ENVIRONMETRICS
Environmetrics 2001; **12**: 347–356 (DOI: 10.1002/env.464)

A grey-box model describing the hydraulics in a creek

Harpa Jónsdóttir*, Judith L. Jacobsen and Henrik Madsen

Department of Mathematical Modelling, Bldg. 321 DTU, DK-2800 Lyngby, Denmark

SUMMARY

The Saint-Venant equation of mass balance is used here to derive a stochastic lumped model describing the dynamics of the flow in a river. The flow dynamics are described by the evolution of the cross-sectional area of the flow at two locations in the river. The unknown parameters of the model are estimated by combining the physical equations with a set of data. This method is known as grey-box modelling. The data consist of water level measurements, taken every minute at two locations in a river, over a period of nine days. The data are sub-sampled to a sampling period of 15 minutes before further processing, and a maximum likelihood method is used to estimate the parameters of the model.

Three different models were applied to the data set. All three are linear reservoir models with an estimate of the dynamic lateral inflow as a function of precipitation. The first model is a single reservoir model, which proved to be too simple to adequately describe the effect of precipitation. The second model is also a single reservoir model, but the data from the downstream station were translated forward in time, corresponding to a time delay in the system (a retention time). This model responds in a physically reasonable manner to precipitation, capturing very well the flow peaks caused by rain events. The third model is based on two reservoirs, and like the second model, it responds reasonably to precipitation. Its description of the dynamics seems quite good, though it does not capture the flow peaks quite as well as the second model. However, it is shown, that this model statistically provides the best description of the system. Copyright © 2001 John Wiley & Sons, Ltd.

KEY WORDS: grey-box modelling; maximum likelihood estimation; linear reservoir; stochastic hydraulic model.

1. INTRODUCTION

The water quality of a river is mainly affected by the chemical composition of the basin. Chemicals are carried into the river either in dissolved or in particulate form and, in most cases, the chemical profile of the inflow differs from that of the river itself, thus affecting the flora of the river. An adequate modelling of the impact of varying chemical composition on the ecology frequently requires a hydraulic model, which permits estimates of the influx into the system.

In the present work, three hydraulic models are applied and compared. We have used the grey-box approach to establish the models, which are simple lumped models along with stochastic terms, resulting in stochastic ordinary differential equations. The lateral inflow is described as a function of precipitation, and by using Kalman filtering it is possible to perform an estimation of the inflow.

*Correspondence to: H. Jónsdóttir, Department of Mathematical Modelling, Building. 321 DTU, DK-2800 Lyngby, Denmark.

Received 24 February 1999
Accepted 25 August 2000

The data were obtained from a slowly flowing creek, and the ultimate purpose of the data collection was to model the water quality by considering the oxygen level. In order to set up such a model, the hydraulics must be reasonably described. This is facilitated by the hydraulic models proposed in this paper. The only measured variable related to water volume is the depth, and therefore our model is not a true water flow model as we do not know the water flow nor the so-called *Q–h* relations. Instead, the models describe the evolution of the cross-sectional area, because it is more related to flow than the depth.

Three linear reservoir models are tested. The first one is a single reservoir model. The second one is also a single reservoir model but with a time delay (the retention time), and the third model contains two reservoirs. Whitehead and Young (1975) developed a grey-box model for stream flow forecasting, using a deterministic single reservoir model and a black-box rainfall runoff model for the remaining variations. Young and Beck (1974) and Beck and Young (1975) used a linear reservoir model with one reservoir and a transportation delay in a BOD-DO model. Furthermore, Jacobson *et al.* (1997) modelled the water level using a stochastic lumped model with three reservoirs, formulated by stochastic differential equations.

2. THE DATA

The data originate from a creek in North Zealand, approximately 25 km north of Copenhagen. Water level measurements were obtained from two measuring stations which are about 2 km apart. The upstream and downstream stations are referred to as U and D, respectively. There are two rainfall runoffs near the upstream station, one located about 10 m further downstream and the other 2 m upstream. The precipitation measuring station is located approximately 2.5 km upstream from station U. Figure 1 shows a sketch of the area and the measuring stations. The data span a period of nine days in August 1996, with a sampling period of one minute. The goal is to model the overall dynamics, especially the increased water flow during rain events. Data have been low-pass filtered and sub-sampled by averaging over 15-minute periods. The choice of this sampling period was made by comparing the data with the phenomena of interest. The data indicate a retention time of approximately one hour between the two stations. To minimize the risk of information loss, a sampling period of 15 minutes was chosen. The equations below are based on mass balance wherefore it was decided to use the cross sectional areas rather than the water level in the modelling work. Figure 2 shows the cross-sectional areas and the rain intensity.

