t|t-1) + \boldsymbol{\epsilon}(t)'\boldsymbol{R}(t|t-1)^{-1}\boldsymbol{\epsilon}(t)\right)\\ &+\ \text{const}\end{aligned} \tag{4.17}$$

In the linear case, the conditional mean and variance in Eq. (4.14) can be calculated recursively by using a Kalman filter. In the non-linear case, an extended Kalman filter is used for calculating the conditional mean and variance. The numerical details of the algorithms can be found in Madsen and Melgaard (1991).

The maximum likelihood estimates are found by maximization of the log likelihood function (4.17), and the uncertainties of the parameter estimates are found using the observed curvature of the log likelihood function, evaluated at the final estimates (Madsen and Melgaard (1991), Melgaard and Madsen (1993a)).

5 Results and discussion

Table 1 show the obtained maximum likelihood estimates for the parameters, and their standard deviations.

TABLE 1: MAXIMUM LIKELIHOOD PARAMETER ESTIMATES

Symbol	Parameter	Estimate	Unit
R_h	Resistance between reservoirs	0.340	[h]
		($0.744\ 10^{-2}$)	
$A_c\,\Phi/A$	Additional water flow	29.750	[–]
		(1.305)	

Table 1: Parameter estimates and their standard deviations.

A positive estimate for the resistance, R_h was to be expected. Since neither the surface area of the river section, nor the catchment area, is known, it is not possible to identify the runoff coefficient, Φ and fully assess this parameter. The ratio between these two areas was estimated to 29.75. With a width of the river about 6 – 7 m and the length of the section about a kilometer and a half, this corresponds roughly to a reduced catchment of about 25 hectars. The standard deviation of both parameters is reasonably small.

The measured, as well as the simulated water level, is seen in Figure 2. The simulation was performed using the estimated parameters. It is obvious that some of the variation is explained by the input (h_{in}); but a simulation without the input from rain has shown that the flow during rain has to be described by the rain input also. Otherwise, the model is not able to simulate the high peaks in the water level satisfactory.

The model was evaluated statistically and several tests for white noise behavior of the residuals have indicated that the model described the dynamics sufficiently. Hence, no more than three reservoirs are needed. On the other hand, less than three reservoirs leads to a model with autocorrelated errors. Hence, exactly three reservoirs are needed for the considered case.

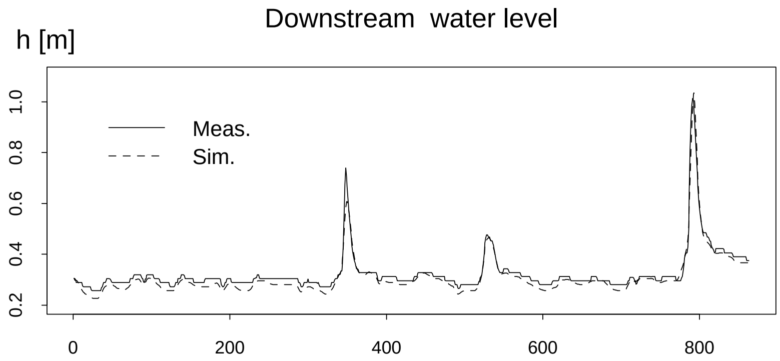


Figure 2: Measured and simulated water level.

It is seen from Figure 2 that the model slightly underestimates the water level. This could be interpreted as some sort of additional diffuse water supply regardless of whether it rains or not. Or it could be that the water level, h, alone cannot adequately describe the flow.

A constant diffuse water supply, Q_c, could easily be included in the description by a small modification of (1.5) to the following extended model:

$$\begin{bmatrix} dh_{out} \\ dh_n \\ . \\ . \\ dh_1 \end{bmatrix} = \begin{bmatrix} \frac{-1}{R_h} & \frac{1}{R_h} & 0 & . & 0 \\ 0 & \frac{-1}{R_h} & \frac{1}{R_h} & . & 0 \\ 0 & . & . & . & 0 \\ . & . & . & . & . \\ 0 & . & . & . & \frac{-1}{R_h} \end{bmatrix} \begin{bmatrix} h_{out} \\ h_n \\ . \\ . \\ h_1 \end{bmatrix} dt$$

$$+ \begin{bmatrix} 0 & 0 & Q_c \\ 0 & 0 & Q_c \\ . & . & . \\ 0 & \frac{A_c \Phi}{A} & Q_c \\ \frac{1}{R_h} & 0 & Q_c \end{bmatrix} \begin{bmatrix} h_{in} \\ P_r \\ 1 \end{bmatrix} \qquad (5.18)$$

Since the extended model is also a state space model of the general form described by (3.7) – (3.8), the parameters of the model (5.18) could easily be estimated by the suggested maximum likelihood method.

6 Conclusions

A description of the hydraulic system for a river receiving water from transient rain events has been proposed. This description is based on a series of linear reservoirs as an idealisation and uses measurements from two stations. The model is formulated as a grey box model in state space form.

A maximum likelihood method for estimating the parameters of state space models in continuous time for hydraulic dynamics is described. An ordinary Kalman filter is used to evaluate the likelihood function. The maximum likelihood approach makes it very easy to evaluate the model by statistical tools as indicated in the paper. Furthermore, the state space formulation is very flexible, and extensions like the incorporation of the diffuse water supply is easy without changing the statistical procedure.

The parameters of a very simple model, with three reservoirs, have been identified. Each of the parameter estimates is within reasonable bounds statistically. The model is based on physical laws for the system, as well as the actual data, making the model a grey box model. The continuous time formulation of the grey box approach has the advantage that a direct physical interpretation of the estimated parameters is possible. For model building, the grey box approach based on continuous time state space models is very attractive and flexible. As described in the paper, it is very easy to introduce and estimate a constant diffuse ground water infiltration, as well as hydraulic models where the resistance against flow varies along the stretch of the river.

The model will subsequently be used as the hydraulic part of a more complex model involving transient loads of water and pollutants and the effects on the oxygen concentration. Measured as dissolved oxygen, this quantity is one of the most universal environmental measures of water quality. As urban rivers and streams are often considered a recreational asset in towns and cities, it is important to be able to model the dynamics, such that an assessment of proposed regulations may be carried out (see e.g. Jacobsen et al. (1996)).

Paper G

A Linear Reservoir Model of the Oxygen Dynamics for a Two-station set-up.

Manuscript.

Abstract

In this paper measurements for the water level and oxygen concentration at two stations of a suburban river has been used to formulate and estimate a dynamic model for the hydraulics and the oxygen dynamics. The hydraulic part is an extension of the hydraulic model reported in Jacobsen et al. (1997b), and it will be shown that a model with two reservoirs is adequate.

The description of the oxygen dynamics is added on top of the hydraulic model. An adequate description of the hydraulics is very important, as the oxygen level in this river is heavily influenced by the transient impact from rain events.

A linear reservoir approximation is used to describe the dynamics of the system. The model is formulated as a grey box – or transparent – model, which means that the formulation is kept in continuous time and (most of) the parameters have a direct physical interpretation. Nine days of data, sampled every 15 minutes is used for estimating the parameters of the embedded continuous time model.

Statistical methods are used to validate the model and the estimated parameters are verified by a comparison with previously assumed values. Five different models are tried for the oxygen part of the model and the best one is based on the suggestion from Odum (1956), that the respiration is dependent on the organic matter in the water. The general conclusion is, however, that the hydraulic part of the models gives a satisfactory description, while the oxygen part of the models needs more attention in future work.

Keywords: Linear reservoir model; Stochastic model; Grey box model; Oxygen dynamics; Hydraulic model; Water quality.

Introduction

In a previous paper (Jacobsen et al. (1997b)) a linear reservoir model was developed to describe the hydraulics between two measuring stations in a small creek. Although the model was very simple it seems to give reasonable prediction and simulation results. However, validating the model it became clear that the model could be improved by a simple extension, i.e. by including a constant diffuse water supply.

