Groundwater Management: Quantity and Quality (Proceedings of the Benidorm Symposium, October 1989). IAHS Publ. no. 188, 1989.

Prediction of the concentration distribution of groundwater pollutants

KENJI JINNO, AKIRA KAWAMURA, TOSHIHIKO UEDA & HIROSHI YOSHINAGA

Department of Civil Engineering, Faculty of Engineering, Kyushu University, 6-10-1 Hakozaki, Hagashi-ku, Fukuoka 812, Japan

Abstract A method based on the extended Kalman filter is developed to identify the parameters of a one-dimensional constant coefficient stochastic convective-dispersive equation and to predict the concentration distribution of groundwater pollutants. To study the various characteristics of this method, it is applied to a synthetically generated concentration distribution. The effects on the accuracy of predictions of the number of observation stations for the physical parameters, the sampling time interval, the number of terms in the Fourier series expansion and the local changes in physical parameters are investigated. This method identifies the physical parameters effectively and predicts the concentration distribution accurately

Prévision de la distribution des concentrations des polluants de l'eau souterraine

Résumé Une méthode fondée sur le filtre étendu de Kalman a été mise au point pour identifier le paramètre de l'équation de dispersion de convection stochastique du coefficient constant à une dimension et pour prévoir la distribution des concentrations des polluants de l'eau souterraine. Pour étudier les diverses caractéristiques de cette méthode, elle est appliquée à une distribution des concentrations qui est générée synthétiquement. L'influence sur l'exactitude des prévisions du nombre de stations d'observations des paramètres physiques, de l'intervalle temporel des échantillons, du nombre de termes dans l'expansion en séries de Fourier, et des changements locaux dans les paramètres physiques fait l'objet de recherches. Cette méthode identifie efficacement les paramètres physiques et prévoit correctement la distribution des concentrations.

NOTATION

A_m, B_m, E_m, F_m Fourier coefficient (g m^{-3})
B_0 mean of the series (g m^{-3})
C concentration (g m^{-3})

$\hat{C}$	prediction value of concentration (g m^{-3})
D	longitudinal dispersion coefficient (m^2 day^{-1})
f	$(2M + 4) \times 1$ system state vector function
H	known $N \times (2M + 4)$ observation matrix
i	observation time interval (day)
I	$(2M + 4) \times (2M + 4)$ identity matrix
J	$(2M + 4) \times (2M + 4)$ Jacobian matrix
J_1, J_2	accuracy standards of prediction
k	time instant
K	$(2M + 4) \times N$ Kalman gain matrix
l	basic wave length (m)
M	maximum number of terms in the Fourier series expansion
m	number of terms in the Fourier Series expansion
N	number of observation stations
N_1	number of evaluation time instants
N_2	number of evaluation stations
P	$(2M + 4) \times (2M + 4)$ system state estimate error covariance matrix
s	known $(2M + 4) \times 1$ time variant vector
t	time (day)
Δt	discrete time interval (day)
u	velocity (m day^{-1})
v	$(2M + 4) \times 1$ system noise vector
V	$(2M + 4) \times (2M + 4)$ covariance matrix of system noise
w	$(N \times 1)$ observation noise vector
W	$(N \times N)$ covariance matrix of observation noise
x	distance (m)
X	$(2M + 4) \times 1$ system state vector
$\hat{X}$	$(2M + 4) \times 1$ system state estimate vector
y	$(N \times 1)$ observation vector
γ	coefficient of adsorption (day^{-1})
ϵ	zero-mean Gaussian white noise (g m^{-3})
ν	$(N \times 1)$ innovation vector
Φ	known $(2M + 4) \times (2M + 4)$ state transition matrix

INTRODUCTION

When a convective-dispersive phenomenon of groundwater pollutants is analysed, its mechanism should be regarded as stochastic (Kawamura *et al.*, 1989; Jinno *et al.*, 1986a). Even though numerical simulations by a finite difference method (FDM) or a finite element method (FEM) are commonly employed to predict the concentration distribution of pollutants, these methods require that parameters of the convective-dispersive differential equation and deterministic initial and boundary conditions should be known beforehand. These assumptions are ambiguous for natural groundwater flow systems. The phenomenon should be considered as stochastic because of many uncertainties in the interaction between pollutant and flow (Jinno *et al.*, 1986b).