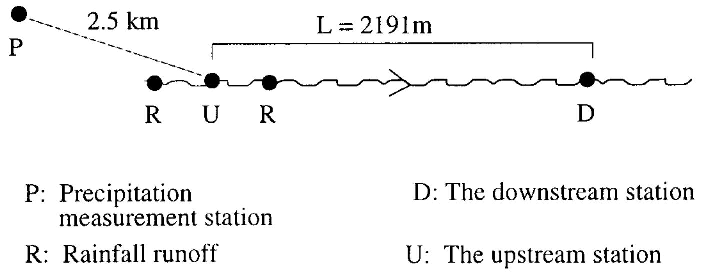
P
2.5 km
L = 2191m
R
U
R
D
P: Precipitation measurement station
D: The downstream station
R: Rainfall runoff
U: The upstream station

Figure 1. Overview of the area

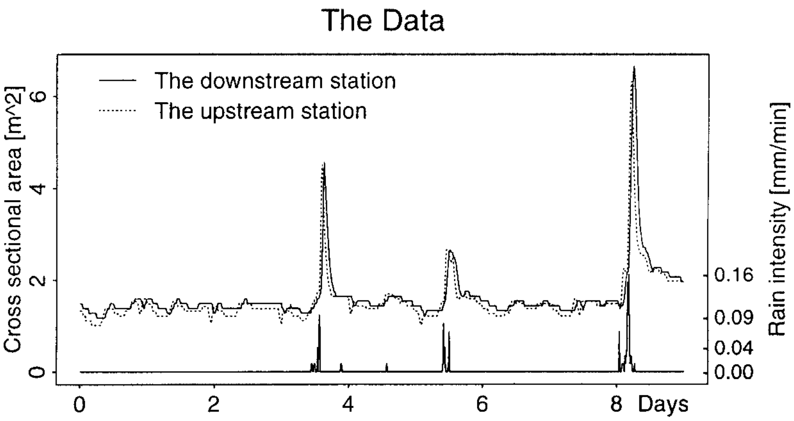

Figure 2. The cross-sectional areas and the rain intensity

3. THE MATHEMATICAL MODEL

The conservation of mass is used to derive a stochastic lumped model describing variations in time of the cross-sectional areas. Our goal is to develop a hydraulic model which can be used as a basis for a water quality model, for example, a model for the oxygen concentration. The only data available related to the hydraulics are water level measurements, so in order to develop a physical model, some simplifications must be made. The flow in a river can be described by the Saint-Venant equation

$$\frac{\partial A(x,t)}{\partial t}+\frac{\partial Q(x,t)}{\partial x}=q(x,t) \tag{1}$$

where A is the cross-sectional area (m^2), Q is the flow of water (m^3/min), q is the lateral inflow per unit length ((m^3/min)/m) = (m^2/min), x is the location along the river (m), and t is the time (min).

Measurements of A are available at two locations in the creek. The flow can be expressed as $Q(x,t)=v(x,t)\ A(x,t)$, where $v(x,t)$ is the average cross-sectional velocity. A constant retention time is assumed. This assumption may be questionable, since an increased water level probably leads to an increase in the average velocity, resulting in a shorter retention time. However, since a period of nine days with relatively small rain events is considered, this simplification is probably reasonable. The velocity $v(x,t)$ is set to be a constant average transportation velocity $v=L/s$, where L is the distance between the two stations and s is the retention time. Equation (1) can thus be rewritten as

$$\frac{\partial}{\partial t}A(x,t)+\frac{L}{s}\frac{\partial}{\partial x}A(x,t)=q(x,t) \tag{2}$$