In this paper, the results of the extended model are reported and the dynamics of dissolved oxygen as well as organic matter, expressed as Biological Oxygen Demand (BOD) is added to the extended hydraulic model. The description is based on the same linear reservoir concept as in Jacobsen et al. (1997b), but several conceptual models will be investigated. Jamieson et al. (1972) used a deterministic rainfall runoff model combined with a stochastic model for the model discrepancy and in effect argued for the grey box concept and Whitehead and Young (1975) used a similar concept where a deterministic flow model, which basically was a linear reservoir model, is coupled with a stochastic black box model for rainfall-runoff. A similar linear reservoir model, as the one proposed here, is reported in Whitehead et al. (1979), Whitehead et al. (1981). In these cases, however, they consider models in discrete time, whereas this paper considers continuous time models. Another example is reported in Young and Beck (1974) and Beck and Young (1975a), who proposed a similar model based on a single reservoir coupled with a transportation delay. Being based on daily measurements, diurnal variations were not observable and thus the assumption of constant volume and uniform flow-rate, for the given river-stretch, was reasonable.

The Data

The data used in this work originates from two measurement stations, 2191 meters apart:

- Station 302 Kokkedal – The upstream station.
- Station 301 Nive mølle – The downstream station.

The period covered was a full 9 day period in August 1996. Measurements are taken at 15 minute intervals and the considered river is heavily influenced by the transient impact from rain events. Hence, the volume of the river stretch in question varies with the water level, which shows large peaks corresponding to the rain events. Figures 1 and 2 show the measured water level at the downstream measuring station and the time series of rain events, respectively.

The input to the model is the measurements from the upstream station, the measured cumulated rain 2.5 kilometers further upstream, and the solar radiation measured in Tåstrup, about 15 km away. The water temperature was also measured at both stations and the average of these temperatures is used to calculate the saturated oxygen concentration.

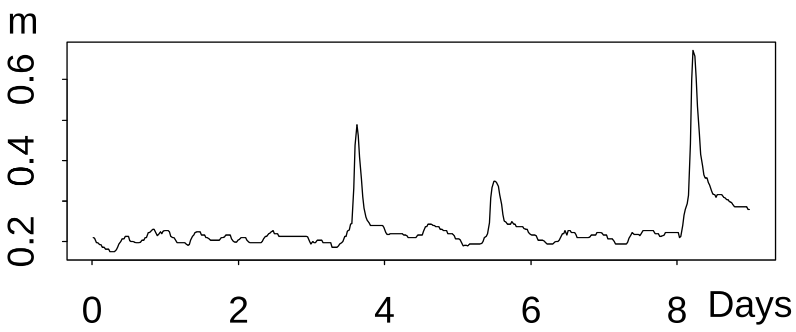


Figure 1: The measured water level at the downstream station.

In order to adequately describe the hydraulics of the system it turns out that more than one reservoir is necessary and the retention time between the two stations is then described by the inherent lag properties of the linear reservoir formulation.

When describing the hydraulics, it will be demonstrated that only two reservoirs are needed. Hence, when considering the models for the oxygen and organic matter part of the model will be based on two reservoirs.

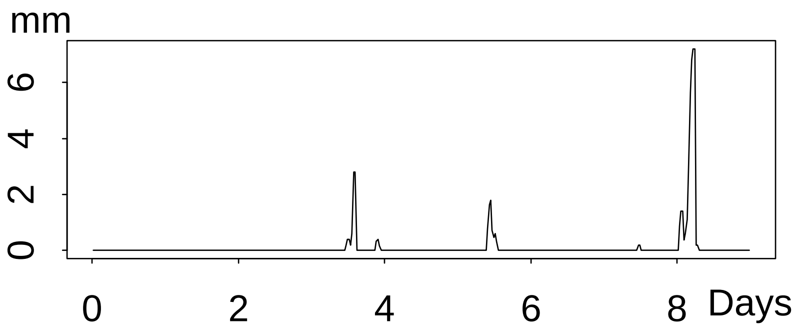


Figure 2: The rain intensity measured a few kilometers upstream. The rain is cumulated over 15 minutes sampling intervals.

1 Part one. The extended hydraulic model

In this part the extended hydraulic model will be presented along with the obtained results. In the previous paper (Jacobsen et al. (1997b)) the measured and the simulated water level differs in a way that indicates that the model could be improved by including a constant supply of water. Hence the model will be extended with a term to describe this.

1.1 An extended two-station hydraulic model

In Jacobsen et al. (1997b) the general matrix equations for n linear reservoirs was presented. Here, the equations will just be briefly introduced. Including a constant diffuse water supply Q_{add}, we obtain the following description of the water volume V for one reservoir:

$$\frac{dV}{dt} = Q_{in} - Q_{out} + Q_{add} \tag{1.1}$$

where Q_{in} and Q_{out} is the water flow in and out of the observed river-stretch, respectively and t is time. Assuming that: the volume $V = A_{surf}\, h$, it is further assumed that

$$Q_{in} = \frac{A_{surf}\, h_{in}}{R_h}; \qquad Q_{out} = \frac{A_{surf}\, h_{out}}{R_h}; \tag{1.2}$$

$$V = A_{surf}\ h \qquad Q_{add} = A_c\ \Phi P_r + Q_c \tag{1.3}$$

where A_{surf} is the surface area of the river-stretch, the water level is denoted h, and R_h is the resistance against water flow between the reservoirs. The additional water flow arises from an additional diffuse term, Q_c plus runoff during rain, which is calculated as the catchment area A_c, multiplied by the runoff coefficient Φ, and the rain input P_r. By substituting (1.2) and (1.3) into (1.1), and dividing by A_{surf}, we obtain the linearized change over time for a single reservoir:

$$\frac{dh_{out}}{dt} = \frac{h_{in} - h_{out}}{R_h} + \frac{A_c\ \Phi\ P_r\ +\ Q_c}{A_{surf}} \tag{1.4}$$

where the term for rain input is only introduced in one of the linear reservoirs. The surface area was allowed to vary with the water level.

For n reservoirs, the differential equations can be written in matrix notation as the state space model:

$$\begin{bmatrix} \frac{dh_{out}}{dt} \\ \frac{dh_n}{dt} \\ . \\ . \\ \frac{dh_1}{dt} \end{bmatrix} = \begin{bmatrix} \frac{-1}{R_h} & \frac{1}{R_h} & 0 & . & 0 \\ 0 & \frac{-1}{R_h} & \frac{1}{R_h} & . & 0 \\ 0 & . & . & . & 0 \\ . & . & . & . & . \\ 0 & . & . & . & \frac{-1}{R_h} \end{bmatrix} \begin{bmatrix} h_{out} \\ h_n \\ . \\ . \\ h_1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ . & . \\ 0 & \frac{A_c\,\Phi}{A} \\ \frac{1}{R_h} & 0 \end{bmatrix} \begin{bmatrix} h_{in} \\ P_r \end{bmatrix} \tag{1.5}$$

For simplicity, it is assumed that the contribution from rain is affecting only the second last reservoir. The observed water level, at the downstream measuring station, is:

$$\boldsymbol{Y} = \begin{bmatrix} 1 & 0 & . & . & 0 \end{bmatrix} \begin{bmatrix} h_{out} & h_n & . & . & h_1 \end{bmatrix}^T \tag{1.6}$$

To include the stochastic variation, an additive noise term, which represents the uncertainty in the input plus the uncertainty in the model, is added to Eq. (1.5), while a term for the measurement error is added to Eq. (1.6). It is assumed that these two stochastic terms are mutually independent.

The resistance, R_h accounts for the time lag property of the linear reservoir model. This may be interpreted as the retention time (or time of residence) between the measuring stations. In Young and Beck (1974) and Beck and Young (1975a) it was argued that the use of several reservoirs would necessitate measurements from a number of measuring stations, commensurate with the number of reservoirs used plus one. However, using the continuous time state space model formulation as in this work, this is not necessary as long as the states are observable. For more detail on this, see Jacobsen et al. (1996).

