Journal of Hydroscience and Hydraulic Engineering
Vol. 3, No. 2, November, 1985, pp.49-59.

ERROR STRUCTURE AND PROBABILISTIC EVALUATION OF SIMULATION ON GROUND WATER BEHAVIORS

By

Fusetsu Takagi

Professor, Department of Civil Engineering
Nagoya University, Nagoya 464, Japan

and

Morihiro Harada

Research Associate, Department of Civil Engineering
Nagoya University, Nagoya 464, Japan

SYNOPSIS

The present paper intends to propose a evaluation method of the ground water simulation by combining the kinematical and probabilistic approaches. For this purpose, the variational formulation has been made on the ground water behaviors, which gives us a useful tool to express the relationships among errors of parameter estimation, boundary and initial conditions and the accuracy of the simulated result in the whole region concerned. This relationship is combined with the probabilistic concept and leads to the evaluation-index of the ground water simulation.

GENERAL REMARKS

The simulation technique has been widely used to investigate and estimate the confined ground water behaviors in various regions. Based on the flow mechanism and geophysical and geological information, usually the model structure is fixed at first and then the model parameters are estimated through the identification process, comparing the observed data with the calculated data. Thereafter, the model is applied to various purposes concerning with the ground water behaviors in real regions for the given boundary and initial conditions, pumping intensity and so forth. Even though the parameters are carefully identified, therefore, the accuracy of the simulated result depends greatly on the accuracy of the conditions and input information. However, in the actual state, it is very difficult to estimate the real states of the boundary and initial conditions etc. themselves.

Numerous research works have been promoted on the simulation technique, and most of the efforts have been devoted to the calculation procedure or the identification problems. However, it seems that we have lack of attention to the objective evaluation of the simulated results themselves. The discretion itself by the professional engineers or researchers in the field is, of course, one of the basic evaluation of the results obtained through the simulation. The problem discussed in the present paper is how to evaluate once more the simulated results, which were made by the professional engineers based on the potential distributions, flow directions, flow regimes, geological information and so forth. The authors intend to propose the evaluation method of the ground water simulation by combining the kinematical and probabilistic approaches. For this purpose, the variational formulation is made on the ground water behaviors, which gives us a useful tool to express the relationships among errors in the estimation of parameters, boundary and initial conditions and also the errors or accuracy of the simulated result over the whole region concerned. This relationship is combined with the probabilistic concept and leads to the probabilistic evaluation of the final result of simulation.

VARIATIONAL FORMULATION OF GROUND WATER BEHAVIORS

Two dimensional fundamental equation of the ground water flow through the confined aquifer takes the following form, integrated the three-dimensional equation over the aquifer with respect to the vertical direction,

$$S\frac{\partial h}{\partial t} = \Sigma\left\{\frac{\partial}{\partial x_i}\ T\frac{\partial h}{\partial x_i}\right\} + r \tag{1}$$

In this equation the symbols denote, x_i: the horizontal coordinates (i=1,2), t: time, h: Piezometric head, r: vertical inflow flux (intensity) to the aquifer including the recharge, leakage and pumping intensity per unit area, S: storage coefficient and T=kb: transmissibility, i.e. the product of the permeability coefficient k and the thickness of the aquifer b.

We write here the piezometric head h as the summation of the real value h* and the small arbitrary variation δh around the real value h*. Both of them, i.e. h* and δh are the functions of space coordinates as well as time; that is,

$$h(x_i,\ t) = h^*(x_i,\ t) + \delta h(x_i,\ t) \tag{2}$$

$$|h^*(x_i,\ t)| >> |\delta h(x_i,\ t)| \tag{3}$$

Multiplying the equation (1) by δh, we obtain the following equation with the help of Eqs. (2) and (3),

$$S\delta h\frac{\partial \delta h}{\partial t} = -S\delta h\frac{\partial h^*}{\partial t} - \frac{1}{2}T\delta\Sigma\left(\frac{\partial h}{\partial x_i}\right)^2 + r\delta h + \Sigma\frac{\partial}{\partial x_i}\left\{T\frac{\partial h}{\partial x_i}\delta h\right\} \tag{4}$$