The next step is to discretize the mass balance equation by $\frac{\partial}{\partial x}A(x,t)\approx\frac{\Delta A(x,t)}{\Delta x}$ and setting x equal to the location of the downstream station and let Δx be the distance L between the two stations. This leads to

$$\frac{\partial}{\partial t}A(x_D,t)+\frac{A(x_D,t)-A(x_U,t)}{s}=q(x_D,t) \tag{3}$$

where x_D is the location of the downstream station and x_U is the location of the upstream station. Using the rule of total differential[1] in Equation (3) and, since this is an Euler description, $\frac{\partial}{\partial t}$ can be replaced with $\frac{\mathrm{d}}{\mathrm{d}t}$, resulting in

$$\frac{\mathrm{d}}{\mathrm{d}t}A(x_D, t) = A(x_U, t)/s - A(x_D, t)/s + q(x_D, t) \tag{4}$$

This equation can also be obtained by simplifying a linear reservoir model, as in Jacobson *et al.* (1997), where a similar model was obtained using that approach. We shall therefore refer to Equation (4) as the single reservoir model. Let the lateral inflow be a function of precipitation, described by a first-order process,

$$\frac{\mathrm{d}}{\mathrm{d}t}q(t) = aq(t) + bP(t) + k \tag{5}$$

where $P(t)$ is the precipitation. In some situations it might be necessary to introduce a time delay in Equation (5) between the rain event and the inflow by using $P(t-\tau)$. This is not the case here, since the precipitation has an almost immediate impact on the lateral inflow through the rainfall runoffs. Finally, a noise term is added to the equations and thereby a stochastic model is obtained. The resulting single reservoir model is expressed in matrix notation as

$$\begin{bmatrix} \mathrm{d}q(t) \\ \mathrm{d}A_D(t) \end{bmatrix} = \begin{bmatrix} a & 0 \\ 1 & -1/s \end{bmatrix} \begin{bmatrix} q(t) \\ A_D(t) \end{bmatrix} \mathrm{d}t + \begin{bmatrix} b & 0 & k \\ 0 & 1/s & 0 \end{bmatrix} \begin{bmatrix} P(t) \\ A_U(t) \\ 1 \end{bmatrix} \mathrm{d}t + \begin{bmatrix} \mathrm{d}w_1(t) \\ \mathrm{d}w_2(t) \end{bmatrix} \tag{6}$$

$$y(t) = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} q(t) \\ A_D(t) \end{bmatrix} + e(t) \tag{7}$$

where the last equation is introduced to describe the fact that only the cross-sectional area is measured, and the measurement $y(t)$ is encountered with the measurement noise $e(t)$. In an abbreviated notation, the equations are written as the following continuous discrete-time state space model,

$$\mathrm{d}\boldsymbol{X}(t) = \boldsymbol{A}\boldsymbol{X}(t)\mathrm{d}t + \boldsymbol{B}\boldsymbol{U}(t)\mathrm{d}t + \mathrm{d}\boldsymbol{w}(t) \tag{8}$$

$$\boldsymbol{Y}(t) = \boldsymbol{C}\boldsymbol{X}(t) + \boldsymbol{e}(t) \tag{9}$$

Equation (8) is known as the system equation which describes the dynamics of the physical system. The vector $\boldsymbol{X}$ is the state vector and $\boldsymbol{U}$ is a vector of inputs. The matrix $\boldsymbol{A}$ characterizes the dynamic behaviour of the system and the matrix $\boldsymbol{B}$ specifies how the input signals enter the system. The noise term $d\boldsymbol{w}(t)$ is the model error, and it is assumed to be a diagonal Wiener process, i.e., the increments are mutually independent and Gaussian distributed. In our formulation, the cross-sectional area at the downstream station and the lateral inflow are state variables, whereas the precipitation and the cross-sectional area at the upstream station are inputs. Equation (9) is known as the measurement equation. The vector $\boldsymbol{Y}(t)$ contains the measured variables and the matrix $\boldsymbol{C}$ specifies which linear combination

[1] $\frac{\mathrm{d}}{\mathrm{d}t}A(x_D, t) = \frac{\partial}{\partial x}A(x_D, t)\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial}{\partial t}A(x_D, t)\frac{\mathrm{d}t}{\mathrm{d}t}$

of the states actually are measured. The vector $\boldsymbol{e}$ contains the measurement errors which are assumed to be mutually independent Gaussian distributed and independent of the model error $d\boldsymbol{w}$.