1.2 Results for the extended model

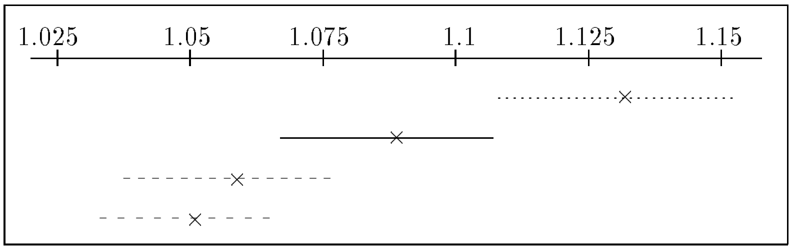
The model was run with increasing numbers of linear reservoirs. However, only the results from the investigations with one to four reservoirs will be used to determine an adequate number of reservoirs. Table 1 shows the obtained maximum likelihood estimates for the parameters and their standard variations.

No. of res.	R_h $[hr^{-1}]$		A_c $[m^2]$		Q_c $[m^3/hr]$	
1	1.130	(0.022)	25.059	(1.351)	$0.209\ 10^{-1}$	$(0.110\ 10^{-2})$
2	1.087	(0.020)	22.805	(1.360)	$0.225\ 10^{-1}$	$(0.965\ 10^{-3})$
3	1.064	(0.019)	21.443	(1.294)	$0.234\ 10^{-1}$	$(0.914\ 10^{-3})$
4	1.048	(0.016)	20.534	(0.895)	$0.240\ 10^{-1}$	$(0.892\ 10^{-3})$

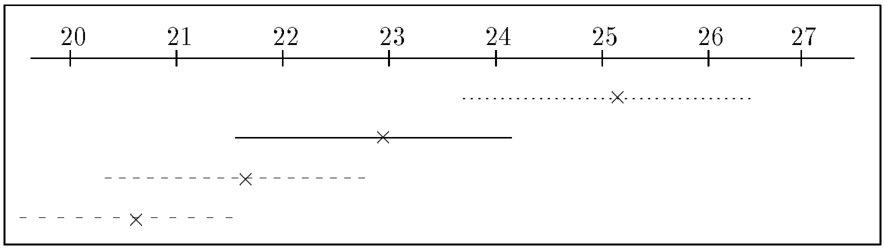
Table 1: Parameter estimates and their standard deviations in paranthesis.

The parameter estimates may also be depicted graphically as in Fig. 3. Notice, that the parameter estimates converge and become more accurate (smaller standard deviations) as the number of reservoirs increases.

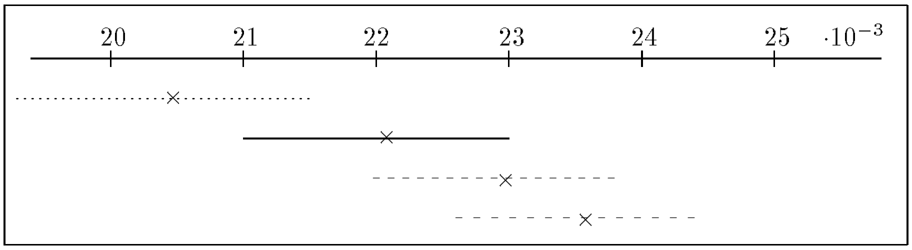
In order to find the best model, a number of criteria are considered. The model fit, viewed graphically, the correlation matrices between the parameter estimates and an R-square test were employed. The model fit, both for one-step prediction error and of the simulated data compared to the measured, were all very good. It was not really possible to make an objective choice on the number of reservoirs based on that.



Parameter estimates for the time constant R_h, and the standard deviations.



Parameter estimates for the catchment area A_c, and the standard deviations.



Parameter estimates for the additional waterflow Q_c, and the standard deviations.

······ 1 reservoir, —— 2 reservoirs, - - - 3 reservoirs, - - - 4 reservoirs.

Figure 3: Graphical presentation of the parameter estimates, and their standard deviation, for the three parameters and four reservoirs.

We introduce the simulation R_s^2 value as:

$$R_s^2 = \frac{V[Y] - V[Y - Y_{sim}]}{V[Y]} \tag{1.7}$$

where the variance for the actual data, $V[Y] = (0.085)^2$. The R_s^2 values are shown in Table 2.

Reservoirs	1	2	3	4
R_s^2	0.9729	0.9766	0.9774	0.9775

Table 2: Simulation R-square values for the different number of reservoirs.

As diagnostic measures, the log-likelihood values and the mean of the prediction error, and its variance is also given. But, because the models are not sub-models of each other, the log-likelihood values and the prediction errors are not directly comparable. Another diagnostic tool was the correlation matrices. It became clear that too many reservoirs resulted in high correlations between some of the estimated initial state values of the intermittent reservoirs. Table 3 shows the correlations and which initial states were highly correlated.

Res.	Which initial states	Correlations
1		
2	(h_1, h_2)	(-0.90)
3	(h_1, h_2); (h_2, h_3); (h_1, h_3)	(-0.99) (-0.99) (0.97)
4	(h_1, h_2); (h_1, h_3); (h_2, h_3)	(-0.97) (0.93) (-0.99)

Table 3: Correlations between estimates of the initial states.

They all occur for the intermediate reservoirs. In this respect, experience has shown that a correlation value of below 0.90 is acceptable. Therefore the model with two reservoirs between the input and the output was chosen as adequate. The correlations between the other parameters in the model were all below 50 %. From Figure 3 it is seen that the difference from going from three reservoirs to two is much less distinct than when going from two reservoirs to one.

The estimation was carried out using the program reported in Jacobsen et al. (1997b). The method is a maximum likelihood method, which uses the Kalman filter for calculating the likelihood function in the linear cases, while the extended Kalman filter (EKF) was employed for the non-linear cases. For a more thorough presentation of the parameter estimation technique, see Jacobsen and Madsen (1996), Madsen and Melgaard (1991), Melgaard and Madsen (1993a) and Melgaard et al. (1994).

The cumulative periodogram for the model is seen in Figure 4. The confi-

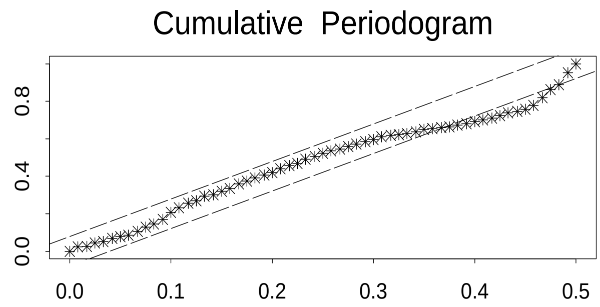


Figure 4: Cumulative periodogram for the hydraulic model. The dashed lines are the 95 % confidence limits.

dence limits are calculated using a Kolmogorov-Smirnov test principle (see Box and Jenkins (1976)). The periodogram indicates that the model is adequate, though with a slight discrepancy in the high frequency area.

In Figure 5 the measured and the simulated water level are shown. They are seen to be in allmost perfect agreement.

2 Part Two: The water quality model

In this part, the model for the oxygen and organic matter will be presented, along with the results for this particular model. This division between a hydraulic model and a water quality model is possible because the extended hydraulic model was included precisely as estimated into the full model. This is because no effect on the water level is expected to arise from the dynamics between oxygen and organic matter.

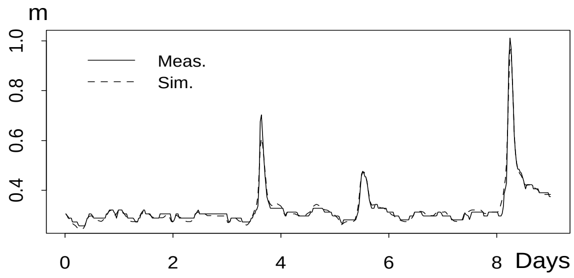


Figure 5: The measured and simulated water level at the downstream station.