The integration of this equation (4) over the whole region concerned G and also the time results in

$$\frac{1}{2}\int_G (\delta h)^2 dx_i = -\delta\int_G \frac{\partial h^*}{\partial t}h + \frac{1}{2}\frac{T}{S}\Sigma\left(\frac{\partial h}{\partial x_i}\right)^2 - \frac{1}{S}rh\ dx_i dt$$

$$+ \oint_C \frac{T}{S}\left\{\frac{\partial h}{\partial x_1}\frac{dx_2}{ds} - \frac{\partial h}{\partial x_2}\frac{dx_1}{ds}\right\}\delta h\ ds dt$$

$$+ \frac{1}{2}\int_I (\delta h)^2\ dx_i \qquad \geq \qquad 0 \tag{5}$$

where C is the boundary around the region G, s is the counterclockwise distance along C, $\int dx_i$ is the areal integration over the region G, i.e. $\iint dx_1 dx_2$, and the last term is the areal integral at the initial time.

For the real behaviors of ground water, since

$$h(x_i,\ t) \equiv h^*(x_i,\ t) \tag{6}$$

all over the region G, the equality in Eq.(5) should be satisfied. Moreover, since δh becomes zero at the boundary and initial states, provided the boundary and initial conditions are correctly given, the line integral and the areal integral at the initial time $\int_I dx_i$ vanish. In this case, then, only the first term remains, and we may write the fundamental equation of ground water as,

$$\delta\int L_g\ dx_i dt = 0 \tag{7}$$

$$L_g = \frac{\partial h^*}{\partial t}h + \frac{1}{2}\frac{T}{S}\Sigma\left(\frac{\partial h}{\partial x_i}\right)^2 - \frac{1}{S}rh \tag{8}$$

In this equation the variation should be taken with respect only to the quantity h, keeping the real quantity h* fixed, according to the assumption of local equilibrium. Thereafter, the principle (7) must be used with the subsidiary condition (6). The assumption of local equilibrium is what we need to apply the variational principle held only for the conservative system to the dissipative one.

The variational principle (7) as introduced above is the equation which governs

the ground water flow in the region G as a whole. Therefore, to solve the differential equation (1) for the given conditions is equivalent to the problem to find the solution h* which satisfies the variational principle (7) for the conditions.

Although the detailed discussions are omitted here on the variational principle, since the more comprehensive presentations on its physical significances and its various applications have been already given by one of the authors (1)(2)(3), it turned out that local potential L_g^* in which the quantity h in L_g is substituted by the real water depth h*, is in the following relationship between the loss of kinematical energy ΔE_g per unit area and unit time, and the kinematical energy carried into a unit area of ground water region by the flow, that is,

$$L_g^* = -\frac{1}{2}\Delta E_g + (\text{energy flux}) \tag{9}$$

In other words, the variational principle may be considered as to be closely related with the energy of flow in the aquifer. Moreover, it should be noted here, that the left hand side of Eq.(5) corresponds to the (half times of) summation of square errors over the whole region G concerned at the specified time, if we consider the quantity δh as the error between the quantity h and the real behaviors h*.

VARIOUS ERRORS INVOLVED IN THE SIMULATION

The errors which come into the simulation may be classified as Table 1. In the table, the errors expressed as "Model structure" are concerned with such the basic problems as to whether the flow mechanism is correctly expressed in the mathematical form, i.e. it is governed by Darcy's law or not and also whether the flow region of ground water is appropriately estimated as the integral domain of the basic equation.

Table 1

Model structures
Parameters
Conditions
Numerical procedures

The next category "Parameters" implies the errors due to the over- and/or under-estimations of parameters T and S. As the third category, if the initial and boundary conditions including the recharge intensity and pumping intensity are not correctly given, the simulation results will have the corresponding errors. At last, some errors may arise in the "Numerical procedures". Although we have to consider all these problems in the general meaning, for the brief treatment in the present paper we will treat only the 2nd the 3rd categories in the Table 1 and we assume that the model structures are correctly formulated, besides the numerical procedures are not considered.