Inspection of Equations (6)–(7) reveals that the parameter s has the unit minutes and is interpreted as the retention time. The parameter b is independent of time, with the units m^2/mm. 1 mm of rain will thus result immediately in $b\,\mathrm{m}^2$ of lateral inflow, while the extra inflow due to rain declines as described by a/min.

4. PARAMETER ESTIMATION AND MODEL VALIDATION

In this section an outline of the estimation procedure is presented, and some validation methods are described.

The unknown parameters are estimated using a maximum likelihood method. A test of significance of the estimated parameters can therefore be performed by using a t-test. It is well known that the likelihood function for time series data becomes a product of conditional densities. Based on the fact that the model in Equations (6)–(7) is linear, and the assumptions about the Gaussian nature of the error components in Equations (8) and (9), it is easily shown that the conditional densities are Gaussian. The Gaussian distribution is completely characterized by the conditional mean and the conditional variance, which can be calculated recursively by the Kalman filter (see, for example, Harvey (1994)).

Because the data are given in discrete time, the stochastic differential equations have to be integrated through the sampling interval in order to evaluate the likelihood function. Given the constant sampling period τ, the solution to Equations (8) is

$$\boldsymbol{X}(t+\tau) = e^{\boldsymbol{A}\tau}\left(\boldsymbol{X}(t) + \int_t^{t+\tau} \boldsymbol{B}\boldsymbol{U}(s)e^{-(s-t)\boldsymbol{A}}\mathrm{d}s + \int_t^{t+\tau} e^{-(s-t)\boldsymbol{A}}\mathrm{d}\boldsymbol{w}(s)\right) \tag{10}$$

(see Kloeden and Platen (1992) for a solution of stochastic differential equations in the form of Equation (8)). Since the matrix $\boldsymbol{B}$ is a constant, the expression can be simplified by moving $\boldsymbol{B}$ outside the integral.

The input $\boldsymbol{U}(s)$ is measured only at discrete time points, t and $t+\tau$. For all t, the input $\boldsymbol{U}$ inside the interval $[t,\ t+\tau]$ is obtained by linear interpolation. Equation (10) can then be written as

$$\boldsymbol{X}(t+\tau) = \boldsymbol{\phi}(\tau)\boldsymbol{X}(t) + \boldsymbol{B}\boldsymbol{\psi}(\tau)\boldsymbol{U}(t) + \boldsymbol{B}\boldsymbol{\Psi}(\tau)(\boldsymbol{U}(t+\tau) - \boldsymbol{U}(t)) + \boldsymbol{v}(t,\tau) \tag{11}$$

where

$$\begin{aligned} &\boldsymbol{\phi}(\tau) = e^{\boldsymbol{A}\tau};\ \ \boldsymbol{\psi}(\tau) = \int_0^{\tau} e^{-(r-\tau)\boldsymbol{A}}dr;\ \ \boldsymbol{\Psi}(\tau) = \int_0^{\tau} e^{-(r-\tau)\boldsymbol{A}}\frac{r}{\tau}\mathrm{d}r \\ &\boldsymbol{v}(t,\tau) = \int_t^{t+\tau} e^{(t+\tau-s)\boldsymbol{A}}\,\mathrm{d}\boldsymbol{w}(s). \end{aligned} \tag{12}$$

Equation (11) is the system equation of the discrete-discrete state space model originated from the continuous-discrete time space model. The estimation procedure has been implemented in a program called CTLSM (continuous time linear stochastic modelling). See Melgaard and Madsen

(1993) for a description of the program, and Jacobsen and Madsen (1996) for a further description of the method.

The main purpose of validating a model must be to verify whether or not the model describes the physical phenomena, and in a statistical approach, whether the model describes the data. We concentrate on the residuals, which are the one-step prediction errors. The assumptions on the error terms previously mentioned should lead to white noise residuals. Hence, all the well-known tests for white noise residuals from time series analysis can be used (see, e.g., Box and Jenkins (1976)).