2.1 The model for oxygen and organic matter

The model for oxygen and organic matter is formulated as a state space model, just as the hydraulic model. For each reservoir, the change in volume and concentrations of the two state variables are:

$$\frac{dV}{dt} C = Q_{in}\, C_{in} - Q_{out}\, C_{out} + Q\, S_c \tag{2.8}$$

$$\frac{dV}{dt} L = Q_{in}\, L_{in} - Q_{out}\, L_{out} - V\, S_l + Q_{add}\, L \tag{2.9}$$

where the same notation as in Eq.s (1.1) and (1.3) is used. The dissolved oxygen concentration is denoted C, and S_c is the source/sink term for reaeration, degradation of organic matter, photosynthesis and respiration. Organic matter (BOD) is denoted L and S_l is the corresponding source/sink term for degradation of organic matter. For simplicity it is assumed that the concentration of organic matter is the same in the river as in the diffuse water supply. Furthermore, we assume that:

$$S_c = K_2(C_m - C) - K_c\, L + P(I) - Resp \tag{2.10}$$

$$K_2 = \frac{K_L}{h} = const\, \frac{V_F}{h} = const \frac{\sqrt{g\, R_w\, I_o}}{h} \tag{2.11}$$

$$C_m = 14.54 - 0.39T + 0.01T^2 \tag{2.12}$$

$$S_l = K_3\, C - K_l\, L \tag{2.13}$$

where K_2 is the reaeration constant and K_L, the adsorption coefficient. C_m is the saturated oxygen as a function of temperature, while $P(I)$ denotes the photosynthesis as a function of solar radiation I, and $Resp$, the respiration. The temperature T is taken as an average of the input and output temperatures. The gravitational constant, $g = 9.82\ m/s^2$. I_o is the slope of the riverbed: $1.01\ ^0\!/_{00} = 1.01$ and R_w is the hydraulic radius, equal to the cross sectional area divided by the wetted perimeter, and calculated as a function of the water level.

In this paper, five different models are tried out for the oxygen dynamics. These include various expressions for photosynthesis, oxygen demand from the river bed and respiration. These models are presented in the next section.

Assuming that

$$\begin{aligned}
Q_{in}\, C_{in} &= \frac{A_{surf}\, h_{in}\, C_{in}}{R_h} \quad \text{and} \quad Q_{out}\, C_{out} = \frac{A_{surf}\, h_i\, C_{out}}{R_h} \\
&Q\, S_c(i) = A_{surf}\, h_i\, [K_2(i)(C_m - C_{out}) - K_c\, L + P(I) - Resp] \qquad (2.14)
\end{aligned}$$

$$\begin{aligned}
Q_{in}\, L_{in} &= \frac{A_{surf}\, h_i\, L_{in}}{R_h} \quad \text{and} \quad Q_{out}\, L_{out} = \frac{A_{surf}\, h_{in}\, L_{out}}{R_h} \\
V\, S_l &= A_{surf}\, h_i\, [K_3\, C - K_l\, L] \\
Q_{add}\, S_{add} &= \frac{A_c\, \Phi P_r}{V}\, L \qquad (2.15)
\end{aligned}$$

where the source/sink terms, S_c with the reaeration constant K_2, are dependent on the waterlevel, c.f.:

$$K_2(i) = K \frac{\sqrt{9.82\, R_w\, 1.01 \cdot 10^{-3}}}{h_i} \qquad (2.16)$$

Analogue with the hydraulic equations, we have the linearized change over time, for one reservoir:

$$\begin{aligned}
\frac{dC_{out}}{dt} &= \frac{C_{in} - C_{out}}{R_h} + K_2(C_m - C_{out}) \\
&\quad -K_c\, L + P(I) - Resp \qquad (2.17) \\
\frac{dL_{out}}{dt} &= \frac{L_{in} - L_{out}}{R_h} + V\, S_l + \frac{A_c\, \Phi P_r}{V} L_{out} \qquad (2.18)
\end{aligned}$$

Finally, the differential equations for the three states are:

$$
\begin{aligned}
\frac{dC_{out}}{dt} &= \frac{C_2 - C_{out}}{R_h} + S_c(C_{out}) \\
\frac{dL_{out}}{dt} &= \frac{L_2 - L_{out}}{R_h} + S_l(L_{out}) \\
\frac{dC_2}{dt} &= \frac{C_1 - C_2}{R_h} + S_c(C_2) \\
\frac{dL_2}{dt} &= \frac{L_1 - L_2}{R_h} + S_l(L_2) + \frac{A_c\, \Phi P_r}{V}\, L \\
\frac{dC_1}{dt} &= \frac{C_{in} - C_1}{R_h} + S_c(C_1) \\
\frac{dL_1}{dt} &= \frac{L_{in} - L_1}{R_h} + S_l(L_1) \qquad (2.19)
\end{aligned}
$$

2.2 Different models for the oxygen dynamics

In this investigation it was decided to try several different models for the oxygen dynamics to compare with the results obtained in previous papers and for another different river.

In Jacobsen and Madsen (1996) and Jacobsen et al. (1996) several different source/sink terms was included to model the following suggestions:

I The oxygen consumption by the riverbed is described by a a half order model. I.e. as the square root of the oxygen concentration.

II An organic matter source from the precipitation term, modelled linearly.

III A controversial organic matter source provided by oxygen partaking in nitrification, which provides nourishments in the form of nutrient salts, resulting in more plants:

Respiration $R(T)$ was corrected for temperature, T, and two photosynthetic functions $P(I)$ was tried out. A two-parameter Michaelis-Menten type function, which was found to perform best in an investigation of the dynamical effect of solar radiation on oxygen dynamics in Jacobsen et al. (1998b) and a linear function of the solar radiation I, which was found adequate in Jacobsen and Madsen (1996), Jacobsen et al. (1996).

During the estimation procedures, it became apparent that estimating the complete model with both organic matter and oxygen gave rise to problems. Many parameters turned out to be not significantly different from zero, and as the parameters from the two-parameter model of photosynthesis as a function of solar radiation were highly correlated, it was decided to try the linear model. This provided more reasonable estimates. Hence, only the linear model is considered in the following. The models were all estimated using the program described in Jacobsen et al. (1997b).

Model A

This model was used to test whether any contribution of organic matter could stem from oxygen (sort of a feedback mechanism). The term involved is K_3C in Eq. (2.15). It was tested whether that oxygen used in nitrification provided nourishments in the form of nutrient salts, thereby giving rise to more plant growth and thus organic matter. In this model, that parameter did not differ significantly from zero (on a level of 5 % and furthermore resulted in complete correlation with the parameter K_l, which governs the degradation of organic matter.

Model B

It was tried to model the half order river bottom oxygen demand independently of the precipitation P_r, as well as multiplied with this term. It was tested whether the rain carried with it particular organic matter which would result in this oxygen demand. Only when coupled by the term $K_b\,P_r$ with rain did the parameter estimate K_b become significantly different from zero. However, there was full correlation (1.00) between K_l and K_b, implying the two terms for degradation of organic matter cannot be separated. The simulation R-square (R_s^2) value were 0.6449999 (see Eq. (1.7)).

Model C

This purpose of this model was to test the idea, first proposed by Odum (1956), that respiration was dependent on the organic matter. Here, just a simple linear function was tried out. The results showed that with this model, the parameter for degradation of organic matter K_l did not differ significantly from zero and K_l and respiration, Rsp was correlated with a correlation coefficient of 0.95. Thus, the model cannot identify whether oxygen is consumed from organic matter with respiration or as a first order degradation. The simulation R_s^2 value for the model was 0.6450015.

Model D
For comparison, it was decided to try a very simple model without river bottom oxygen consumption during rain ($K_b = 0$) or contribution from organic matter ($K_3 = 0$) and respiration dependence of organic matter. All the remaining parameters turned out to be significantly different from zero in the t-tests, but there was a quite high correlation between respiration and the degradation constant K_l. The simulation R_s^2 value were 0.6450006. The residual analysis showed that there were dynamics not accounted for, especially in the lower frequency area.

Model E
Finally, it was decided to try modelling only oxygen, excluding organic matter, but trying both types of photosynthetic functions. Both models seemed fine, with very little outwardly difference. R_s^2 values for the two models were 0.6445545 for the model with the linear photosynthetic function and 0.6445544 for the other. Simulation plots and model diagnostics looked indistinguishable, as well. However, the residual analysis showed that there were dynamics not accounted for, especially in the lower frequency area. Thus, it can be inferred that the uncertainty of this model is larger and covers the uncertainty between the two sub-models for photosynthesis.

3 Results for the total model

The five models were estimated by providing reasonable initial values and a ranges, they were expected to be confined within. The software used, provides a number of model diagnostics, such as t-tests to test hypotheses that a given parameter is not significantly different from zero, auto and partial auto correlation functions and cumulative periodogram of the residuals (see Melgaard and Madsen (1993a)). Table 4 show the estimated parameters, pertaining to the dynamics of the oxygen and the organic matter.