Errors due to the estimation of the parameters

In this article at first we deal with such the case that the simulation is carried out with the approximated parameters S_o and T_o and all the other factors are correctly given. The simulation result with the parameters including errors, that is, the approximated value of the potential distribution is written by the notation $h_o(x_i,t)$ and its error by the symbol with prime, namely

$$h(x_i, t) = h_o(x_i, t) + h'(x_i, t) \tag{10}$$

Moreover, for the other factors we write here the approximated values and the errors by the subscripts o and prime, respectively,

$$(S/T) = (S/T)_o + (S/T)', \qquad (1/S) = (1/S)_o = (1/S)' \tag{11}$$

In this simulation, the approximated value of the potential distribution h_o corresponds to the solution h_o^* which satisfies the following variational principle

$$\delta \int L_{go}\, dx_i dt = 0 \tag{12}$$

with the local potential

$$L_{go} = \frac{\partial h_o^*}{\partial t} h_o + \frac{1}{2}\left(\frac{T}{S}\right)_o \Sigma \left(\frac{\partial h_o}{\partial x_i}\right)^2 - \left(\frac{1}{S}\right)_o r h_o \tag{13}$$

instead of the original equation (7) and (8). As easily understood, the variational principle (12) should be treated with the subsidiary condition.

$$h_o(x_i,\ t) \equiv h_o^*(x_i,\ t) \tag{14}$$

Introducing the local potential L_{go}, we may rewrite Eq.(7) in the following form

$$\frac{1}{2} \int (\delta h)^2\, dx_i = - \delta \int L_{go}\, dx_i dt - \delta \int (L_g - L_{go})\, dx_i dt \geq 0 \tag{15}$$

In this equation, since the approximated solution h_o satisfies Eq.(12), the error distributions h' of the approximated solution should satisfy

$$\delta \int L_g'\, dx_i dt = 0 \tag{16}$$

$$\begin{aligned} L_g' &= L_g - L_{go} \\ &= \frac{\partial h'^*}{\partial t} h' + \frac{1}{2} \Sigma \frac{T}{S}\left(\frac{\partial h'}{\partial x_i}\right)^2 - \frac{1}{S} r h' + \Sigma \left(\frac{T}{S}\right)' \frac{\partial h_o^*}{\partial x_i} \frac{\partial h'}{\partial x_i} + \begin{matrix}\text{terms which}\\ \text{vanish in the}\\ \text{variational}\\ \text{procedure}\end{matrix} \end{aligned} \tag{17}$$

In this step L_{go} is a definite known function of h_o^*, because the approximated solution h_o^* is already given and the treatment of the variational calculus (16) is same for Eqs. (8) and (12). It goes without saying that Euler-Lagrangian condition of the variational principle (16) with help of the subsidiary condition $h'(x_i,t) \equiv h'^*(x_i,t)$ gives the following differential equation with respect to the error h'.

$$\frac{\partial h'^*}{\partial t} - \Sigma \frac{\partial}{\partial x_i}\left\{\left(\frac{T}{S}\right)\frac{\partial h'^*}{\partial x_i}\right\} - \Sigma \frac{\partial}{\partial x_i}\left\{\left(\frac{T}{S}\right)' \frac{\partial h_o^*}{\partial x_i}\right\} - \left(\frac{1}{S}\right)' r = 0 \tag{18}$$

As discussed above, in the case only the paremeters involve the errors, if we obtain the estimated solution h_o^*, the summation of square errors all over the region, that is, the error of the simulation is conclusively written in the form

$$\frac{1}{2} \int (\delta h)^2\, dx_i = - \delta \int L_g'\, dx_i dt = E_1 \geq 0 \tag{19}$$