A method based on the extended Kalman filter is developed in which the concentration at each observation station can be effectively used for both prediction of the transport of pollutants and identification of parameters of the one-dimensional, constant coefficient stochastic convective-dispersive differential equation. To study the various characteristics of this method, it is applied to synthetically generated concentration distributions. The effects on the accuracy of identification of physical parameters and prediction of concentration distribution of the number of observation stations, the sampling time interval, the number of terms in the Fourier series expansion and the spatial changes in physical parameters are investigated.

KALMAN FILTER FORMULATION

The Kalman filter formulation considers the following system and observation equations (Athans *et al.*, 1968; Kawamura *et al.*, 1984a, 1984b, 1989):

$$X(k+1) = \Phi(k)\, X(k) + s(k) + v(k) \tag{1}$$

$$y(k+i) = H(k+i)\, X(k+i) + w(k+i) \tag{2}$$

The estimate of the state vector at time step $k+i$, calculated using the observation obtained at time step k, is denoted as $\hat{X}(k+i|k)$ and at time step $k+i$ as $\hat{X}(k+i|k+i)$. If $\hat{X}(k|k)$ is known after the observation is obtained at time step k, $\hat{X}(k+i|k)$ and $\hat{X}(k+i|k+i)$ are computed at time step $k+i$ as follows:

$$\hat{X}(k+i|k) = \Phi(k+i-1)\, \hat{X}(k+i-1|k) + s(k+i-1) \tag{3}$$

$$\hat{X}(k+i|k+i) = \hat{X}(k+i|k) + K(k+i)\, v(k+i) \tag{4}$$

where

$$K(k+i) = P(k+i|k)\, H^T(k+i)\, [H(k+i)\, P(k+i|k)\, H^T(k+i) + W(k+i)]^{-1} \tag{5}$$

$$v(k+i) = y(k+i) - \hat{y}\,(k+i|k) \tag{6}$$

$$\hat{y}(k+i) = H(k+i)\, \hat{X}(k+i|k) \tag{7}$$

Similarly, the state estimate error covariance matrix at time step $k+i$, calculated using the observation obtained at time step k, is denoted as $P(k+i|k)$ and at time step $k+i$ as $P(k+i|k+i)$. They are calculated as follows:

$$P(k+i|k) = \Phi(k+i-1)\, P(k+i-1|k)\, \Phi^T(k+i-1) + V(k+i-1) \tag{8}$$

$$P(k+i|k+i) = [I - K(k+i)\, H(k+i)]\, P(k+i|k) \tag{9}$$

PREDICTION MODELLING FOR CONSTANT PARAMETERS

In this paper, the following one-dimensional, constant coefficient stochastic convective-dispersive equation is considered:

$$\frac{\partial C(x,t)}{\partial t} + u\frac{\partial C(x,t)}{\partial x} = D\frac{\partial^2 C(x,t)}{\partial x^2} - \gamma C(x,t) + \epsilon(x,t) \tag{10}$$

The physical parameters u, D and γ are assumed to be constant in time and space. The last term in equation (10) is added to take into account the uncertainties which are inherent in modelling the phenomenon.

The stochastic partial differential equation is transformed into an ordinary differential equation using a Fourier series expansion. Specifically, $C(x,t)$ and $\epsilon(x,t)$ are expanded as:

$$C(x,t) = B_0(t) + \sum_{m=1}^{M} [A_m(t)\ \sin(2\pi mx/l) + B_m(t)\ \cos(2\pi mx/l)] \tag{11}$$

$$\epsilon(x,t) = \sum_{m=1}^{M} [E_m(t)\ \sin(2\pi mx/l) + F_m(t)\ \cos(2\pi mx/l)] \tag{12}$$

When a filtering approach is used for spatial systems, one is faced with discretizing a partial differential equation. FDM and FEM are two discretization methods commonly used. However, they require imposed boundary conditions even if these boundary conditions are difficult to specify and randomly disturbed. Moreover, the accuracy of numerical calculation will be lost if an improper mesh size is used in either FDM or FEM simulation. On the other hand, the Fourier series expansion assumes a periodic boundary condition which can be altered easily to accommodate any situations (Jinno *et al.*, 1986b).