Par.	K_{ad}	K_c	β	K_l	K_b	Rsp
Unit	$[m/h]$	$[h^{-1}]$	$[d^{-1}]$	$[mg/l]$	$[-]$	$[mg/l]$
B	$0.507\ 10^{-1}$	0.586	$0.378\ 10^{-2}$	$0.267\ 10^{-3}*$	$0.271\ 10^{-1}$	$0.152\ 10^{-2}$
	$(.758\ 10^{-2})$	$(.766\ 10^{-1})$	$(.271\ 10^{-3})$	$(.173\ 10^{-3})$	$(.108\ 10^{-1})$	$(.254\ 10^{-3})$
C	$0.507\ 10^{-1}$	0.718	$0.378\ 10^{-2}$	$0.179\ 10^{-3}*$	NA	0.116
	$(.528\ 10^{-2})$	(.295)	$(.269\ 10^{-3})$	$(.417\ 10^{-3})$	NA	$(.428\ 10^{-1})$
D	$0.507\ 10^{-1}$	.743	$.378\ 10^{-2}$	$0.702\ 10^{-3}$	NA	$0.152\ 10^{-2}$
	$(.512\ 10^{-2})$	$(.750\ 10^{-1})$	$.267\ 10^{-3})$	$(.343\ 10^{-3})$	NA	$(.262\ 10^{-3})$
E	$0.507\ 10^{-1}$	NA	$0.377\ 10^{-2}$	NA	NA	0.768
	$(.609\ 10^{-2})$	NA	$(.265\ 10^{-3})$	NA	NA	$(.742\ 10^{-2})$

Table 4: The estimated parameters for the various models. Standard deviations are shown in paranthesis. An $*$ denotes that the parameter was not significantly different from zero, while the signature NA denotes that this parameter was not included.

In several cases the degradation coefficient K_l, for organic matter did not become significantly different from zero. One reason is that it may not be possible to include more than one term for each component (oxygen and organic matter) as the model cannot distinguish between the two. Another possibility is that perhaps the degradation coefficients, K_c and K_l for consumption of oxygen and the degradation of organic matter, respectively, in reality should be identical. These were estimated as two different values in the water quality model, reported in Jacobsen and Madsen (1996), Jacobsen et al. (1996). However, that was a model for data from only one measuring station. It was thought that if the two should be identical, this would be revealed in the estimation process (i.e. they would have identical estimates).

As for the extended hydraulic model, the R-square values for the simulation and prediction error was calculated between the models.

Model	Simulation	Prediction
B	0.6449999	0.9149555
C	0.6450015	0.9149558
D	0.6450006	0.9149556
E	0.6445545	0.9148689

Table 5: R-square values for simulation and prediction error between the models.

Thus, it may be inferred that model C gave the best model. However, the differences between the models were small and the residual analysis was not satisfactory, as shown in the cumulative periodogram for best model (C) is seen in Figure 6.

The periodogram shows that the model is not adequate. There is too much cumulation in the low frequency area. Thus, the model needs additional description for the lower frequencies. In Figure 7 the measured and the simulated oxygen concentration is shown.

It is seen that some dynamics are not accounted for. It occurs on the days with rain and may be caused by the spatial variability that is not accounted for by a point measure of precipitation. Interestingly, in Young and Beck (1974), it was prolonged sunny weather that was thought to cause discrepancies in the model. Solar radiation data was not available in their study, hence a mechanism for 'sustained sunlight effect' was included by

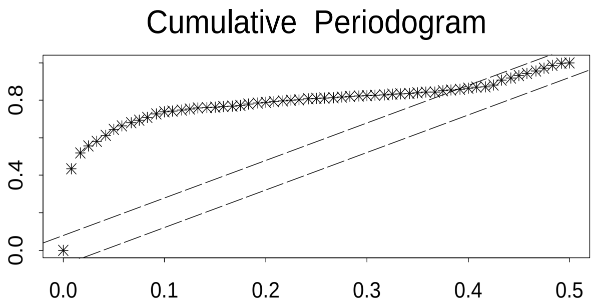


Figure 6: Cumulative periodogram for model C. The dashed lines are the 95 % confidence limits.

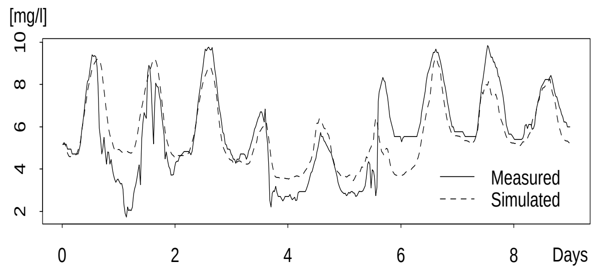


Figure 7: Measured and simulated oxygen at Station 301 in Usserød Å for 9 days.

way of a low-pass filter. In the case of the river modelled in this paper, the solar radiation data were measured some 15 km away. Thus, spatial variations are included, which are not accounted for.

There are several theories that could be tried out to improve the oxygen model besides improving the description of the processes during the transient impact from rain. Portielje et al. (1996) have shown that gross production is positively related to the level of external nutrient input. Perhaps a model with a separate state for the bottom oxygen demand from benthic algae could improve the model. Another alternative is a variation of model C, with respiration as a function of the oxygen content instead of a function of the organic matter. Such a model have proved successful in laboratory studies (Owens and Marris (1964), McDonnell and Weeter (1972)). Monitoring of the river is ongoing. If it were possible to obtain a new data series in which, CO_2 was also measured, it would be possible to try out theories as to photosynthesis being dependent on the pools of CO_2 (Sand-Jensen (1983), Sand-Jensen (1989), ?).

4 Conclusion

In this paper an extended model for hydraulics is proposed and the parameters of that model was estimated. The model was investigated for various number of reservoirs and a model with two reservoirs was deemed adequate by judging the simulation performance and the supplied model diagnostics and it is concluded that this model gave a reasonable description of the hydraulics.

Then the extended hydraulic model, with two reservoirs, was further expanded to include the dynamics of the oxygen and organic matter for that river. This model was based on the same linear reservoir concept as the hydraulic model. Hence, the extended hydraulic model was included without alterations into the final water quality model. Five different models were tried out and the best one determined from the modelling diagnostics and the simulation performance.

Overall, the best model for this stretch of river was a model where respiration was made linearly dependent on the organic matter. There were only five parameters, but the degradation constant K_l was not accepted as being significantly different from zero. Also, that parameter was correlated with respiration, with a correlation coefficient of 0.95. Hence, it is possible that the model cannot distinguish whether oxygen is consumed from organic matter with respiration or as a first order degradation although it seems that the consumption is best described with respiration. It should also be investigated further whether the consumption constant, K_c for oxygen and the degradation constant for organic matter, K_l, in fact should be identical.

However, none of the models performed completely convincing, as the residual analyses showed that the models were not fully adequate and that the residuals did not seem to be independent. This may be caused by spatial variability, not accounted for in the rain and solar radiation measurements. Further work will have to be done to determine how to extend the model to obtain a satisfying model. However, this preliminary work shows that a linear reservoir model may be very attractive when modelling water quality, when data from two measuring stations are available. The continuous time state space model formulation facilitates that states not measured may be incorporated, which is a clear advantage when multiple reservoirs is warranted to model the dynamics of the oxygen and organic matter interactions. Organic matter, expressed as the biological oxygen demand over five days, is still not feasible to measure in a monitoring program where short sampling intervals are wanted to include the diurnal variations in the oxygen dynamics.

References

Åstrøm, K.-J. and Bohlin, T. (1965). Numerical identification of linear dynamic systems from normal operating records. *Proc. IFAC Symposium for Theory of Self-adaptive Control Systems, Teddington, England*, 96–111.

Beck, M. and Young, P. (1975a). A dynamic model for DO-BOD relationships in a non-tidal stream. *Water Research*, **9**, 769–776.

