ENVIRONMETRICS
Environmetrics 2000; **11**: 1–12

GREY BOX MODELING OF FIRST FLUSH AND INCOMING WASTEWATER AT A WASTEWATER TREATMENT PLANT

HENRIK BECHMANN[1,2]*, HENRIK MADSEN[1], NIELS KJØLSTAD POULSEN[1]
AND MARINUS K. NIELSEN[2]

[1]*Department of Mathematical Modelling, Technical University of Denmark, Building 321, DK-2800 Lyngby, Denmark*
[2]*Krüger A/S, Gladsaxevej 363, DK-2860 Søborg, Denmark*

SUMMARY

On-line measurements of turbidity, UV absorption and flow in the inlet to a Danish wastewater treatment plant are used to establish a dynamic model of the deposition of pollutants in the sewer system and the pollutant mass flow to the treatment plant. The modelling is made using the grey box approach, which is a statistical method that uses known physical relations to formulate the model. The dynamics of the sewer are modelled by means of continuous time stochastic differential equations combined with dry weather diurnal pollutant mass flows. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS on-line measurements; sewer systems; wastewater treatment plant; dynamic systems; grey box modelling; stochastic differential equations; first flush; diurnal pollutant load

1. INTRODUCTION

Modelling of sewer systems is necessary to gain a better understanding of the dynamics of the system, and to verify if deposition of pollutants occurs. If deposition occurs then the model can be used to predict and quantify the first flushes. When the dynamics of the first flushes are quantified, it can be used to calculate the necessary size of a storage basin in the sewer system or at the wastewater treatment plant.

When modelling sewer systems, these are often described using a configuration of storage volumes and pipes with different dimensions. This approach results in a mathematical model with many parameters. Unfortunately the data seldom provides sufficient information to uniquely identify all parameters.

In this paper a data-based grey box modelling approach is used. A grey box model is a stochastic model which only describes the most important physical relations. The benefits of this approach are that the resulting model has few parameters and states which are possible to identify, using appropriate statistical methods. Due to the small number of parameters, few computational resources are needed to estimate the parameters, which makes the parameter estimation applicable for on-line use, e.g. for on-line control purposes (Carstensen *et al.* 1996).

* Correspondence to: H. Bechmann, Department of Mathematical Modelling, Technical University of Denmark, Building 321, DK-2800 Lyngby, Denmark.

CCC 1180–4009/2000/010001–12$17·50

Received 20 April 1998
Accepted 8 April 1999

The present work is based on on-line measurements of UV absorption, turbidity and flow. The measurements of UV absorption and turbidity are used to estimate the actual level of chemical oxygen demand (COD) and suspended solids (SS) in the wastewater (Kanaya *et al*. 1985; Matsché and Stumwöhrer 1996; Ruban *et al*. 1993; Nowack and Ueberbach 1995). Based on these data a model including the diurnal profiles of COD and SS mass flow, as well as the amount of COD and SS deposits in the sewer, is proposed and identified. The model parameters are estimated using one set of data, and cross validated using another data set.

The deposits, which are not practically measurable, are states in the model and hence estimated by means of the procedure. By use of the model, the estimated amounts of deposited pollutants in the sewer are identified and shown.

2. THE MEASUREMENT SYSTEM

A measuring box, developed by Krüger A/S, Denmark, collected measurements from the inlet to the wastewater treatment plant of the town of Skive in central Jutland, Denmark, in the early spring of 1997. The Krüger measuring box is a compact and portable unit which consists of a datalogger, a Dr. Lange UV absorbance sensor model LXV 109 and a Dr. Lange SOLITAXplus LXV 121 turbidity sensor. Furthermore, inlet flow estimates were collected from the supervisory control and data acquisition (SCADA) system of the WWTP. The flow estimates were computed in the SCADA system on the basis of measurements from the inlet pumping station. Off-line measurements of the precipitation and laboratory analyses of COD and SS were also available.

The influent to Skive WWTP is separated into industrial and municipal wastewater. The sensors were placed in the municipal wastewater.

3. COD AND SS

Linear relationships between the concentration of COD, denoted C_{COD}, and UV absorption (UV) and between the concentration of SS, denoted C_{SS}, and turbidity (Turb) have been suggested by several authors (Kanaya *et al*. 1985; Ruban *et al*. 1993; Nowack and Ueberbach 1995; Matsché and Stumwöhrer 1996), i.e.

$$C_{COD} = \alpha_{\mathrm{U}}\mathrm{UV} + \beta_{\mathrm{U}} \tag{1}$$

$$C_{\mathrm{SS}} = \alpha_{\mathrm{T}}\mathrm{Turb} + \beta_{\mathrm{T}} \tag{2}$$

where α_{U}, β_{U}, α_{T} and β_{T} are constants. The concentrations of COD and SS are measured in gO_2/m^3 and gSS/m^3. The models are calibrated on the basis of laboratory measurements of C_{COD} and C_{SS} and the corresponding on-line measurements of UV and Turb. When the concentrations are multiplied by the wastewater flow Q (in m^3/h), the results are the mass flows of COD and SS in gO_2/h and gSS/h, respectively.