Errors due to the estimation of the initial condition

In the case that only the initial condition involves a certain error, Eq.(5) is written as

$$\frac{1}{2} \int (\delta h)^2\, dx_i = - \delta \int L_g\, dx_i dt + \frac{1}{2} \int_I (\delta h)^2\, dx_i \geq 0 \tag{20}$$

Introducing the quantities h_o, h', L_{go} and L_g' in the same way as in the last article, we obtain

$$L_{go} = \frac{\partial h_o^*}{\partial t} h_o + \frac{1}{2} \Sigma \left(\frac{T}{S}\right)\left(\frac{\partial h_o}{\partial x_i}\right)^2 - \left(\frac{1}{S}\right) r h_o \tag{21}$$

$$L_g' = \frac{\partial h'^*}{\partial t} h' + \frac{1}{2} \Sigma \left(\frac{T}{S}\right)\left(\frac{\partial h'}{\partial x_i}\right)^2 + \frac{\partial h'^*}{\partial t} h_o^* + \Sigma \frac{\partial}{\partial x_i}\left\{\left(\frac{T}{S}\right)\frac{\partial h_o^*}{\partial x_i} h'\right\} \tag{22}$$

and Eq.(20) may be rewritten as follows.

$$\frac{1}{2}\int (\delta h)^2 \, dx_i = -\delta \int L_g' \, dx_i dt + \frac{1}{2}\int_I (\delta h)^2 \, dx_i = E_2 \geq 0 \tag{23}$$

In this way, it may be concluded that the errors of the simulation over the whole region G appears in the form E_2 of the above equation (23), if only the initial condition is estimated with some errors.

Errors due to the estimation of the boundary conditions

We write the approximated solution as h_o, which is given by the simulation in such the situation that only the boundary conditions have a certain error and all the other factors are correctly estimated. In this case we may deduce the following equation

$$\frac{1}{2}\int (\delta h)^2 \, dx_i = -\delta \int L_g' \, dx_i dt + \oint \frac{T}{S}\left\{\frac{\partial h^*}{\partial x_1}\frac{dx_2}{ds} - \frac{\partial h^*}{\partial x_2}\frac{dx_1}{ds}\right\}\delta h \, ds dt$$

$$= E_3 \geqq 0 \tag{24}$$

in which the local potential $L_g' = L_g - L_{go}$ and L_{go} are of the same mathematical forms as Eqs.(22) and (21), respectively. This equation (24) gives the error structure of the simulation for the approximated boundary conditions.

Error of simulation

We have considered how the errors in various factors affect mathematically on the error of simulation all over the region concerned. Although the error distributions over the region G may be essentially obtained by the equation

$$\delta \int L_g' \, dx_i dt = 0 \tag{25}$$

to solve the equation we need the correct parameters involved in L_g' and the correct boundary and initial conditions. If we know these parameters and the conditions, we do not have to obtain the approximate solution, and actually it is impossible to express the real error distributions. To estimate the error of the simulation we should then try the another approach as discussed later.

Anyway, the discussions made in this chapter implies that in the case that errors of various factors are simultaneously involved in the simulation, the synthetical error of simulated results as a whole may be expressed by the following form and it is closely related with the energy of ground water flow.

$$\frac{1}{2}\int (\delta h)^2 \, dx_i = E \geq 0 \tag{26}$$

IDENTIFICATION AND ERROR OF SIMULATION

In the usual simulation procedure, the model parameters are identified so that the calculated values of the potential become equal to the observed potentials at several well positions as well as possible. In this identification process, based on the considerations about the extent of the error, flow direction and the time and spatial distributions of flow, the parameters are finally defined. While some approaches have been tried so as to make this decision rational---for example, such an approach minimizing the summation of square errors at each well point (4), (5), (6). These approaches have been made from the statistical stand point. The equations (5) and (25) introduced here suggest that such the identification approaches are rational. Moreover, identification of the parameters based on the summation of the square errors is no more than to identify them in the energy dimension in the kinematical meanings. In other words, the discussion in the last chapter may clarify the kinematical significances of the evaluation standard based on the minimization of square error.