Substituting equations (11) and (12) into equation (10), a set of homogeneous ordinary differential equations for the Fourier coefficients of the frequency component m is obtained:

$$\begin{bmatrix} dA_m/dt \\ dB_m/dt \end{bmatrix} = \begin{bmatrix} -R_m & Q_m \\ -Q_m & -R_m \end{bmatrix} \begin{bmatrix} A_m(t) \\ B_m(t) \end{bmatrix} + \begin{bmatrix} E_m(t) \\ F_m(t) \end{bmatrix} \tag{13}$$

where:

$$R_m = D(2\pi m/l)^2 + \gamma \tag{14}$$

$$Q_m = u(2\pi m/l) \tag{15}$$

The system state vector X is modelled as follows:

$$X(t) = [u\ D\ \gamma\ B_0(t)\ A_1(t)\ B_1(t)\ \ldots\ A_M(t)\ B_M(t)]^T \tag{16}$$

Here, the physical parameters u, D and γ are assumed constant in time and space such that:

$$du/dt = 0, \quad dD/dt = 0, \quad d\gamma/dt = 0 \tag{17}$$

Combining equations (13) and (17), the system equation is obtained as follows:

$$dX/dt = f(X(t)) + v(t) \tag{18}$$

$$= \begin{bmatrix} 0 \\ 0 \\ 0 \\ -B_0(t)\gamma \\ -R_1A_1(t) + Q_1B_1(t) \\ -Q_1A_1(t) - R_1B_1(t) \\ \vdots \\ -R_MA_M(t) + Q_MB_M(t) \\ -Q_MA_M(t) - R_MB_M(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ E_1(t) \\ F_1(t) \\ \vdots \\ E_M(t) \\ F_M(t) \end{bmatrix} \tag{19}$$

Equation (19) is a set of nonlinear functions of u, D, γ, B_0, A_m and B_m. Here, the Taylor series expansion of $f(X(t))$ about $X(t) = X^*(t)$ is carried out to first order terms, and $f(t)$ is linearized as:

$$f(X(t)) = f(X^*(t)) + J(X^*(t)) \; [X(t) - X^*(t)] \tag{20}$$

where

$$\{J(X^*(t))\}_{ij} = [\partial f_i/\partial X_j(t)]_{X_j(t)=X_j^*(t)} \tag{21}$$

Substituting equation (20) into equation (18) and arranging, equation (22) is obtained as:

$$dX/dt = J(X^*(t))\, X(t) + f(X^*(t)) - J(X^*(t)) \; X^*(t) + v(t) \tag{22}$$

The left-hand side of equation (22) is discretized as follows:

$$dX/dt = [X(t + \Delta t) - X(t)]/\Delta t \tag{23}$$

Substituting equation (23) into equation (22) and denoting t and $t + \Delta t$ by k and $k + 1$, equation (24) is obtained:

$$X(k + 1) = J(X^*(k))\, \Delta t + I]\, X(t) + [f(X^*(t)) - J(X^*(k))\, X^*(k)]\, \Delta t + v(k)\, \Delta t \tag{24}$$

Comparing with equation (1):

$$\Phi(k) = J(X^*(k))\,\Delta t + I \tag{25}$$

$$s(t) = [f(X^*(k)) - J(X^*(k))\,X^*(k)]\,\Delta t \tag{26}$$

$$v(t) = v(k)\,\Delta t \tag{27}$$

$X^*(k)$ is replaced by $\hat{X}(k|k-j)$, the estimate of the state vector at time step k based on the observation at time step $k-j$.

The observation vector y is the concentration distribution measured at the randomly, spatially arranged observation stations at sampling time interval i. The concentration distribution given by equation (11) is written as:

$$C(x,k) = [0\ 0\ 0\ 1 \sin(2\pi x/l) \cos(2\pi x/l) \ \ldots\ \sin(2\pi Mx/l) \cos(2\pi Mx/l)]$$
$$\cdot\,[u\ D\ \gamma\ B_0(k) A_1(k) B_1(k) \ldots A_M(k) B_M(k)]^T + w(x,k) \tag{28}$$

representing the observation matrix H in equation (2). The observation matrix H is:

$$H = \begin{bmatrix} 0\ 0\ 0\ 1\ \ldots & \sin(2\pi Mx_1/l) & \cos(2\pi Mx_1/l) \\ & \vdots & \vdots \\ 0\ 0\ 0\ 1\ \ldots & \sin(2\pi Mx_N/l) & \cos(2\pi Mx_N/l) \end{bmatrix} \tag{29}$$

The concentration distribution is predicted using the extended Kalman filter on the basis of information from the observation stations. The unknown physical parameters and Fourier coefficients are identified in the prediction process.