4. A DYNAMICAL MODEL OF THE DEPOSITED POLLUTANTS

It is assumed that pollutants, quantified as masses of COD and SS, deposit gradually in the sewer system and on impervious areas of the catchment in dry weather (low wastewater flow). Similarly, it is assumed that the deposited pollutants are flushed out during rain incidents and into the inlet of the treatment plant. This means that the model does not explicitly describe reactions occurring in the sewers, but these processes are in part accounted for through the statistical calibration of the model.

The deposited amounts of COD and SS are denoted X_{COD} and X_{SS}, respectively. The time derivatives, $\mathrm{d}X_{\mathrm{COD}}/\mathrm{d}t$ and $\mathrm{d}X_{\mathrm{SS}}/\mathrm{d}t$ are the growth rates at which COD and SS are built up in the depots. The growth rates are assumed to be described by the first order linear differential equations:

$$\frac{\mathrm{d}X_{\mathrm{COD}}}{\mathrm{d}t} = a_{\mathrm{COD}}(X_{\mathrm{COD}} - \bar{X}_{\mathrm{COD}}) + b_{\mathrm{COD}}(Q - \bar{Q}) \tag{3}$$

$$\frac{\mathrm{d}X_{\mathrm{SS}}}{\mathrm{d}t} = a_{\mathrm{SS}}(X_{\mathrm{SS}} - \bar{X}_{\mathrm{SS}}) + b_{\mathrm{SS}}(Q - \bar{Q}) \tag{4}$$

where Q is the wastewater flow in the inlet and $\bar{X}_{\mathrm{COD}}$ and $\bar{Q}$ are the mean values of X and Q. The proposed model is thus a simple first order storage model. When the parameters a_{COD}, b_{COD}, a_{SS} and b_{SS} are assumed to be negative constants, a flow larger than the mean flow will cause a decrease of the amount of deposits and a flow lower than the mean flow will cause an increase of the deposits. Likewise a deposited amount of pollutants larger than the mean will cause a decrease in the corresponding growth rate, and vice versa.

Furthermore we assume that the pollutant mass flow entering the sewer system is following a diurnal profile and that the rain water does not contain any COD and SS. The diurnal profiles are modelled by periodic functions with a 24 h period, which are described by nth order harmonic functions. A simple mass balance of the sewer system shows that the mass flows of COD and SS at the WWTP are the corresponding diurnal profile minus the contribution to the depots. The pollutant mass flows $Q_{\mathrm{COD}} = Q \times \mathrm{COD}$ and $Q_{\mathrm{SS}} = Q \times \mathrm{SS}$ in the inlet to the WWTP are then modelled by:

$$Q_{\mathrm{COD}} = a_0 + \sum_{k=1}^{n}\left(a_k \sin\left(2\pi k \frac{t}{24\mathrm{h}}\right) + b_k \cos\left(2\pi k \frac{t}{24\mathrm{h}}\right)\right) - \frac{\mathrm{d}X_{\mathrm{COD}}}{\mathrm{d}t} \tag{5}$$

$$Q_{\mathrm{SS}} = c_0 + \sum_{k=1}^{n}\left(c_k \sin\left(2\pi k \frac{t}{24\mathrm{h}}\right) + d_k \cos\left(2\pi k \frac{t}{24\mathrm{h}}\right)\right) - \frac{\mathrm{d}X_{\mathrm{SS}}}{\mathrm{d}t} \tag{6}$$

where a_0, a_k, b_k, c_0, c_k and d_k ($1 \leqslant k \leqslant n$), are the parameters of the harmonic functions. Combining (3) with (5) and (4) with (6) leads to:

$$\begin{aligned} Q_{\mathrm{COD}} = & -a_{\mathrm{COD}}(X_{\mathrm{COD}} - \bar{X}_{\mathrm{COD}}) - b_{\mathrm{COD}}(Q - \bar{Q}) \\ & + a_0 + \sum_{k=1}^{n}\left(a_k \sin\left(2\pi k \frac{t}{24\mathrm{h}}\right) + b_k \cos\left(2\pi k \frac{t}{24\mathrm{h}}\right)\right) \end{aligned} \tag{7}$$

$$\begin{aligned} Q_{\mathrm{SS}} = & -a_{\mathrm{SS}}(X_{\mathrm{SS}} - \bar{X}_{\mathrm{SS}}) - b_{\mathrm{SS}}(Q - \bar{Q}) \\ & + c_0 + \sum_{k=1}^{n}\left(c_k \sin\left(2\pi k \frac{t}{24\mathrm{h}}\right) + d_k \cos\left(2\pi k \frac{t}{24\mathrm{h}}\right)\right) \end{aligned} \tag{8}$$

