Pergamon

Wat. Sci. Tech. Vol. 33, No. 2, pp. 177–185, 1996.

0273–1223/96 $15·00 + 0·00

PII:S 0273-1223(96)00199-0

MODELLING THE TRANSIENT IMPACT OF RAIN EVENTS ON THE OXYGEN CONTENT OF A SMALL CREEK

J. L. Jacobsen*, H. Madsen** and P. Harremoës***

* *Institute of Mathematical Modelling, Building 321 DTU, DK-2800 Lyngby, Denmark and PH-Consult, Ordruphøjvej 4, DK-2920 Charlottenlund, Denmark*
** *Institute of Mathematical Modelling, Building 321 DTU, DK-2800 Lyngby, Denmark*
*** *Institute of Environmental Science and Engineering, Building 321 DTU, DK-2800 Lyngby, Denmark and PH-Consult, Ordruphøjvej 4, DK-2920 Charlottenlund, Denmark*

ABSTRACT

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 ultimate 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.

KEYWORDS

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

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 (1975), Whitehead and Young (1975), Thyssen (1980), Cosby (1984), Cosby et al. (1984), Erlandsen and Thyssen (1983)). Recently, new work is emerging (Jacobsen and Voss (1994), Jacobsen and Madsen (1994), 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 venture, 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 (1975) argued that further research on the diurnal variation were 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.

DATA DESCRIPTION

Nearly two months 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 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 depth 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 (1994) and Jacobsen and Madsen (1994). All input and output variables are listed in Table 1. Temperature is measured in deg. Celsius.

The model

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

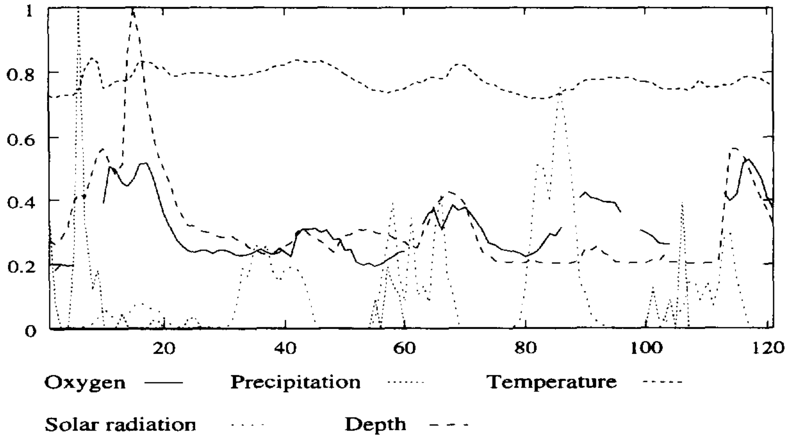

Fig. 1. The scaled measurements of oxygen, solar radiation, water depth, temperature and precipitation. Note the missing values for the oxygen measurements.

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 depth	$[L]$

Symbols are, T: Time; M: Mass; L: Length; and E: Energy

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 (1994), it is argued that the variation of oxygen and organic matter can be described by the following non-linear state space model:

$$\begin{bmatrix} dC \\ dL \end{bmatrix} = \begin{bmatrix} -\frac{K}{h\sqrt{h}} & -K_c \\ K_3 & -K_l \end{bmatrix} \begin{bmatrix} C \\ L \end{bmatrix} dt + \begin{bmatrix} \beta & \frac{\sqrt{C}\,K_b}{h} \\ 0 & \kappa \end{bmatrix} \begin{bmatrix} I \\ P_r \end{bmatrix} dt$$

$$+ \begin{bmatrix} \frac{K}{h\sqrt{h}} C_m(T) \; - R(T) \\ 0 \end{bmatrix} dt + \begin{bmatrix} dw_1 \\ dw_2 \end{bmatrix} \tag{1}$$

$$\text{where} \quad C_m(T) = 14.54 - 0.39T + 0.01T^2 \qquad [mg/l] \tag{2}$$

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

The processes considered for dissolved oxygen are reaeration, consumption from degradation of organic matter, net production from photosynthesis 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

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	$[M\,T^{-1}]$

Symbols are, T: Time; M: Mass; L: length; and E: Energy

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 (2). It was found by trying out all reasonable candidates for the time delay (see Jacobsen and Voss (1994)).

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 (1975), includes 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:

$$C_r(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} C(t) \\ L(t) \end{bmatrix} + e(t) \tag{3}$$

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 (1994) or Jacobsen and Madsen (1994)).

Modelling the photosynthesis

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 photosynthesis, 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.

$$P(I) = \beta\, I \tag{4}$$

$$P(I) = P_m \, E_0 \frac{I}{P_m + E_0 \, I} \tag{5}$$

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 used 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 (4).

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 mentioned above, were introduced into the equations:

$$K_2 = \kappa \, \theta^{T-20} \qquad \theta = 1.015\text{-}1.047 \tag{6}$$

$$\beta = \pi \, \theta^{T-20} \qquad \theta = 1.0241 \tag{7}$$

$$R_T = \rho \, \theta^{T-20} \qquad \theta = 1.0713 \tag{8}$$

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

In other studies, Beck and Young (1975) 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. (2) and (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.

The estimation method

To estimate the parameters of the model, a program (CTLSM) for modelling continuous time systems (Melgaard and Madsen (1993)) 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, $e(t)$ is assumed to be a Gaussian white noise process. Furthermore, it is assumed that $w_1(t)$, $w_2(t)$ and $e(t)$ 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, that 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.