EXAMPLE 1

In order to study the various characteristics of the proposed method, it was applied to a synthetically generated concentration distribution. Figure 1 shows the generated concentration distribution (solid lines) of a one-dimensional constant coefficient stochastic convective-dispersive equation (10).

The following values were used for the synthetic generation: u = 1.0 m day^{-1}, D = 0.5 m^2 day^{-1}, γ = 0.01 day^{-1}, Δt = 0.1 day^{-1}, l = 100 m, M = 20 and the Gaussian random noises E_m and F_m were assumed to be zero-mean with standard deviation equal to 0.0005 g m^{-3}. Equation (13) was solved numerically using the Runge-Kutta-Gill method (Kawamura *et al.*, 1989). Here, the initial concentration distribution for the simulation study was the analytical solution of the deterministic equation (10) (excluding the term $\epsilon(x,t)$) at t = 10 days, given the following initial

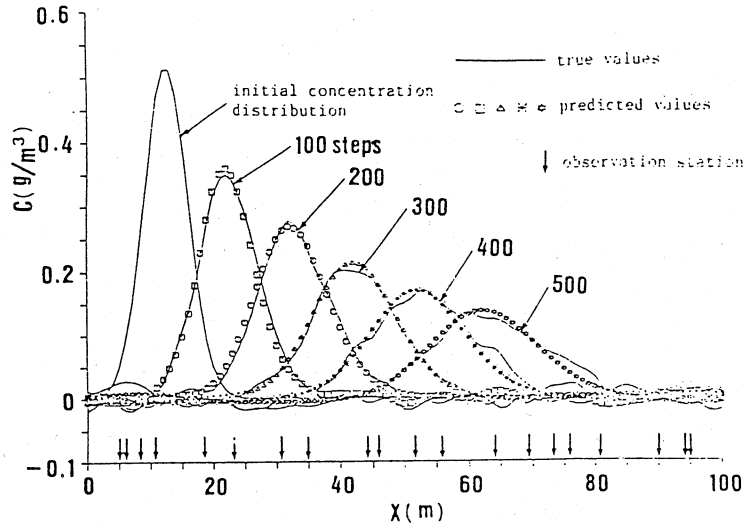

Fig. 1 Prediction of the concentration distribution (Example 1).

condition:

$$C(x,0) = 1 \quad (0 \leqslant x \leqslant 5)$$

$$C(x,0) = 0 \quad (-\infty < x < 0,\ 5 < x < \infty)$$

When the generated concentration distribution was assumed observable, an observation noise with zero-mean and standard deviation equal to 0.01 g m^{-3} was added as shown in equation (28). Likewise, when the physical parameters u, D and γ were observable at a certain station, an observation noise with zero-mean and standard deviation equal to 10% of the true value was added.

An extended Kalman filter was used to predict the concentration distribution and to identify the parameters. The following initial conditions required for the recursive applications of the extended Kalman filter were used in this example. The initial values of the physical parameters and the Fourier coefficients were assumed to be 50% of the true values. The diagonal elements of P, for u, D were taken as 0.01, for γ and Fourier coefficients as 0.001 and off-diagonal as 0.0001. The diagonal elements of V were all taken as 0.0005^2 and off-diagonal as zero. Downward-arrows in Fig. 1 indicate the 20 observation stations which were located randomly in space. Observed values of the concentration distribution were obtained every 10 time steps of one day. Figure 1 show the predicted concentration distribution.