Now we consider the limitation of the accuracy of the simulation. Assume that we can completely fit the simulation results to the observed potentials, for

example, at the three well positions A, B and C in the region G as shown in Fig. 1-a. Considering the region G' shown in Fig. 1-b, the region surrounding the wells A, B and C be infinitely small. Equation (5) becomes then with respect to the region G' as follows

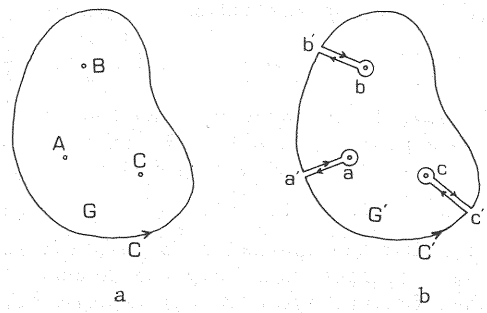

Fig. 1 Schematic representstion of the Domain

$$\frac{1}{2}\int_{G'}(\delta h)^2 dx_i = -\delta\int_{G'} L_g\, dx_i dt + \frac{1}{2}\int_{I}(\delta h)^2\, dx_i + \oint_{C'} + \oint_{a} + \oint_{b} + \oint_{c} + \int_{aa'} + \int_{a'a} + \int_{bb'} + \int_{b'b} + \int_{cc'} + \int_{c'c} \geq 0 \tag{27}$$

in which the integrands of the integral such as $\oint$, $\int_{aa'}$, ... are of the same form as those of the boundary integral in Eq.(5).

It corresponds to make the first term in the right hand side of Eq.(27) zero that we solve the ground water behaviors within the region. (although actually the factor concerning with L_g' remains). And since the simulation expresses the potentials at the three wells completely without errors, $\delta h=0$ at the neighbourhoods of the wells.

$$\oint_a = \oint_b = \oint_c = 0 \tag{28}$$

Moreover,

$$\int_{aa'} = -\int_{a'a}\ , \ldots. \tag{29}$$

Therefore, Eq.(27) is rewritten as

$$\frac{1}{2}\int (\delta h)^2\, dx_i = -\delta\int L_g'\, dx_i dt + \oint \frac{T}{S}\left\{\frac{\partial h^*}{\partial x_1}\frac{dx_2}{ds} - \frac{\partial h^*}{\partial x_2}\frac{dx_1}{ds}\right\}\delta h\, ds dt + \frac{1}{2}\int_I (\delta h)^2\, dx_i = E \tag{30}$$

The left hand side is the error of the simulation over the region and the right hand side is the summation of the errors of boundary and initial conditions and $-\delta\int L_g' dx_i dt \geq 0$. In other words, even though we have identified with error zero at the well-positions, if we have some errors in the boundary and initial conditions and so forth, the errors are distributed all over the region in the simulation. That is, Eq.(30) shows that "the accuracy of the simulation can not be much more than the error of the boundary and initial conditions etc." This is quite in the nature of things.

PROBABILITY OF ERROR IN THE REGION CONCERNED

As already shown in Eq.(26), the errors due to the various factors are distributed all over the region G. Of course, in the rigorous meaning, the distributions

of the error should be governed by such the equation as (25), however, in the actual situation since we have lack of the knowledge on the correct errors of the factors, we cannot help dealing with the probabilistic concepts in the discussions on the error distributions. In this chapter, we will discuss the statistical structure of the error of simulation over the whole region G.