5. RESULTS AND DISCUSSION

In this section the results corresponding to the suggested model Equations (6)–(7) are first discussed. These results point to some extensions of the model which are then introduced. The results of the parameter estimation are shown in Table I. The results in the first column correspond to the single reservoir model given by Equations (6)–(7). The parameter b is estimated to be less than zero, which means that the lateral inflow abates and actually becomes negative because of rain. Physical intuition tells us that this cannot be true, and the conclusion is that Equations (6)–(7) are too simple to describe the system. The model must therefore be modified and this is done in two ways. The first modification is as follows: instead of using the input (the cross-sectional area at the upstream station and the precipitation) at time t, the input at time $t - s$ is considered where s is the retention time. The same method is used by Young and Beck (1974) in a DO-BOD model. In the second modification, an unobserved state between the upstream station and the downstream station is introduced, i.e., the reservoir model is extended to contain two reservoirs. Jacobsen *et al.* (1997) presented a grey-box model with several reservoirs using the same data set. The main difference between the models here and the models presented in Jacobsen *et al.* (1997) is that here we have the non-observed lateral inflow as a state variable. Table I shows the parameter estimates for all the models. It is readily seen that all the parameters are significantly different from zero.

Each model will now be discussed. As mentioned before, using the single reservoir model Equations (6)–(7), the 'precipitation response' parameter b is estimated to be negative. The explanation for this is as follows. From Equations (6)–(7) it is seen that the lateral inflow $q(t)$ is a function of precipitation, but it also depends on the value of $A_D(t) - A_U(t)$. Figure 3 shows this difference. It is seen that $A_D(t) - A_U(t)$ is negative at the beginning of the rain events. The reason for this is that almost all the lateral inflow enters the system through the rainfall runoffs located

Table I. Maximum likelihood estimates of the unknown parameters, the numbers in parentheses are the standard deviation of the parameter estimates. All the parameters are significantly different from zero

Parameter	Single reservoir (Equations (6)–(7))	Single reservoir with delay (Equations (13)–(14))	Two reservoirs (Equations (16)–(17))	Units
a	– 0.0293	– 0.0411	– 0.0115	(1/min)
	(0.0042)	(0.0032)	(0.0026)	
b	– 0.0031	0.0271	0.0044	(m^2/mm)
	(0.0011)	(0.0033)	(0.0011)	
s	66.603	22.769	61.33	(min)
	(3.635)	(2.0992)	(1.72)	
k	3.5×10^{-5}	1.04×10^{-4}	2.3×10^{-5}	(m^2/mm)
	(1.1×10^{-5})	(1.78×10^{-5})	(7.3×10^{-6})	

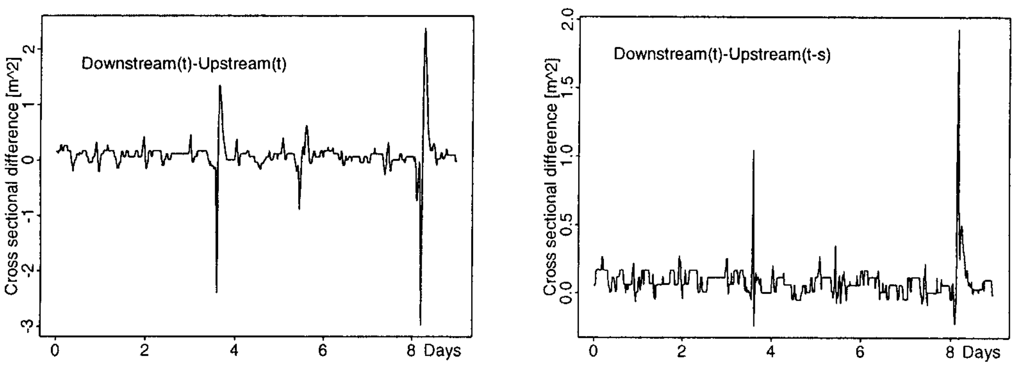

Figure 3. The cross-section difference. To the left is the actual difference $A_D(t) - A_U(t)$ and to the right is the delayed difference, $A_D(t) - A_U(t - s)$, where the retention time s is set to one hour

near the upstream station, and it takes some time for this extra amount of water to reach the downstream station. Hence, the model (6)–(7) obtains the best fit to the data by a negative response to rain.