Figure 2 shows some of the identified parameters. The accuracy of the prediction of concentration distribution was quantitatively expressed as follows:

$$J_1 = \sum_{k=1}^{N_1} J_2(k)/N_1 \tag{30}$$

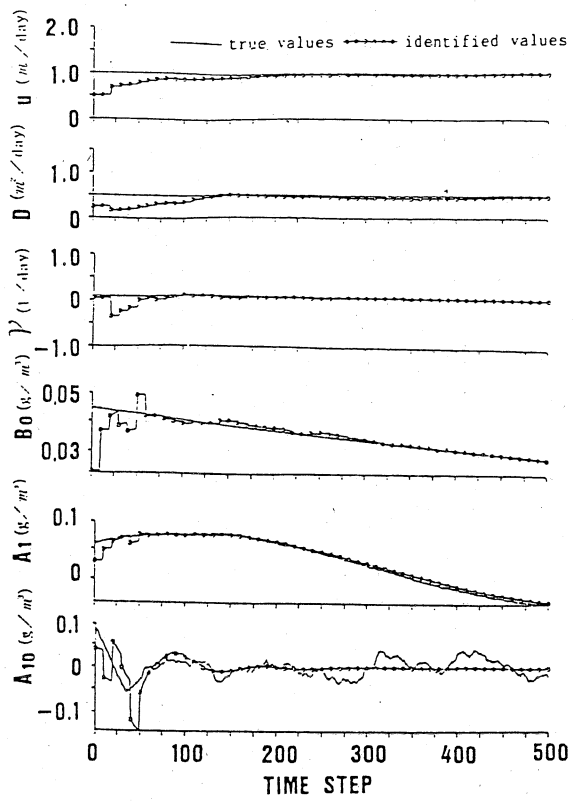

Fig. 2 Identification of the physical parameters and Fourier coefficients (Example 1).

$$J_2(k) = \left\{\sum_x [\hat{C}(x,k|k-i) - C(x,k)]^2\right\}^{1/2}/N_2 \qquad (30)$$

where $\hat{C}(x,k|k-i)$ was evaluated for every one metre at time step k (N_2 = 101). Figure 3 shows the variation of J_1 as a function of N when the physical parameters were observable. Figure 4 shows the effect of the different sampling time intervals on the accuracy of prediction.

EXAMPLE 2

The proposed prediction method was applied to the case where the physical parameters u, D and γ changed at x = 30 m as follows:

$$u_1 = 1.0 \text{ m day}^{-1}, \quad D_1 = 0.5 \text{ m}^2 \text{ day}^{-1}, \quad \gamma_1 = 0.01 \text{ day}^{-1} \quad (0 \leq x < 30)$$

$$u_2 = 1.5 \text{ m day}^{-1}, \quad D_2 = 0.1 \text{ m}^2 \text{ day}^{-1}, \quad \gamma_2 = 0.0 \text{ day}^{-1} \quad (30 \leq x < 100)$$

Since the parameters change spatially in this example, the method of characteristics was used in generating the synthetic concentration distribution

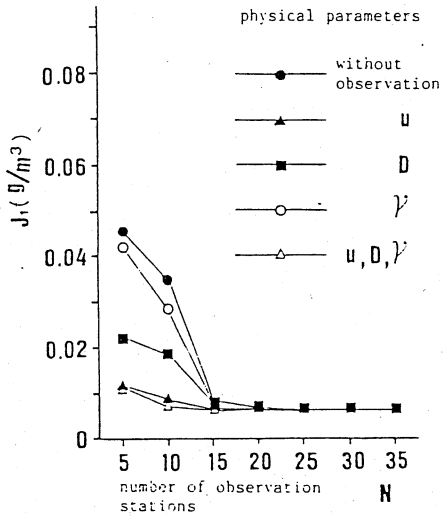

Fig. 3 Variation of J_1 as a function of N with observation of the physical parameters (Example 1).

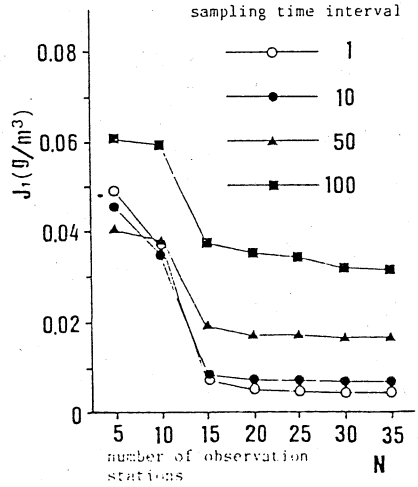

Fig. 4 Variation of J_1 as a function of N for various sampling time intervals (Example 1).

(Jinno & Ueda, 1978). Figure 5 shows the synthetic concentration distribution (solid lines) generated under the following conditions: Δt = 0.1 day and the Gaussian random noise ϵ (see equation (1)) was assumed to be zero-mean with standard deviation equal to 0.02. The extended Kalman filter was used with the same initial conditions as in example 1, except for the matrix V. The diagonal elements of V for u, D and γ were set equal to 0.00001 times the square of the amount of change.