Error probability in a small cell region

At first, we assume here the error E as known and decompose the concerned total region G into N-small cell regions with same area g=G/N. Then, the equation introduced in the previous chapter (26) may be rewritten as

$$\frac{1}{2}\sum \int_{g_i} (\delta h)^2 \, dx_i = E \geq 0 \tag{31}$$

If we assume the errors in the individual cell regions are mutually independent, then the problem results in such the problem that the total error E is distributed to N-cell regions. Moreover, it is also assumed here that the errors in the individual regions have only the discrete values. Write the number of cell regions as n_i, in which the value of error is given by

$$\frac{1}{2}\int_{g} (\delta h)^2 \, dx_i = \frac{1}{2}(\delta h)^2 g = \varepsilon_i \tag{32}$$

then the following relationships hold.

$$\sum n_i = N, \qquad \sum \varepsilon_i n_i = E \tag{33}$$

The probability of error ε_i in a specified cell region is obtained as follows (7), if the error ε_1 is distributed to the n_1 cell regions, ε_2 to n_2, ε_3 to n_3,....

$$p(\varepsilon_i : E) = \frac{n_i}{N} = \frac{1}{f}\exp(-\beta\varepsilon_i)$$
$$f = \sum \exp(-\beta\varepsilon_i) \tag{34}$$

in which the value β is defined by the condition (33). This expression (34) gives the probability of the event that a certain cell region has the error of ε_i, provided the total error E is known.

Measure of error and parameter β

We will make some reference to the parameter β. Differentiating the partition function f with respect to β, we obtain

$$\frac{\partial f}{\partial \beta} = -\sum \varepsilon_i \exp(-\beta\varepsilon_i) \tag{35}$$

Then the mean value of error $\bar{\varepsilon}$ is written as

$$\bar{\varepsilon} = \frac{E}{N} = \sum \varepsilon_i p(\varepsilon_i : E) = \frac{\sum \varepsilon_i \exp(-\beta\varepsilon_i)}{f} = -\frac{\partial f}{\partial \beta}/f = -\frac{\partial}{\partial \beta}\log f(\beta) \tag{36}$$

If we write the unit measure to observe the error as ε_o, the error can take only the discrete values such as

$$0, \ \varepsilon_o, \ 2\varepsilon_o, \ 3\varepsilon_o, \ 4\varepsilon_o, \ \cdots\cdots$$

In such the case, $f(\beta)$ may be expressed

$$f(\beta) = \sum \exp(-\beta\varepsilon_i) = 1 + \exp(-\beta\varepsilon_o) + \exp(-2\beta\varepsilon_o) + \cdots\cdots$$
$$= 1/\{1 - \exp(-\beta\varepsilon_o)\} \tag{37}$$

that is,

$$\log f(\beta) = -\log\{1 - \exp(-\beta\varepsilon_o)\} \tag{38}$$

This equation gives the following expression of the parameter β, with help of Eqs. (36) and (38).

$$\beta = \frac{1}{\varepsilon_o} \log \left\{1 + \frac{\varepsilon_o}{\bar{\varepsilon}}\right\} \tag{39}$$

This is the equation which gives the parameter β in terms of the unit measure ε_o and the mean error $\bar{\varepsilon}=E/N$.

ERROR PROBABILITY OF SIMULATED RESULT

In the actual simulation, the parameters S and T are identified by comparing the simulated values and the observed values at gauging wells in the concerned region G. In this chapter we evaluate the error probability involved in the final simulation results given through the identification and the various discretion.

In the previous chapter, the total error of the simulation E has been assumed as known, however, the value E itself is unknown in the actual simulation. In other words, the evaluation of simulation is nothing less than evaluating the value E. Moreover, as a clue to do so we have only the information of the observed and calculated values of potential at the gauging well points. Therefore, our task in this chapter is to evaluate the total error E in terms of those errors at the gauging points.

Here we consider such the situation that in the total region G consist of N-cell regions exist the three gauging wells J, K and L, of which observed potentials represent those in the each cell region including the wells. We write the each error in those well points as ε_j, ε_k and ε_l, that is,

$$\frac{1}{2}(\delta h)^2 g = \varepsilon_j,\ \varepsilon_k,\ \varepsilon_l \tag{40}$$

and the simultaneous event of these errors as ε_{jkl}. If the total error $E=E_m$ is given, then, the probability of this simultaneous event is written by Eq. (34).

$$p(\varepsilon_{jkl}: E_m) = e^{-\beta(\varepsilon_j+\varepsilon_k+\varepsilon_l)} / f^3 \tag{41}$$