Figure 4 shows the measured and simulated values of the cross-sectional area and the estimate of the non-observed lateral inflow. Considering the single reservoir model defined by Equations (6)–(7), it is seen that the simulation results are quite accurate under stationary conditions, but the peaks due to rain are not adequately described. The estimated lateral inflow $q(t)$ is far from being realistic. The variations are too sharp and the peaks are too narrow compared with the peaks of the cross-sectional area.

The single reservoir model with the time delay is considered next. In the previous model the retention time was estimated to be 63 minutes. The data intervals are 15 minutes, and consequently the delay is approximated by 4 time steps. Figure 3 shows the cross-sectional difference $A_D(t) - A_U(t - 60)$. This difference does not reach as large negative values as the undelayed difference $A_D(t) - A_U(t)$. The model becomes:

$$\begin{bmatrix} dq(t) \\ dA_D(t) \end{bmatrix} = \begin{bmatrix} a & 0 \\ 1 & -1/s \end{bmatrix} \begin{bmatrix} q(t) \\ A_D(t) \end{bmatrix} \mathrm{d}t + \begin{bmatrix} b & 0 & k \\ 0 & 1/s & 0 \end{bmatrix} \begin{bmatrix} P(t-60) \\ A_U(t-60) \\ 1 \end{bmatrix} \mathrm{d}t + \begin{bmatrix} \mathrm{d}w_1(t) \\ \mathrm{d}w_2(t) \end{bmatrix} \tag{13}$$

$$y(t) = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} q(t) \\ A_D(t) \end{bmatrix} + e(t) \tag{14}$$

The parameter s can no longer be interpreted as a retention time. The 'precipitation response' parameter b is now estimated to be positive, i.e., the model responds to rain by a positive lateral inflow. Figure 4 shows the measured and simulated values of the cross-sectional area, and the estimated lateral inflow $q(t)$. The simulation captures the peaks very well, but still the lateral inflow does not behave as intuitively expected. As in the previous model, the variations are too sharp. Note that the estimated values of $q(t)$ during rain periods are much higher than the in previous model. This is because $q(t)$ has a different interpretation. Here $q(t)$ depends on the value of $A_D(t) - A_U(t - s)$ and this value actually represents the integrated lateral inflow during the retention time. This can be seen by solving