Figure 5 shows the results. Figure 6 shows the identified parameters u,

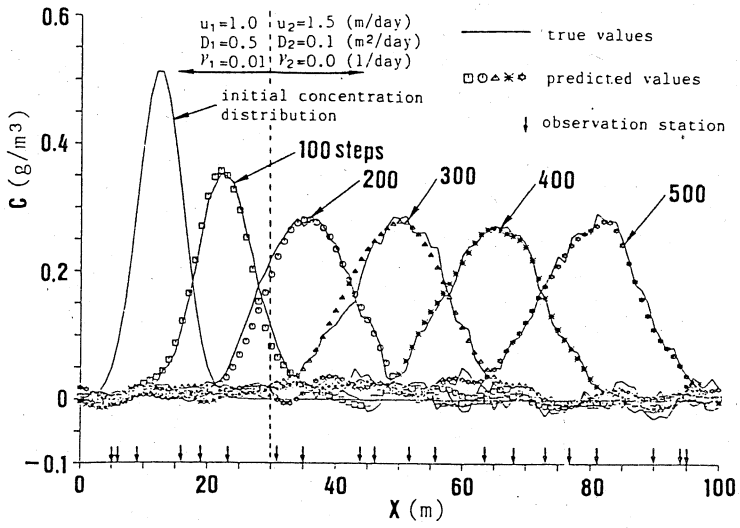

Fig. 5 Prediction of the concentration distribution (Example 2).

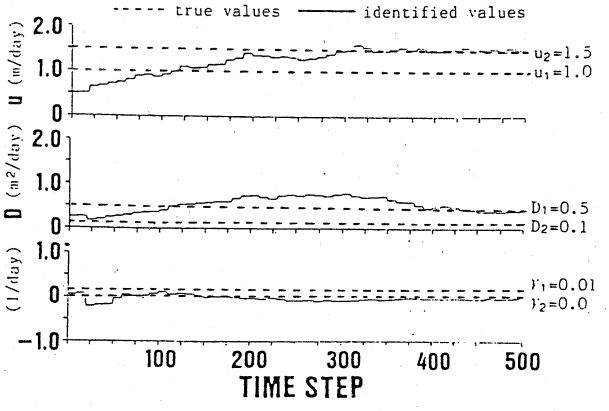

Fig. 6 Identification of the physical parameters (Example 2).

D and γ, and Fig. 7 the variation of J_1 as a function of N for different sampling time intervals.

DISCUSSION

Figure 1 shows the synthetically generated concentration distribution disturbed by noise ϵ. This disturbance became dominant as time passed. The 10-step ahead predicted values were shown to be accurate, except the contributions by the high frequency components.

Figure 2 shows the changes in the parameters and some Fourier coefficients. They converged to the true values after 150 time steps. The power spectrum of the identified Fourier coefficient for the wave

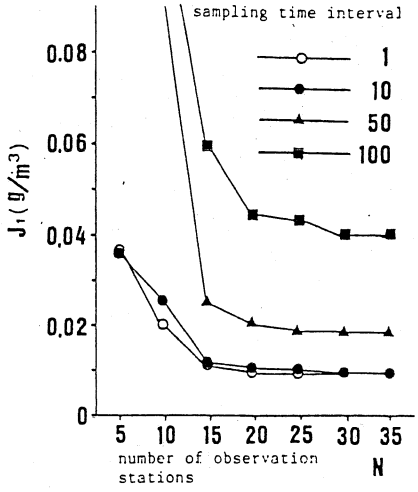

Fig. 7 Variation of J_1 as a function of N for various sampling time intervals (Example 2).

number m = 10 became less than that of the system noise ϵ (Kawamura *et al.*, 1989) as time passed. This means that the contribution of the high wave numbers (more than m = 10) to the concentration distribution is small.

As shown in Fig. 3, the accuracy of the prediction became high and J_1 converged to 0.008 independent of the observation of the physical parameters when N was more than 15. On the other hand, the accuracy depended on the observation of the physical parameters when N was less than 15. If all the parameters were observed, the accuracy was still reasonable even if the number of observation stations was less than 15. The observation of u gave the highest accuracy for the prediction when one of the three physical parameters was to be measured. In order to predict the concentration distribution accurately with fewer observation stations, it would be necessary to get a reliable observation for u.