In order to use a matrix notation we introduce **X**, **U** and **Y** as the state vector, the input vector and the observation vector, respectively; i.e.:

$$\mathbf{X} = [X_{\mathrm{COD}} - \bar{X}_{\mathrm{COD}}, X_{\mathrm{SS}} - \bar{X}_{\mathrm{SS}}]' \tag{9}$$

$$\mathbf{U} = \left[Q - \bar{Q}, 1, \sin\left(2\pi \frac{t}{24\mathrm{h}}\right), \cos\left(2\pi \frac{t}{24\mathrm{h}}\right), \ldots, \cos\left(2\pi k \frac{t}{24\mathrm{h}}\right)\right]' \tag{10}$$

$$\mathbf{Y} = [Q_{\mathrm{COD}}, Q_{\mathrm{SS}}]' \tag{11}$$

and the matrices **A**, **B**, **C** and **D** defined by:

$$\mathbf{A} = \begin{bmatrix} -a_{\mathrm{COD}} & 0 \\ 0 & -a_{\mathrm{SS}} \end{bmatrix} \tag{12}$$

$$\mathbf{B} = \begin{bmatrix} -b_{\mathrm{COD}} & 0 & 0 & 0 & \ldots & 0 \\ -b_{\mathrm{SS}} & 0 & 0 & 0 & \ldots & 0 \end{bmatrix} \tag{13}$$

$$\mathbf{C} = \begin{bmatrix} a_{\mathrm{COD}} & 0 \\ 0 & a_{\mathrm{SS}} \end{bmatrix} \tag{14}$$

$$\mathbf{D} = \begin{bmatrix} b_{\mathrm{COD}} & a_0 & a_1 & b_1 & a_2 & b_2 & \ldots & a_k & b_k \\ b_{\mathrm{SS}} & c_0 & c_1 & d_1 & c_2 & d_2 & \ldots & c_k & d_k \end{bmatrix} \tag{15}$$

After addition of a noise term (3) and (4) can now be described by the stochastic differential equation:

$$\mathrm{d}\mathbf{X} = \mathbf{AX}\,\mathrm{d}t + \mathbf{BU}\,\mathrm{d}t + \mathrm{d}\mathbf{w}(t) \tag{16}$$

where the stochastic process $\mathbf{w}(t)$ is assumed to be a vector Wiener process. The noise term is included to describe the deviations between the model equations (3) and (4) and the true system.

The observation equation in the matrix notation is found by combining (3) with (5) and (4) with (6):

$$\mathbf{Y}(t) = \mathbf{CX}(t) + \mathbf{DU}(t) + \mathbf{e}(t) \tag{17}$$

Here the term $\mathbf{e}(t)$ is the measurement error, which is assumed to be a zero mean Gaussian white noise sequence independent of $\mathbf{w}(t)$.

5. THE PARAMETER ESTIMATION METHOD

This section briefly describes the method used to estimate the parameters of the stochastic differential equation (16) for the dynamics of the sewer system. The estimation method is a maximum likelihood method for estimating parameters in stochastic differential equations based on discrete time data. For a more detailed description of the method we refer to Madsen and Melgaard (1991) or Melgaard and Madsen (1993).

The observations are given in discrete time, and, in order to simplify the notation, we assume that the time index t belongs to the set $\{0, 1, 2, \ldots, N\}$, where N is the number of observations. Introduce

$$\mathcal{Y}(t) = [\mathbf{Y}(t), \mathbf{Y}(t-1), \ldots, \mathbf{Y}(1), \mathbf{Y}(0)]' \tag{18}$$

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

Using the matrix notation, the continuous time stochastic differential equation describing the dynamics of the sewer system can be written as the so-called Itô differential equation (Øksendal 1995)

$$\mathrm{d}\mathbf{X}(t) = f(\mathbf{X}, \mathbf{U}, t)\,\mathrm{d}t + \mathbf{G}(\mathbf{U}, t)\,\mathrm{d}\mathbf{w}(t) \tag{19}$$

where $\mathbf{X}$ is the state vector, $\mathbf{U}$ an input (e.g. control) vector, $\mathbf{w}$ a vector standard Wiener process (see e.g. Kloeden and Platen 1995), and f and $\mathbf{G}$ are known functions. The matrix $\mathbf{G}(\mathbf{U}, t)$ describes any input or time dependent variation related to how the variation generated by the Wiener process enters the system.