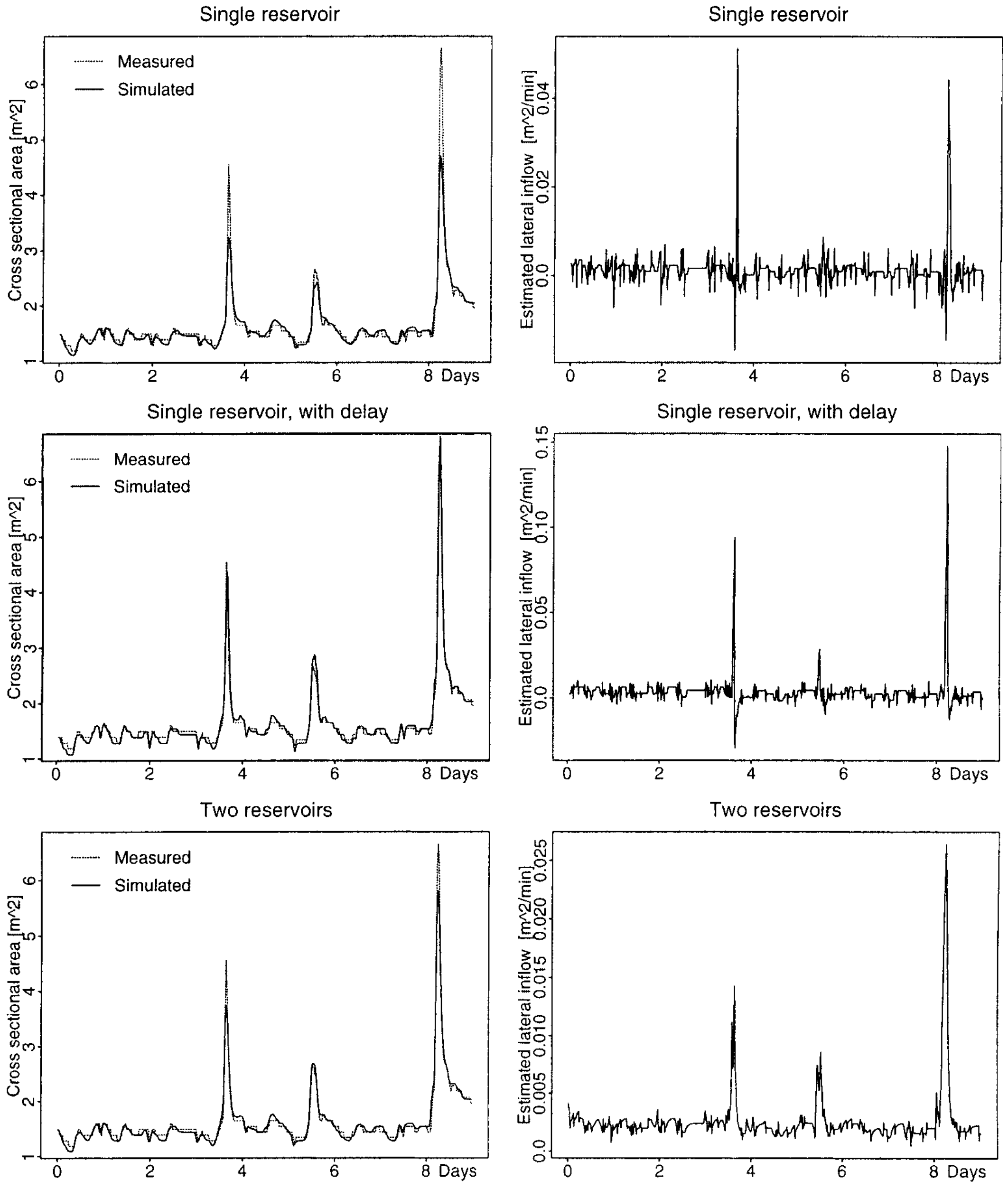

Figure 4. To the left are measurements and simulations using the three models and to the right are the estimated lateral inflows $q(t)$

the Saint-Venant equation of mass balance along with all our simplifications (i.e., Equation (2)), which gives the relation

$$A_D(t) = A_U(t - s) + \int_{t-s}^{t} q(\tau)\mathrm{d}\tau \tag{15}$$

Finally the single reservoir model Equations (6)–(7) is extended to a two reservoir model. This is done by introducing a virtual station between the two measurement stations, denoted $A_F(t)$. The equations then become:

$$\begin{bmatrix} \mathrm{d}q(t) \\ \mathrm{d}A_F(t) \\ \mathrm{d}A_D(t) \end{bmatrix} = \begin{bmatrix} a & 0 & 0 \\ 1 & -2/s & 0 \\ 0 & 2/s & -2/s \end{bmatrix} \begin{bmatrix} q(t) \\ A_F(t) \\ A_D(t) \end{bmatrix} \mathrm{d}t + \begin{bmatrix} b & 0 & k \\ 0 & 2/s & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} P(t) \\ A_U(t) \\ 1 \end{bmatrix} \mathrm{d}t + \begin{bmatrix} \mathrm{d}w_1(t) \\ \mathrm{d}w_2(t) \\ \mathrm{d}w_3(t) \end{bmatrix} \tag{16}$$

$$y(t) = [0 \quad 0 \quad 1] \begin{bmatrix} q(t) \\ A_F(t) \\ A_D(t) \end{bmatrix} + e(t) \tag{17}$$

We have chosen to let the lateral inflow $q(t)$ affect only the first reservoir, since it is known that almost all the transient inflow between the two stations enters the river stretch close to the upstream station. Notice that this is another way of dealing with the time delay. Figure 4 shows the measured cross-sectional area along with the simulation and the estimated lateral inflow. The simulation is quite good but it does not capture the peaks as well as the single reservoir model with time delay. However, the lateral inflow $q(t)$ seems realistic, since it varies around some value close to zero and then increases when it rains. The duration of the extended inflow corresponds to the duration of the increased cross-sectional area. The parameter b is estimated to be positive, which indicates that the system responds to rain by a positive inflow.