Figure 4 shows that, when N was more than 15, J_1 became independent of the number of observation stations for any sampling time interval. The higher accuracy was obtained for the shorter sampling time interval. This means that more information was obtained when the sampling time interval was shorter. Nor N less than 15, the prediction was poor for any sampling time interval. J_1 became smaller than 0.01 (g m^{-3}) for any number of observation stations if sampling was made at a frequency less than 10 time steps, being independent of the number of terms in the Fourier series expansion.

The accuracy of the prediction of concentration distribution with 10 terms in the Fourier series expansion did not differ so much from that with 15 terms.

Figure 5 shows the prediction of the concentration distribution when the physical parameters u, D and γ changed at x = 30 (m). Accurate predictions were still obtained even though the parameters changed locally. This means that the present method would be robust for such a case.

As shown in Fig. 6, the parameter u was identified very accurately. When the peak of the concentration distribution still remained before x = 30 m where

the physical parameters change (i.e. around time step 100), u seemed to converge to u_1. On the other hand, u converged to u_2 after the peak passed by x = 30 m (i.e. around time step 300). The parameter γ converged to γ_2 after the peak of the concentration distribution passed by x = 30 m. It seems, however, that γ could not converge to γ_1 before the peak passed by x = 30 m. The identification of D seemed to be difficult in this example because D did not converge to D_2 within 500 time steps. However, an accurate concentration distribution was obtained compensating for the inaccurate D by the Fourier coefficients.

Figure 7 shows that J_1 as a function of N with different sampling time intervals was similar to that in Fig. 4 if N was more than 15. The present method still predicted the concentration distribution accurately even though the physical parameters changed spatially. The accuracy of prediction and identification for different values of V was also studied. The results showed that V was more effective to the accuracy of the identification of physical parameters than that of the prediction of concentration distribution.

CONCLUSIONS

In this paper, a method that uses the Fourier series expansion and the extended Kalman filter is proposed for the identification of the parameters of a one-dimensional constant coefficient stochastic convective-dispersive equation and the on-line prediction of the concentration distribution of groundwater pollutants. This method was compared with two synthetically generated concentration distributions: one with constant physical parameters and the other with spatial change in the physical parameters. The results obtained herein showed that the proposed method would be able to predict the concentration distribution and identify the parameters accurately.

REFERENCES

Athans, M., Wishner, R. P. & Bertolini, A. (1968) Suboptimal state estimation for continuous-time nonlinear system from discrete noisy measurement. *IEEE Trans. on Automatic Control* AG-13 (5), 504-514.

Jinno, K. & Ueda, T. (1978) On the numerical solutions of convective-dispersion equation by shifting particles (in Japanese). *Proc. Japan Soc. Civ. Engrs* no. 271, 45-53.

Jinno, K., Ueda, T. & Yasuda, H. (1986a) Parameter identification of two-dimensional convective-dispersion equation and prediction of concentration distribution based on on-line observations (in Japanese). *Proc. 30th Japan Conf. Hydraulics* 30, 313-318.

Jinno, K. Ueda, T. & Yasuda, H. (1986b) Analytical approach to monitoring pollution by chlorinated solvents. In: *monitoring to detect Changes in Water Quality Series* (Proc. Budapest Symp., July 1986), 185-196. IAHS Publ. no. 157.

Kawamura, A., Jinno, K. & Ueda, T. (1984a) On the characteristics of the adaptive Kalman filter for detecting abnormality in a periodic time sequence (in Japanese). *Proc. 28th Japan Conf. Hydraulics* 28, 383-390.

Kawamura, A., Jinno, K. & Ueda, T. (1984b) Detection of abnormality by the adaptive Kalman filter (in Japanese). *Proc. Japan Soc. Civ. Engrs* no. 345/II-1, 111-121.

Kawamura, A., Jinno, K., Ueda, T. & Yoshinaga, H. (1989) On the on-line prediction of the concentration distribution of one-dimensional constant coefficient stochastic convective-dispersion equation based on Kalman filter (in Japanese). *Tech. Reports of Kyushu Univ.* vol. 62, no. 1, January 1989.