For the observations we assume the discrete time relation

$$\mathbf{Y}(t) = h(\mathbf{X}, \mathbf{U}, t) + \mathbf{e}(t) \tag{20}$$

where $\mathbf{e}(t)$ is assumed to be a Gaussian white noise sequence independent of $\mathbf{w}$. All the unknown parameters, denoted by the vector $\boldsymbol{\theta}$, are embedded in the continuous-discrete time state space model (equations (19) and (20)).

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

$$\begin{aligned} L'(\boldsymbol{\theta}; \mathcal{Y}(N)) &= p(\mathcal{Y}(N)|\boldsymbol{\theta}) \\ &= p(\mathbf{Y}(N)|\mathcal{Y}(N-1), \boldsymbol{\theta})p(\mathcal{Y}(N-1)|\boldsymbol{\theta}) \\ &= \left(\prod_{t=1}^{N} p(\mathbf{Y}(t)|\mathcal{Y}(t-1), \boldsymbol{\theta})\right) p(\mathbf{Y}(0)|\boldsymbol{\theta}) \end{aligned} \tag{21}$$

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

In order to evaluate the likelihood function, it is assumed that all the conditional densities are Gaussian. In the case of a linear state space model as described by (16) and (17), it is easily shown that the conditional densities are actually Gaussian (Madsen and Melgaard 1991). In the more general non-linear case the Gaussian assumption is an approximation.

The Gaussian distribution is completely characterized by the mean and covariance. Hence, in order to parameterize the conditional distribution, we introduce the conditional mean and the conditional covarance as

$$\hat{\mathbf{Y}}(t|t-1) = E[\mathbf{Y}(t)|\mathcal{Y}(t-1), \boldsymbol{\theta}] \text{ and } \mathbf{R}(t|t-1) = V[\mathbf{Y}(t)|\mathcal{Y}(t-1), \boldsymbol{\theta}] \tag{22}$$

respectively. It should be noted that these correspond to the one-step prediction and the associated covariance, respectively. Furthermore, it is convenient to introduce the one-step prediction error (or innovation)

$$\epsilon\,(t) = \mathbf{Y}(t) - \hat{\mathbf{Y}}(t|t-1) \tag{23}$$

For calculating the one-step prediction and its variance, an iterated extended Kalman filter is used. The extended Kalman filter is simply based on a linearization of the system equation (19) around the current estimate of the state (see Gelb 1974). The iterated extended Kalman filter is obtained by local iterations of the linearization over a single sample period.

Using (21)–(23) the conditional likelihood function (conditioned on $\mathbf{Y}(0)$) becomes

$$L(\boldsymbol{\theta}; \mathcal{Y}(N)) = \prod_{t=1}^{N}\left((2\pi)^{-m/2} \det \mathbf{R}(t|t-1)^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{\epsilon}(t)'\mathbf{R}(t|t-1)^{-1}\boldsymbol{\epsilon}(t)\right)\right) \tag{24}$$

where m is the dimension of the $\mathbf{Y}$ vector. Traditionally, the logarithm of the conditional likelihood function is considered

$$\log L(\boldsymbol{\theta}; \mathcal{Y}(N)) = -\frac{1}{2}\sum_{t=1}^{N}\left(\log \det \mathbf{R}(t|t-1) + \boldsymbol{\epsilon}(t)'\mathbf{R}(t|t-1)^{-1}\boldsymbol{\epsilon}(t)\right) + \text{const} \tag{25}$$

The maximum likelihood estimate (ML-estimate) is the set $\hat{\boldsymbol{\theta}}$, which maximizes the likelihood function. Since it is not, in general, possible to optimize the likelihood function analytically, a numerical method has to be used. A reasonable method is the quasi-Newton method.

An estimate of the uncertainty of the parameters is obtained by the fact that the ML-estimator is asymptotically normally distributed with mean $\boldsymbol{\theta}$ and covariance

$$\mathbf{D} = \mathbf{H}^{-1} \tag{26}$$

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

$$\{h_{lk}\} = -E\left[\frac{\partial^2}{\partial\theta_l\partial\theta_k}\log L(\boldsymbol{\theta}; \mathcal{Y}(N))\right] \tag{27}$$

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

$$\{h_{lk}\} \approx -\left(\frac{\partial^2}{\partial\theta_l\partial\theta_k}\log L(\boldsymbol{\theta}; \mathcal{Y}(N))\right)_{|\theta=\hat{\theta}} \tag{28}$$

The above equation can be used for estimating the variance of the parameter estimates. The variances serves as a basis for calculating t-test values for test under the hypothesis that the parameter is equal to zero. Finally, the correlation between the parameter estimates is readily found based on the covariance matrix $\mathbf{D}$.