RESULTS AND DISCUSSION

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

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)	

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

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 3 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 time series 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 (1975), Whitehead and Young (1975)). Especially reaeration has 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, has a mean temperature of 8 ^{0}C, while the mean temperature in Harrestrup is nearly 15 ^{0}C.

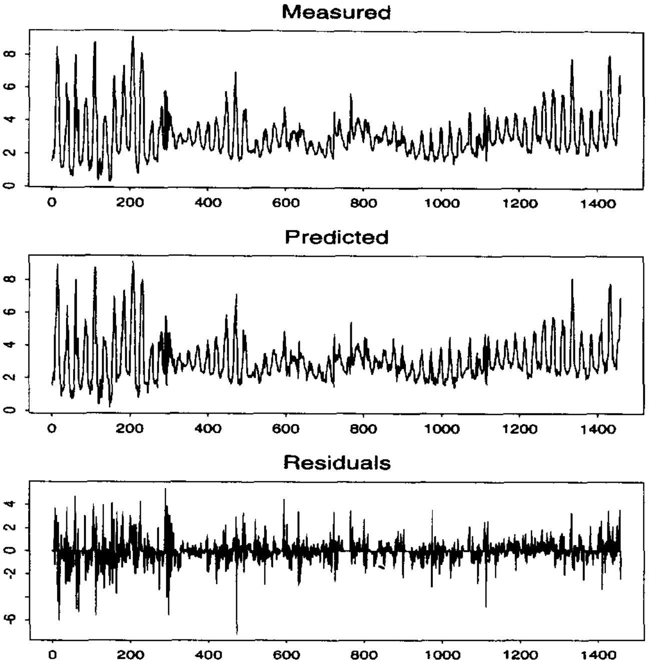

Fig. 2. The prediction and the standardized residuals.

Respiration has not been the subject of most of the known investigations, but it is intuitively known that it should not be zero. In most modelling efforts (deterministic or stochastic) this term has always been considered constant, which seems to be mostly for the sake of obtaining a linear model. Odum (1956) proposed that respiration may be dependent on organic matter, but this has not been tried specifically here. However, the parameters for degradation of organic matter, K_c and K_l, which cause 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 (1993). Furthermore, the cross correlation functions between the residuals and various input signals are calculated.

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 (1994)) 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

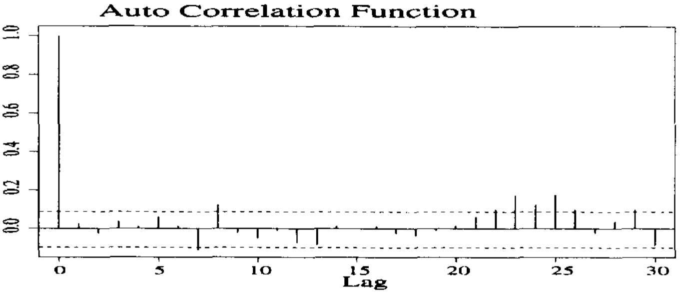

Fig. 3. The autocorrelation (ACF) function for the residuals.

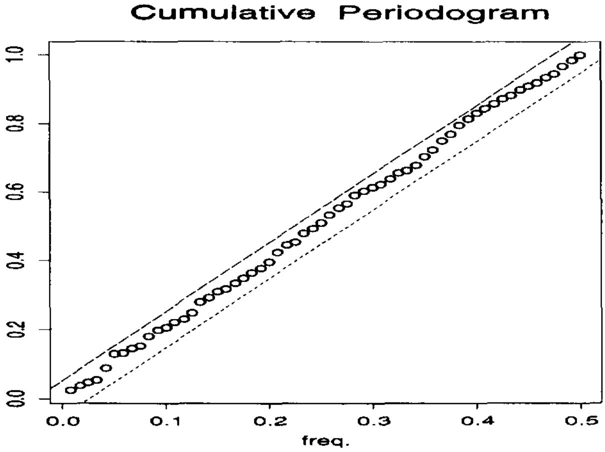

Fig. 4. The cumulative periodogram for the residuals.

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 no macrophytes to speak of 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 did not succeed. 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.

References

Beck, M. and Young, P. (1975). A dynamic model for DO-BOD relationships in a non-tidal stream. *Water Research*, *9*, 769–776.

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.

Dobbins, W. (1964). BOD and oxygen relationships in streams. *Journal of the Sanetary Engineering Division*, ***SA3***, 53–78. Proceedings of the American Society of Civil Engineers.

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.

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 vand-hygiejne).

Jacobsen, J. L. and Madsen, H. (1994). Grey box modelling of oxygen levels in a small stream. *Submitted.*

Jacobsen, J. L. and Voss, L. (1994). Time series analysis and modelling of corrupted data. Master's report, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby, Denmark. no. 12/94.

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.

Melgaard, H. and Madsen, H. (1993). 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.

Odum, H. T. (1956). Primary production in flowing waters. *Limnology Oceanography*, ***1***, 102–117.

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.

Simonsen, J. (1974). *Oxygen fluctuation in streams*. Ph.D. thesis, Department of Technical Hygiene, Technical University of Denmark.

Stigter, J. and Beck, M. (1994). A new approach to the identification of model structure. *EnvironMetrics*, ***5***, 315–333.

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.).

Whitehead, P. and Young, P. (1975). *A Dynamic-Stochastic Model for Water Quality in part of the bedford-Ouse River System. In: Computer Simulation of Water Resources Systems, Ed. G.C. Vansteenkiste*. North Holland, Amsterdam.

Young, P. and Beck, B. (1974). The modelling and control of water quality in a river system. *Automatica*, ***10***, 455–468.

Young, P. (1974). Recursive approaches to time series analysis. *Bulletin, Institute of Mathematics and its Applications*, ***10***, 209–224.