Beck, M. and Young, P. (1975b). Identification and parameter estimation of DO-BOD models using the extended Kalman filter. Tech. rep. CUED/F - Control/TR 93(1975), University Engineering dept., Cambridge.

Beck, M., Jakeman, A., and McAlleer, M. (1993). Construction and evaluation of models of environmental systems. In Jakeman, A., Beck, M., and McAlleer, M. (Eds.), *Modelling Change in Environmental Systems*, pp. 3–35. John Wiley and Sons.

Beck, M. (1991). Principles of modelling. *Water, Science and Technology*, **24**(6), 1–8.

Beven, K. (1989). Changing ideas in hydrology – the case of physically based models. *Journal of Hydrology, 105*, 157–172.

Bjørnø and Hartig (1984). *Textbook for Streamflow phenomena*. Polyteknisk Forlag. (In Danish: Strømningslære).

Bloomfield, P. (1976). *Fourier Analysis of Time Series: An Introduction*. Wiley.

Bohlin, T. (1984). Computer-aided grey-box validation. Tech. rep. TRITA-REG-8403, Department of Automatic Control, Royal Institute of Technology, Stockholm, Sweden.

Bohlin, T. (1994). A case study of grey box identification. *Automatica*, **30**(2), 307–318.

Box, G. E. P. and Jenkins, J. M. (1976). *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco.

Boyle, J. and Scott, J. (1984). The role of benthic films in the oxygen balance in an east Devon river. *Water Resources*, **18**(9), 1089–1099.

Bramslev, J. P. (1989). *Treatment of precipitation data from a precipitation measuring system, covering the country.* Department of Technical Hygiene, Technical University of Denmark. (In Danish: Bearbejdning af nedbørsdata fra et landsdækkende regnmålersystem).

Camp, T. R. (1965). Field estimates of oxygen balance parameters. *Journal of the Sanitary Engineering Division, Proceedings of the American Society of Civil Engineers*, **91**(SA5), 1–16. (In: River basin management, proceedings of a conference).

Chow, V. T., Maidment, D. R., and Mays, L. W. (1988). *Applied Hydrology.* McGraw-Hill International Editions.

Cosby, B. and Hornberger, G. (1984). Identification of photosynthesis-light models for aquatic systems. I. Theory and simulations. *Ecological Modelling*, **23**, 1–24. Elsevier Science Publishers B.V., Amsterdam.

Cosby, B., Hornberger, G., and Kelly, M. (1984). Identification of photosynthesis-light models for aquatic systems. II. Application to a macrophyte dominated stream. *Ecological Modelling*, **23**, 25–51. Elsevier Science Publishers B.V., Amsterdam.

Cosby, B. (1984). Dissolved oxygen dynamics of a stream: Model discrimination and estimation of parameter variability using an extended Kalman filter. *Water, Science and Technology*, **16**(5-7), 561–570. (In: River basin management, proceedings of a conference).

Di Toro, D. M. (1969). Stream equations and method of characteristics. *Journal of the Sanitary Engineering Division*, **SA4**, 699–703. Proceedings of the American Society of Civil Engineers.

Dobbins, W. (1964). Bod and oxygen relationships in streams. *Journal of the Sanitary Engineering Division*, **SA3**, 53–78. Proceedings of the American Society of Civil Engineers.

El-Shaarawi, A., Maul, A., and Brownlee, B. (1988). Modelling the dissolved oxygen change in streams using nonlinear regression analysis. Tech. rep. Scientific Series No 163, Directorate Research Institute for Inland Waters, Burlington Ontario, Canada.

Erlandsen, M. and Sørensen, M. (1984). Statistical analysis of the variation of the oxygen concentration in a river by means of diffusion processes. In *Symposium i Anvendt Statistik*, pp. 421–431. RECAU, Aarhus. (Symposium in Applied Statistics).

Erlandsen, M. and Thyssen, N. (1983). Modelling the community oxygen production in lowland streams dominated by submerged macrophytes. In *Analysis of Ecological: State-of-the-art in Ecological modelling*, pp. 855–860. Elsevier Scientific Publishing Company.

Frandsen, P. E., Jonasson, K., and Tingleff, O. (1995). Robust subroutines for non-linear optimization. Tech. rep. 69-2, Institute for Numerical Analysis, Technical University of Denmark, Lyngby, Denmark.

Gordon, D. M. and Sand-Jensen, K. (1990). Effects of O_2, pH and DIC on photosynthetic net O_2 evolution by marine macroalgae. *Marine Biology*, **106**, 445–451.

Hansen, L. H., Jacobsen, J. L., Nielsen, H. A., and Nielsen, T. S. (1996). Approximating building components using stochastic differential equations. In Bloem, J. (Ed.), *In: System Identification competition*, pp. 149–162. Institute for Systems Engineering and Informatics, Joint Research Centre, Ispra (VA), Italy.

Harremoës, P. and Malmgren-Hansen, A. (1990). *Textbook for Water pollution*. Polyteknisk Forlag. (In Danish: Lærebog i Vandforurening).

Harremoës, P., Henze, H., Arvin, E., and Dahi, E. (1989). *Theoretical Water Sanitation*. Polyteknisk Forlag. (In Danish: Teoretisk vandhygiejne).

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, New York.

Hornberger, G. M. and Kelly, M. G. (1972). The determination of primary production in a stream using an exact solution to the oxygen balance equation. *Water Resources Bulletin, American Water Resources Association*, **8**(4), 795–801.

Hornberger, G. M. and Kelly, M. G. (1975). Atmospheric reaeration in a river using productivity analysis. *Proceedings, American Society of Civil Engineers, ASCE.*

Hornberger, G. M., Kelly, M. G., and Eller, R. M. (1976). The relationship between light and photosynthetic rate in a river community and implications for water quality modelling. *Water Resources Research*, **12**(4).

Hunter, J. (1978). Accounting for the effects of water temperature in aerator test procedures. In *Proceedings workshop toward an oxygen standard, US–EPA*, p. 85.

Jacobsen, J. L. and Madsen, H. (1996). Grey box modelling of oxygen levels in a small stream. *EnvironMetrics*, **7**, 109–21.

Jacobsen, J. L. and Voss, L. (1994a). Robust estimation and time series analysis of oxygen levels in a small stream. In *Symposium i Anvendt Statistik*, pp. 152–167. (Symposium in Applied Statistics).

Jacobsen, J. L. and Voss, L. (1994b). Time series analysis and modelling of corrupted data. Master's thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby, Denmark. No. 12/94.

Jacobsen, J. L., Madsen, H., and Harremoës, P. (1996). Modelling the transient impact of rain events on the oxygen content of a small creek. *Water, Science and Technology*, **33**(2), 177–187. (In: Uncertainty, Risk and Transient Pollution Events. Selected proceedings of the IAWQ Interdisciplinary Symposium).

Jacobsen, J. L., Madsen, H., and Harremoës, P. (1997a). Estimation of a low pass filter for solar radiation data. In *Proceedings of the 9th Conference of the European Consortium for Mathematics in Industry ECMI.* Institute of Mathematics, Technical University of Denmark, Lyngby, Denmark.

Jacobsen, J. L., Madsen, H., and Harremoës, P. (1997b). A stochastic model for two-station hydraulics exhibiting transient impact. *Water, Science and Technology*, **36**(5), 19–26.

Jacobsen, J. L., Madsen, H., and Harremoës, P. (1998a). Modelling the dynamical effect of solar radiation on the oxygen content of a small creek. *EnvironMetrics*, **9**(1), 3–14. (Special Issue: Statistical Methods for Environmental Sciences).

Jacobsen, J. L., Madsen, H., and Harremoës, P. (1998b). Modelling the dynamical effect of solar radiation on the oxygen content of a small creek. *EnvironMetrics*, **9**(1), 3–14. (Special Issue Statistical Methods for Environmental Sciences).

Jacobsen, J. L. (1997). Technology management of the transfer of knowledge within stochastic modelling. Tech. rep. No. 24/1997, Institute of Mathematical Modelling, Technical University of Denmark.

Jakeman, A. and Hornberger, G. (1993). How much complexity is warranted in a rainfall–runoff model. *Water Resources Research*, **29**(8), 2637–2649.