6. RESULTS AND DISCUSSION

In dry weather the relations between COD and UV absorption and between SS and turbidity were found by regression analysis to be:

$$C_{COD} = 5{\cdot}0\frac{\text{g}}{\text{m}^3\text{m}^{-1}}\text{UV} - 26\frac{\text{g}}{\text{m}^3} \tag{29}$$

$$C_{\text{SS}} = 1{\cdot}52\frac{\text{g}}{\text{m}^3\text{FTU}}\text{Turb} - 7{\cdot}8\frac{\text{g}}{\text{m}^3} \tag{30}$$

with degrees of explanation $R^2 = 0{\cdot}90$ and $R^2 = 0{\cdot}85$, respectively. In dry weather equations (29) and (30) describe the relations well, and it is assumed that it is reasonable to extrapolate to rainy situations. More thorough research will have to be carried out to refine the relationships between the measurements of UV absorption and turbidity, and COD and SS, respectively. This investigation could include flow dependent parameter values. Such an approach will increase the need for laboratory analysis of COD and SS enormously.

The parameters in the model were estimated using the CTLSM program (Madsen and Melgaard 1991; Melgaard and Madsen 1991). Different orders of the harmonic functions were tried, and it was found that a second order harmonic was reasonable in describing the diurnal profiles for both COD and SS mass flow. The estimated parameters are shown in Table I, together with the estimated standard deviations of the parameter estimates.

It was found that c_2 was not significantly different from zero, and it was therefore excluded from the final estimation.

In the model presented it is not possible to estimate the mean level of the non-measured deposited pollutants $\bar{X}_{\mathrm{COD}}$ and $\bar{X}_{\mathrm{SS}}$. In the estimations these were fixed at zero, meaning that X_{COD} and X_{SS} are differences from unknown mean values. This does not put any limitations on the use of the model, as it is still possible to quantify the amount of pollutants in a first flush.

Validation of the resulting model was done by cross-validation, i.e. by applying the model on a data set (the validation data set), which differs from that used for the parameter estimation. Using all the estimated parameter values, except for the initial states, the initial values of the states for the validation data set were found by the estimation software. In Figures 1 and 2 the measured

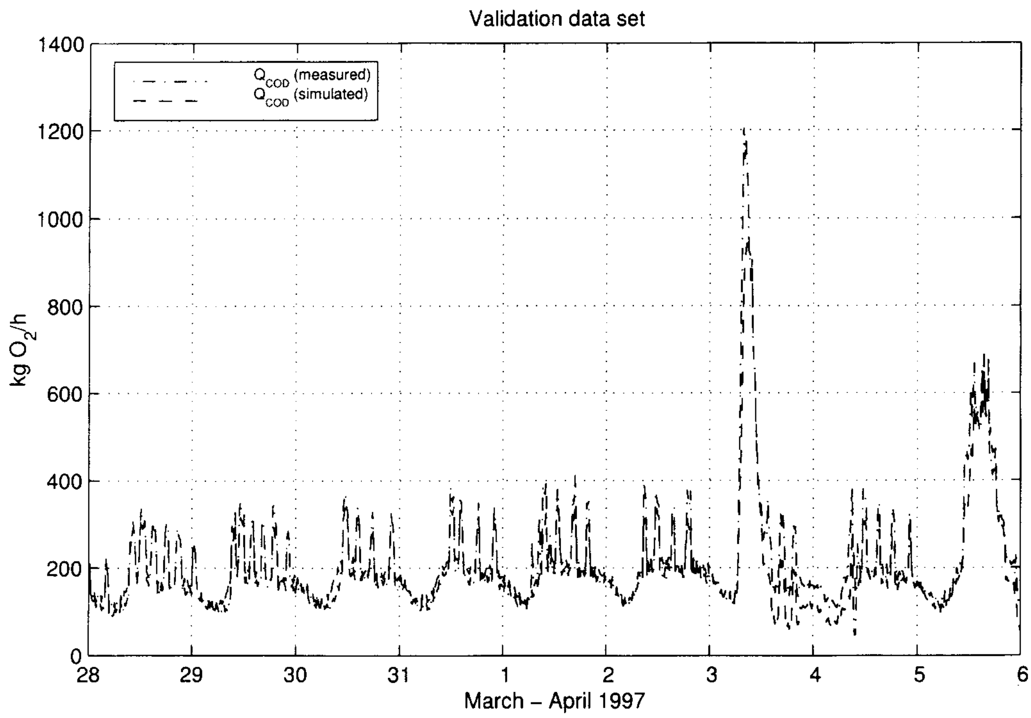

Figure 1. Measured and simulated COD mass flow – validation data set

Table I. Maximum likelihood estimates of the parameters of the system

	a_{COD} (h^{-1})	b_{COD} (kgO_2/m^3)	a_0 (kgO_2/h)	a_1 (kgO_2/h)	b_1 (kgO_2/h)	a_2 (kgO_2/h)	b_2 (kgO_2/h)	a_{SS} (h^{-1})	b_{SS} ($kgSS/m^3$)	c_0 ($kgSS/h$)	c_1 ($kgSS/h$)	d_1 ($kgSS/h$)	c_2 ($kgSS/h$)	d_2 ($kgSS/h$)
Estimate	−0·05432	−0·2969	198·3	−26·58	−16·05	−6·710	20·76	−0·03597	−0·1962	62·14	−12·27	−6·008	–	2·713
Standard deviation	0·00378	0·0041	1·2	1·50	1·62	1·560	1·47	0·00259	0·0032	0·71	1·09	1·061	–	1·054