For a comparison of the models, several model validation statistics where carried out, and these are shown in Table II. The correlations in time of the single-step prediction errors are studied in the frequency domain but are not shown in the table. A white noise error is equally distributed in the frequency domain and by using a Kolmogorov–Smirnov test in the frequency domain, it was concluded that none of the model errors are white noise processes, although the model with two reservoirs is quite close. White noise test statistics have also been performed by using the Portmanteau lack-of-fit test for white noise, see e.g., Box and Jenkins (1976), as shown in the bottom row of Table II.

Table II. The following model validation statistics are shown: the mean value of the one step prediction error and its standard deviation, the mean value of the simulation error and the Portmanteau lack-of-fit test statistic

	Single reservoir Equations (6)–(7)	Single reservoir with delay Equations (13)–(14)	Two reservoirs Equations (16)–(17)	Units
Mean of prediction error	0.0012	0.0011	0.0006	(m^2)
Variance of prediction error.	0.0012	0.0014	0.0010	(m^2)
Mean of simulation error	0.0187	0.0065	0.0066	(m^2)
Portmanteau test statistic	73.33	203	67.45	

The single reservoir model with a time delay is far from having a white noise prediction error, whereas the simple single reservoir model is much closer. The mean of the prediction error is smallest for the two reservoir model, and even though the simulation seemed to be best in the single reservoir model with time delay, the mean value of the simulation error of the two reservoir model is almost the same. From a statistical point of view, the two reservoir model is surely the best. This is also the case from a physical point of view, since this model is the only one capable of estimating the lateral inflow $q(t)$ reasonably well. Even though there are some unexplained dynamics in the two reservoir model, it is expected that for most practical purposes it is acceptable.

6. CONCLUSION

In this paper, three lumped parameter models describing the variation in the cross-sectional area in a creek are established. These are a single reservoir model, a single reservoir model with a time delay and a two reservoir model. All the models are based on the mass balance equation, and the main simplification is that the retention time is assumed to be constant. All the models provide a dynamic estimate of the lateral inflow, which was not measured, this can be very useful in an environmental context. A comparison of the models shows that the two reservoir model gives a better description than the single reservoir models. The prediction error is close to white noise, which was not the case for the other two models, and the mean value of the prediction error is lowest. The two reservoir model is the only model which provides a physically consistent estimate of the lateral inflow. It is concluded that for most practical purposes the two reservoir model describes the data reasonably well.

In this study it has been found advantageous to use the grey-box approach. Contrary to black-box models, the parameters of a grey-box model have a physical meaning. In contrast to traditional physical modelling, or white-box modelling, we have been able to estimate the coefficients of the differential equations. Furthermore, the stochastic approach makes it possible to provide uncertainty bounds on predictions.

ACKNOWLEDGEMENTS

We thank PH-consult and the Hørsholm City Council for providing data.

REFERENCES

Beck MB, Young PC. 1975. A dynamic model for DO-BOD relationships in a non-tidal stream, *Water Research* **9**: 769–776.

Box GEP, Jenkins GM. 1976. *Time Series Analysis, Forecasting and Control*; Holden-Day: San Francisco.

Harvey AC. 1994. *Forecasting, Structural Time Series Models and the Kalman Filter*; Cambridge University Press: Cambridge.

Jacobsen JL, Madsen H. 1996. Grey Box Modelling of oxygen levels in a small stream. *Environmetrics* **7**: 109–121.

Jacobsen JL, Madsen H, Harremöes P. 1997. A stochastic model for two-station hydraulics exhibiting transient impact. *Water, Science and Technology* **36**(5): 19–26.

Kloeden PE, Platen E. 1992. *Numerical Solutions of Stochastic Differential Equations*, 2nd edn. Applications of Mathematics, Stochastic Modelling and Applied Probability. Springer-Verlag: Heidelberg.

Melgaard H, Madsen H. 1993. CTLSM version 2.6 – a program for parameter estimation in stochastic differential equations, Technical Report No. 1/1993, IMM, Building 321, DTU, DK-2800 Lyngby.

Whitehead PG, Young PC. 1975. A dynamic-stochastic model for Water quality in part of the Bedford–Ouse river system. In *Computer Simulation of Water Resources Systems*, Vansteenkiste GC (ed). North Holland: Amsterdam; 417–438.

Young PC, Beck MB. 1974. The modelling and control of water quality in a river system. *Automatica* **10**: 455–468.