Jamieson, D. G., Wilkinson, J. C., and Ibbitt, R. P. (1972). Hydrological forecasting with sequential deterministic and stochastic stages. In *International Symposium on uncertainties in hydrologic and water resource systems.* University of Arizona, Tucson, Arizona 85721. (Volume I).

Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory.* Academic Press, New York.

Jeppesen, E. and Thyssen, N. (1984). Modelling the seasonal variation in structural biological components and oxygen in macrophyte dominated streams: A summary of work in progress. *Water, Science and Technology*, **16**(5-7), 533–537. (In: River basin management, proceedings of a conference).

Kaiser, J. F. and Reed, W. A. (1977). Data smoothing using low-pass digital filters. *Rev. Sci. Instrum.*, **11**.

Kendrick, D. A., Rao, H. S., and Wells, C. H. (1970). Optimal operation of a system of waste water treatment facilities. In *Proceedings of the 9th IEEE Symposium on Adaptive Processes.* Electronics Research Center, The University of Texas, Austin.

Ljung, L. and Caines, P. E. (1979). Asymptotic normality of prediction error estimators for approximate system models. *Stochastics*, **3**, 29–46.

Ljung, L. and Glad, T. (1994). On global identifiability for arbitrary model parameterisations. *Automatica*, **30**(2), 265–276.

Ljung, L. (1979). Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. *IEEE Transactions on Automatic Control*, *AC-24*(1), 36–50.

Ljung, L. (1987). *System Identification: Theory for the User.* Prentice-Hall, Englewood Cliffs, New Jersey.

Madsen, H. and Holst, J. (1995). Estimation of continuous-time models for the heat dynamics of a building. *Energy and Buildings*, **22**, 67–79.

Madsen, H. and Holst, J. (1996). *Modelling Non-Linar and Non-Stationary Time Series.* Institute of Mathematical Modelling, Technical University of Denmark. (Lecture notes).

Madsen, H. and Melgaard, H. (1991). The mathematical and numerical methods used in CTLSM - a program for ML-estimation in stochastic, continuous time dynamical models. Tech. rep. No. 7/1991, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby, Denmark.

Madsen, K. and Tingleff, O. (1990). Robust subroutines for non-linear optimization. Tech. rep. NI-90-06, Institute for Numerical Analysis, Technical University of Denmark, Lyngby, Denmark.

Madsen, H., Spliid, H., and Thyregod, P. (1985). Markov models in discrete and continuous time for hourly observations of cloud cover. *Journal of Climate and Applied Meteorology*, **24**, 629–639.

Madsen, H., Melgaard, H., and Holst, J. (1990). Identification of building performance parameters. In Bloem, J. (Ed.), *Workshop on Advanced Identification Tools in Solar Energy Research*, Non Nuclear Energy, pp. 37–60. Commission of the European Communities, DG XII.

Madsen, H. (1989). *Time Series Analysis.* Institute of Mathematical Statistics and Operations Research, Technical University of Denmark. (In Danish: Tidsrækkeanalyse).

Martin, R. D. and Yohai, V. J. (1985). Robustness in time series and estimating arma models. In Hannan, E. J., Krishnaiah, P. R., and Rao, M. M. (Eds.), *Handbook of Statistics*, pp. 119–155. Elsevier Science Publishers, Amsterdam.

McDonnell, A. J. and Weeter, D. W. (1972). Variable respiration in aquatic plant communities. *Journal of the Sanitary Engineering Division*, **SA2**, 325–339. Proceedings of the American Society of Civil Engineers.

Melgaard, H. and Madsen, H. (1993a). CTLSM continuous time linear stochastic modelling. In Bloem, J. (Ed.), *Workshop on Application of System Identification in Energy Savings in Buildings*, pp. 41–60. Institute for Systems Engineering and Informatics, Joint Research Centre, Ispra (VA), Italy.

Melgaard, H. and Madsen, H. (1993b). CTLSM version 2.6 - a program for parameter estimation in stochastic differential equations. Tech. rep. No. 1/1993, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby, Denmark.

Melgaard, H., Madsen, H., and Nielsen, B. (1994). Identification of the heat dynamics of greenhouses: a course on using CTLSM. In Bloem, J. (Ed.), *Workshop on Application of System Identification in Energy Savings in Buildings*, pp. 61–76. Institute for Systems Engineering and Informatics, Joint Research Centre, Ispra (VA), Italy.

Melgaard, H. (1994). *Identification of Physical Models.* Ph.D. thesis, Institute of Mathematical Modelling, Technical University of Denmark. No. 1994-4.

Odum, H. T. (1956). Primary production in flowing waters. *Limnology Oceanography*, **1**, 102–117.

Øksendal, B. (1995). *Stochastic Differential Equations* (4 edition). Springer Verlag, Berlin.

Owens, M. and Marris, P. J. (1964). Some factors affecting the respiration of some aquatic plants. *Hydrobiologica*, **23**, 533–543.

Pedersen, A. R. (1994). Statistical analysis of gaussian diffusion processes based on incomplete discrete observations. Tech. rep. 297/October

1994, Department of Theoretical Statistics, Institute of Mathematics, University of Århus.

Portielje, R., Kersting, K., and Lijklema, L. (1996). Primary production estimation from continuous oxygen measurements in relation to external nutrient input. *Water Research*, **30**(3), 625–643.

Rauch, W. and Harremoës, P. (1995). Integrated urban water quality management. *Med. Fac. Landbouww. Univ. Gent. 60/4b.*

Rauch, W. and Harremoës, P. (1996). The importance of the treatment plant performance during rain to acute water pollution. *Water, Science and Technology*, pp. 1–8.

Sand-Jensen, K. (1983). Photosynthetic carbon sources of stream macrophytes. *Journal of experimental Botany*, **34**(139), 198–210.

Sand-Jensen, K. (1989). Environmental variables and their effect on photosynthesis of aquatic plant communities. *Aquatic Botany*, **34**, 5–25.

Sand-Jensen, K. (1995). Personal communications. *University of Copenhagen, Fresh Water lab., Hillerød, Denmark.*

Shanahan, P. and Somlyódy, L. (1995). Modelling the impact of diffuse pollution on receiving water quality. Tech. rep. WP-95-2, International Institute for Applied Systems Analysis, IIASA, Laxenburg, Austria.

Simonsen, J. (1974). *Oxygen fluctuation in streams.* Ph.D. thesis, Department of Technical Hygiene, Technical University of Denmark.

Söderström, T. and Stoica, P. (1989). *System Identification.* Prentice Hall, London.

Söderström, T. (1994). Theory of system identification. In Bloem, J. (Ed.), *System Identification Aplied to Building Performance Data*, EUR 15885 EN. Institute for Systems Engineering and Informatics, Joint Research Centre, Ispra (VA), Italy.

Spildevandskommiteen (1985). *Pollution of streams fra overflow structures.* Skrift nr. 22. DIF. (In Danish: Forurening af vandløb fra overløbsbygværker).

Stigter, J. and Beck, M. (1994). A new approach to the identification of model structure. *EnvironMetrics*, **5**, 315–333.

Streeter, H. W. and Phelps, E. B. (1925). *A Study of the Pollution and Natural Purification of the Ohio River.* Bulletin no. 146. U. S. Public Health Service, Washington.

Thackston, E. L. and Krenkel, D. (1969). Reaeration prediction in natural streams. *Journal of Sanitary Engineering Division.* Proceedings of the American Society of Civil Engineers.

Thomsen, E. and Thyssen, N. (1979). Monitoring of important state variables in rivers. *Vand*, **29**(2), 46–53. (In Danish: Måling og registrering af vigtige tilstandsvariabler i vandløb).

Thyssen, N., Erlandsen, M., Jeppesen, E., and T.F.Holm (1983). Modelling the reaeration capacity of lowland streams dominated by submerged macrophytes. In *Analysis of Ecological: State-of-the-art in Ecological modelling*, pp. 861–867. Elsevier Scientific Publishing Company.

Thyssen, N., Erlandsen, M., Kronvang, B., and Svendsen, L. (1990). *Stream models – biological structure and exchanges*, Vol. **C10**. NPo–forskning fra Miljøstyrelsen. (In Danish: Vandløbsmodeller – biologisk struktur og stofomsætning. NPo–forskning fra Miljøstyrelsen.).