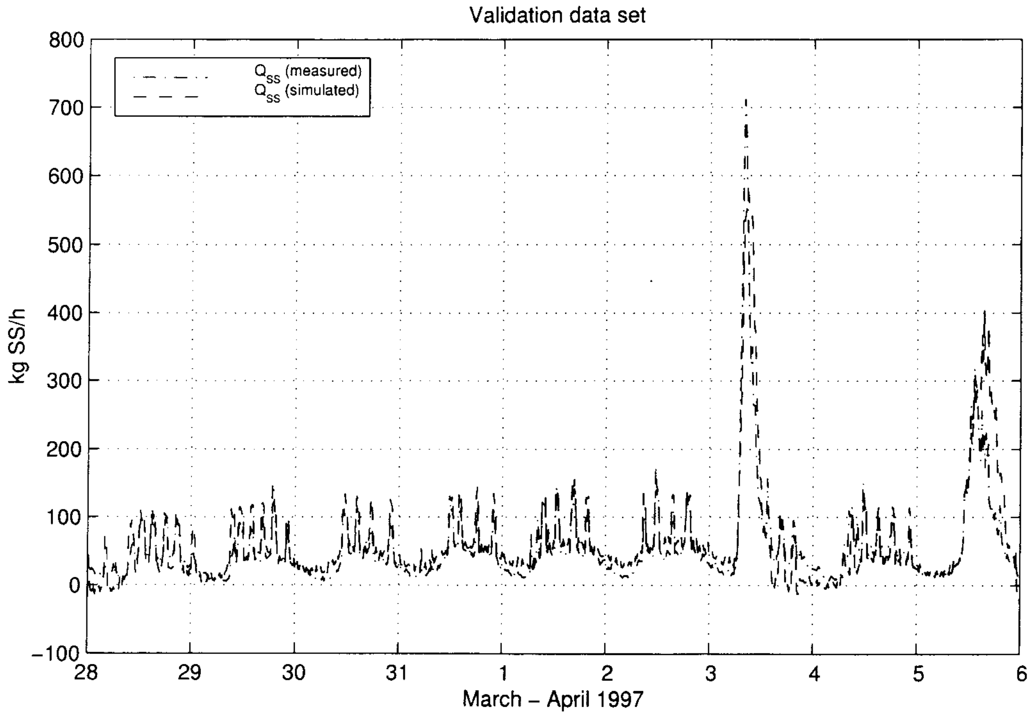

Figure 2. Measured and simulated SS mass flow – validation data set

and simulated pollutant mass flows are shown. They are seen to be in good agreement. Furthermore tests on the white noise assumption based on the autocorrelations and on the cumulative residual periodograms were carried out. It was seen that the white noise assumptions were not perfectly satisfied, but since the cross-validations shown in Figures 1 and 2 are very fine, it is concluded that the model describes the data well.

In Figures 3 and 4 the estimated amounts of deposited COD and SS are shown for the estimation data set and for the validation data set, and it is clear that COD and SS are deposited in dry weather periods and flushed out during rain incidents. From Figure 4 it appears that the amount of COD first flush during the rain incident on April 3rd is approximately 2500 kgO_2 and that the SS first flush is approximately 1800 kgSS. A comparison of these amounts with the mean loads of COD and SS to the WWTP of 5030 kgO_2/24 h and 1527 kgSS/24 h shows that this first flush contains approximately half a 24-h load of COD and 20% more than a 24-h load of SS. This is considered realistic, as it is expected that COD stabilizes in the sewer depots. When COD stabilizes in the sewer the amount of COD in a first flush compared to the normal load of COD will be smaller than the amount of SS compared to the normal load of SS, as SS are not expected to stabilize between rain incidents.

The time constants $\tau_{COD} = -1/a_{COD}$ and $\tau_{SS} = -1/a_{SS}$ are found to be $\tau_{COD} = 18{\cdot}4$ h and $\tau_{SS} = 27{\cdot}8$ h. This means that after a considerable change in the flow, the masses of COD and SS in the sewer depots will reach 63% of the equilibrium in 18·4 and 27·8 h, respectively. The time constants are thus found to be reasonable.

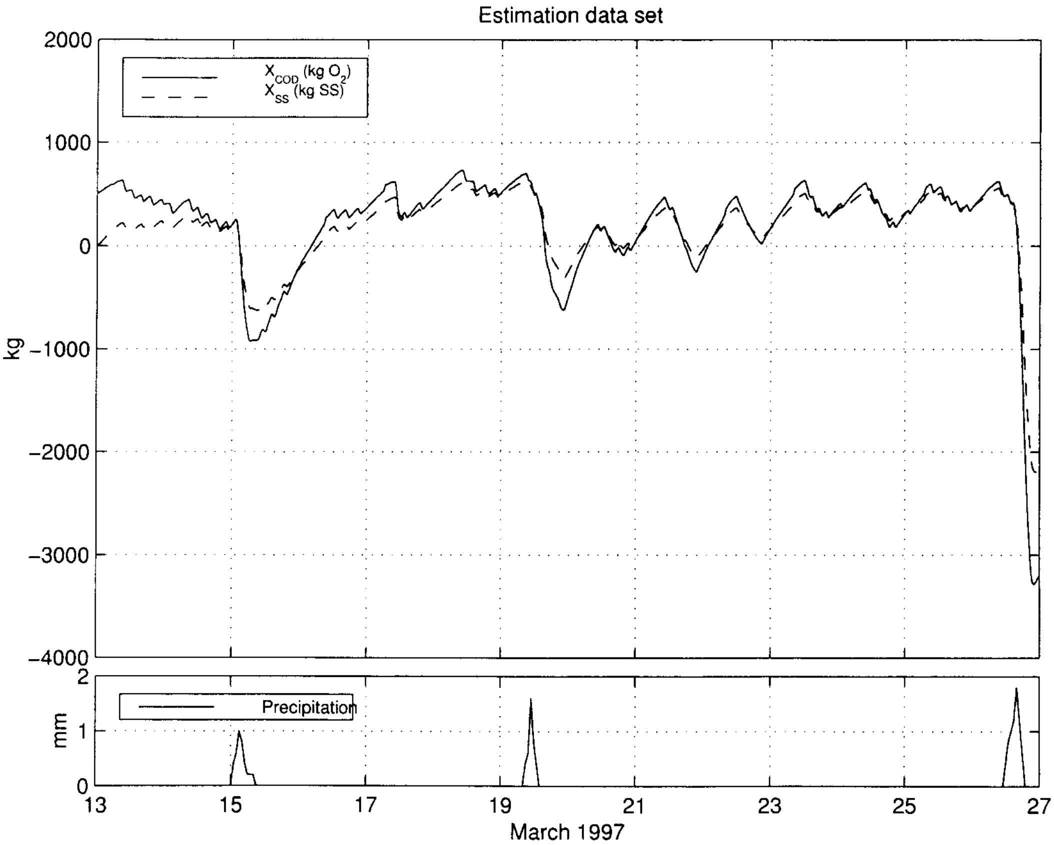

Figure 3. Estimated amounts of COD and SS deposits in the sewer system – estimation data set

7. CONCLUSION

In this paper a stochastic grey box model of the deposition of pollutants in the sewer system and on impervious areas of the catchment is proposed. Furthermore, the model provides a characterization of the influent wastewater to the WWTP. The parameters of the model are estimated using the maximum likelihood method. The modelling is based on measurements of concentrations of COD and SS, which are calculated from measurements of UV absorption and turbidity. It is concluded that the models, which relate the measurements of UV absorption and turbidity to COD and SS concentrations, describe the relations well in dry weather situations, and it is assumed that it is reasonable to extrapolate to rainy conditions.

Using the grey box approach, it is possible to identify a model for a complex dynamical system based on simple physical assumptions combined with statistical modelling tools. The model is formulated in continuous time with discrete time measurements and due to the rather small number of parameters, the model is operational for on-line applications.

The model makes it possible to estimate the amount of COD and SS in a first flush during a rain incident. It is found that these amounts are dependent on the time interval since the previous rain event and on the intensity of the current rain.