Thyssen, N. (1980). *Effects of physical/chemical and biological processes on the oxygen balance of a macrophyte dominated stream, with special attention to the oxygen processes: Reaeration and photosynthesis.* Ph.D. thesis, Botanical Institute, University of Århus. (In Danish: Effecten af fysisk/kemiske og biologiske processer på et macrofytrigt vandløbs oxygenbalance, med særlig relation til de oxygentilførende processer: Genluftning og fotosyntese.).

Thyssen, N. (1982). Aspects of the oxygen dynamics of a macrophyte dominated lowland streams. In Symoens, J., Hooper, S., and Compére, P. (Eds.), *Studies on Aquatic Vascular Plants*, pp. 202–213.

Tong, H. (1990). *Non-linear Time Series, A Dynamical System Approach.* Clarendon Press, Oxford.

Tulleken, H. J. A. F. (1993). Grey-box modelling and identification using physical knowledge and bayesian techniques. *Automatica*, **29**(2), 285–308.

Walter, E. and Pronzato, L. (1990). Qualitative and quantitative experiment design for phenomenological models – a survey. *Automatica*, **26**, 195–213.

Whitehead, P. and Young, P. (1975). A dynamic-stochastic model for water quality in part of the Bedford-Ouse river system. In Vansteenkiste, G. (Ed.), *Computer Simulation of Water Resources Systems*, pp. 417–438. North Holland, Amsterdam.

Whitehead, P., Young, P., and Hornberger, G. M. (1979). A systems model of stream flow and water quality in the Bedford-Ouse river – I. Stream flow modelling. *Water Research*, **13**, 1155–1169.

Whitehead, P., Beck, M., and O'Connel, E. (1981). A systems model of stream flow and water quality in the Bedford-Ouse river – II. Water quality modelling. *Water Research*, **15**, 1157–1171.

Whitehead, P. (1980). Real time monitoring of water quality in the Bedford

Ouse river system. *Hydrological forecasting*, 333–341. Proceedings of the Oxford Symposium, April 1980: IAHS–AISH (Publ. no. 129).

Winter, J. (1988). *Der Einfluß von Emmisionen aus Mishentwässerungssystemen auf den Sauerstoffhaushalt in Fließ gewässern.* Scriftenriehe für Stadtentwässerung und Gewässerschutz. Hannover. (In German: The influence from emmision of combined systems on the oxygen balance in streams).

Wittenberg, D. (1992). *Urbane Gewässer – Das Immisionsprinzip als Plannungsansatz für die Stadtentwässerung.* Scriftenriehe für Stadtentwässerung und Gewässerschutz, vol 5. SuG, Hannover. (In German: Urban water –The principle for immision as a planningtool for the city water).

Wright, R. M. and McDonnell, A. J. (1986). Macrophyte growth in shallow streams: Biomass model. *Journal of Environmental Engineering*, **112**, 967–9829.

Young, P. and Beck, M. (1974). The modelling and control of water quality in a river system. *Automatica*, **10**, 455–468.

Young, P. and Whitehead, P. (1975). *A Recursive Approach to Time-series Analysis for Multivariable Systems.* North Holland, Amsterdam.

Young, P. (1974). Recursive approaches to time series analysis. *Bulletin, Institute of mathematics and its applications*, **10**, 209–224.

Zetterquist, L. (1991). Statistical estimation and interpretation of trends in water quyality time series. *Water Resources Research*, **27**(7), 1637–1648.

X

Ph. D. theses from IMM

1. **Larsen, Rasmus.** (1994). *Estimation of visual motion in image sequences.* *xiv* + 143 pp.
2. **Rygard, Jens Moberg.** (1994). *Design and optimization of flexible manufacturing systems.* *xiii* + 232 pp.
3. **Lassen, Niels Christian Krieger.** (1994). *Automated determination of crystal orientations from electron backscattering patterns.* *xv* + 136 pp.
4. **Melgaard, Henrik.** (1994). *Identification of physical models.* *xvii* + 246 pp.
5. **Wang, Chunyan.** (1994). *Stochastic differential equations and a biological system.* *xxii* + 153 pp.
6. **Nielsen, Allan Aasbjerg.** (1994). *Analysis of regularly and irregularly sampled spatial, multivariate, and multi-temporal data.* *xxiv* + 213 pp.
7. **Ersbøll, Annette Kjær.** (1994). *On the spatial and temporal correlations in experimentation with agricultural applications.* *xviii* + 345 pp.
8. **Møller, Dorte.** (1994). *Methods for analysis and design of heterogeneous telecommunication networks.* Volume 1-2, *xxxviii* + 282 pp., 283-569 pp.
9. **Jensen, Jens Christian.** (1995). *Teoretiske og eksperimentelle dynamiske undersøgelser af jernbanekøretøjer.* ATV Erhvervsforskerprojekt EF 435. *viii* + 174 pp.
10. **Kuhlmann, Lionel.** (1995). *On automatic visual inspection of reflective surfaces.* ATV Erhvervsforskerprojekt EF 385. Volume 1, *xviii* + 220 pp., (Volume 2, *vi* + 54 pp., fortrolig).
11. **Lazarides, Nikolaos.** (1995). *Nonlinearity in superconductivity and Josephson Junctions.* *iv* + 154 pp.
12. **Rostgaard, Morten.** (1995). *Modelling, estimation and control of fast sampled dynamical systems.* *xiv* + 348 pp.
13. **Schultz, Nette.** (1995). *Segmentation and classification of biological objects.* *xiv* + 194 pp.
14. **Jørgensen, Michael Finn.** (1995). *Nonlinear Hamiltonian systems.* *xiv* + 120 pp.
15. **Balle, Susanne M.** (1995). *Distributed-memory matrix computations.* *iii* + 101 pp.
16. **Kohl, Niklas.** (1995). *Exact methods for time constrained routing*

and related scheduling problems. xviii + 234 pp.
17. **Rogon, Thomas.** (1995). *Porous media: Analysis, reconstruction and percolation. xiv* + 165 pp.
18. **Andersen, Allan Theodor.** (1995). *Modelling of packet traffic with matrix analytic methods. xvi* + 242 pp.
19. **Hesthaven, Jan.** (1995). *Numerical studies of unsteady coherent structures and transport in two-dimensional flows.* Risø-R-835(EN) 203 pp.
20. **Slivsgaard, Eva Charlotte.** (l995). *On the interaction between wheels and rails in railway dynamics. viii* + 196 pp.
21. **Hartelius, Karsten.** (1996). *Analysis of irregularly distributed points. xvi* + 260 pp.
22. **Hansen, Anca Daniela.** (1996). *Predictive control and identification - Applications to steering dynamics. xviii* + 307 pp.
23. **Sadegh, Payman.** (1996). *Experiment design and optimization in complex systems. xiv* + 162 pp.
24. **Skands, Ulrik.** (1996). *Quantitative methods for the analysis of electron microscope images. xvi* + 198 pp.
25. **Bro-Nielsen, Morten.** (1996). *Medical image registration and surgery simulation. xxvii* + 274 pp.
26. **Bendtsen, Claus.** (1996). *Parallel numerical algorithms for the solution of systems of ordinary differential equations. viii* + 79 pp.
27. **Lauritsen, Morten Bach.** (1997). *Delta-domain predictive control and identification for control. xxii* + 292 pp.
28. **Bischoff, Svend.** (1997). *Modelling Colliding-pulse mode-locked semiconductor lasers. xxii* + 217 pp.
29. **Arnbjerg-Nielsen, Karsten.** (1997). *Statistical analysis of urban hydrology with special emphasis on rainfall modelling.* Institut for Miljøteknik, DTU. *xiv* + 161 pp.
30. **Jacobsen, Judith L.** (1997). *Dynamic modelling of processes in rivers affected by precipitation runoff. xix* + 213 pp.
31. **Sommer, Helle Mølgaard.** (1997). *Variability in microbiological degradation experiments. - Analysis and case stydy. xiv* + 211 pp.
32. **Ma, Xin.** (1997). *Adaptive extremum control and wind turbine control. xix* + 293 pp.