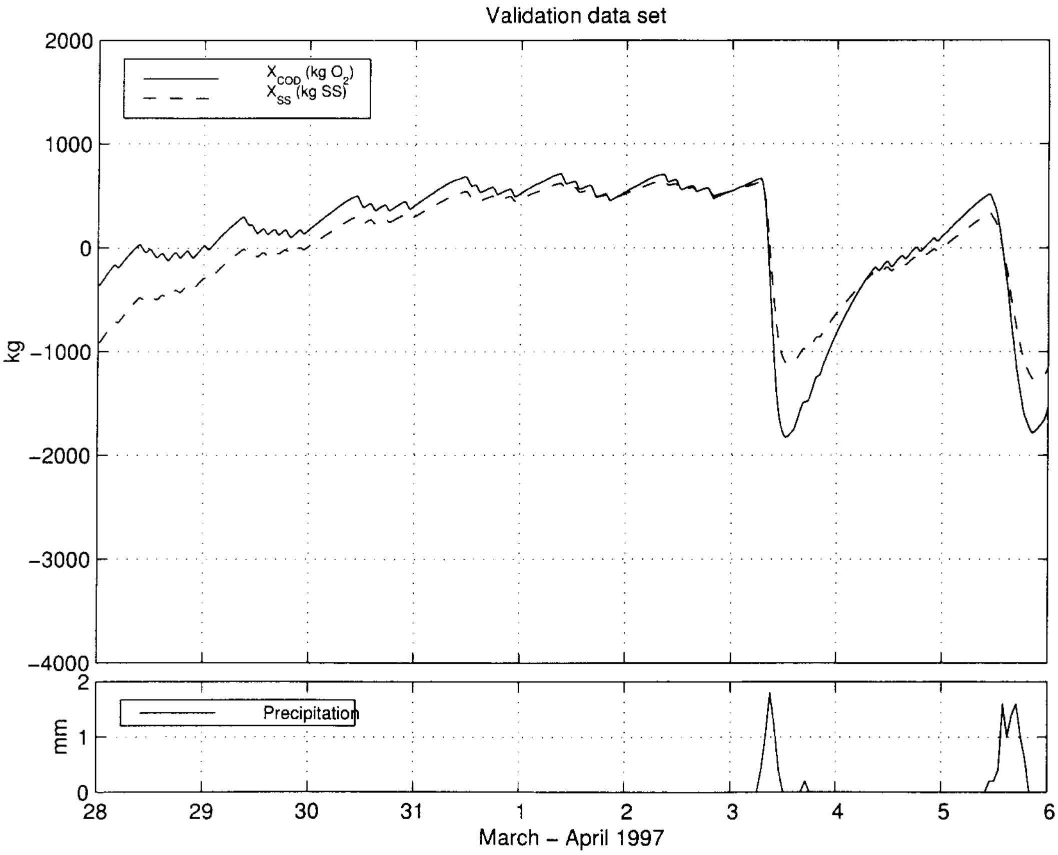

Figure 4. Estimated amounts of COD and SS deposits in the sewer system – validation data set

ACKNOWLEDGEMENTS

The authors wish to thank Skive WWTP for its cooperation during data collection and for making the laboratory analyses, and Henrik A. Thomsen and Kenneth Kisbye for their kind assistance in data collection. Thanks are also due to the Danish Academy of Technical Sciences for funding the project in which the present work has been carried out, under grant EF-623.

REFERENCES

Carstensen, J., Nielsen, M. and Harremoës, P. (1996). 'Predictive control of sewer systems by means of grey-box models'. *Water Science and Technology* **34**(3–4), 189–194.

Gelb, A. (1974). *Applied Optimal Estimation*. New York: MIT Press.

Kanaya, T., Fujita, I., Hayashi, H., Hiraoka, M. and Tsumura, K. (1985). 'Automatic measurement and analysis of process data in a combined wastewater treatment plant'. In *Instrumentation and Control of Water and Wastewater Treatment and Transport Systems*, *Advances in Water Pollution Control*, IAWPRC. ed. R. Drake. Oxford: Pergamon Press, 77–84

Kloeden, P. and Platen, E. (1995). *Numerical Solutions of Stochastic Differential Equations*, 2nd edn. Berlin: Springer.

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'. Technical Report no. 7/1991. Lyngby: Institute of Mathematical Statistics and Operations Research, Technical University of Denmark.

Matsché, N. and Stumwöhrer, K. (1996). 'UV absorption as control-parameter for biological treatment plants'. *Water Science and Technology* **33**(12), 211–218.

Melgaard, H. and Madsen, H. (1991). 'CTLSM version 2.6 – a program for parameter estimation in stochastic differential equations'. Technical Report no. 1/1993. Lyngby: Institute of Mathematical Statistics and Operations Research, Technical University of Denmark.

Melgaard, H. and Madsen, H. (1993). 'CTLSM continuous time linear stochastic modelling'. In *Workshop on Application of System Identification in Energy Savings in Buildings*, ed. F. Bloom. Ispra, Italy: Institute for Systems Engineering and Informatics, Joint Research Centre, 41–60.

Nowack, G. and Ueberbach, O. (1995). 'Die kontinuierliche SAK-Messung Aussagekraft, Statistik und Anwendungen'. *Korresspondenz Abwasser* **42**(11), 2020–2030.

Øksendal, B. (1995). *Stochastic Differential Equations*, 4th edn. Berlin: Springer.

Ruban, G., Marchandise, P. and Scrivener, O. (1993). 'Pollution measurement accuracy using real time sensors and wastewater samples analysis'. *Water Science and Technology* **28**(11–12), 67–78.