Use the symbol $p(E_m)$ for the probability of the event $E=E_m$, then "the probability of such the event that total error $E=E_m$ and moreover the simultaneous event ε_{jkl} in the wells J, K and L occur" may be given by $p(\varepsilon_{jkl}:E_m)p(E_m)$. Therefore, provoded the total error has the possibility to have as its value the discrete M values E_1, E_2, ... E_M, then the posterior probability $p(E_m:\varepsilon_{jkl})$ of the event $E=E_m$ under the condition of the event ε_{jkl} is obtained by Bayse theory as follows.

$$p(E_m: \varepsilon_{jkl}) = \frac{p(E_m)p(\varepsilon_{jkl}: E_m)}{\Sigma\, p(E_i)p(\varepsilon_{jkl}: E_i)} \tag{42}$$

In this equation is involved the prior probability $p(E_m)$ but at this stage we have no knowledge on the event $E=E_m$, except its possibility to have M discrete values E_i (i=1,2,...,M). In other words, $p(E_m)$ is unknown quantity. For this problem, however, provided we have no knowledge on the event E_m, the most rational evaluation is to give the same probability 1/M to each M event E_i. Then we come conclusively to the following mathematical form

$$p(E_m: \varepsilon_{jkl}) = \frac{p(\varepsilon_{jkl} : E_m)}{\Sigma\, p(\varepsilon_{jkl}: E_i)} \tag{43}$$

This is what gives the probabilistic evaluation of the total simulation error E over the region concerned, based on the information of appearance of the simultaneous error event ε_{jkl} at the three gauging stations. As will be easily understood, in this equation $p(\varepsilon_{jkl}:E_i)$ is given by Eq.(41) with $\beta=\beta_i$ defined by Eq.(39) for the value $\bar{\varepsilon}=E_i/N$.

In this equation the probability is defined for the total error $\underline{E}$ over the region, but if we substitute the values E_i and E_m in the equation by $\bar{\varepsilon}_i = E_i/N$ and $\bar{\varepsilon}_m = E_m/N$, respectively, the error probability is expressed in terms of the mean error $\bar{\varepsilon}_m$.

Moreover, if we have k observation wells j1, j2, ..., jk and in those points the errors are observed as ε_{j1}, ε_{j2}, ..., ε_{jk}, (the simulataneous event $\varepsilon_{j12...k}$), then we obtain

$$p(\bar{\varepsilon}_m : \varepsilon_{j12..k}) = \frac{p(\varepsilon_{j12..k} : \bar{\varepsilon}_m)}{\sum_i^M p(\varepsilon_{j12..k} : \bar{\varepsilon}_i)} = \frac{(1-e^{-\beta\varepsilon_m\varepsilon_o})^k \exp(-\beta_m \cdot \sum_s^k \varepsilon_{js})}{\sum_i^M (1-e^{-\beta_i\varepsilon_o})^k \exp(-\beta_i \cdot \sum_s^k \varepsilon_{js})} \tag{44}$$

in which

$$\beta_i = \frac{1}{\varepsilon_o} \log_e \left\{1 + \frac{\varepsilon_o}{\bar{\varepsilon}_i}\right\} \tag{45}$$

As described above, we have obtained "the probabilistic index which evaluates the simulation error E or its mean value $\bar{\varepsilon}$ all over the region based on the errors ε_j between the observed and simulated values".

NUMERICAL EXAMPLES AND DISCUSSION

Figure 2 shows a numerical examples of the probability $p(\bar{\varepsilon} : \varepsilon_{jkl})$ in the case of three gauging well points. Although the probability is essentially defined only for the discrete values $\bar{\varepsilon}=\bar{\varepsilon}_i$ $(=i\varepsilon_o,\ i=1,2...M)$, it is shown here in the figure as the continuous curve for the unit measure ε_o=0.5 and 0.25 in Figs. 2-a and 2-b, respecrively. The probabilities appear very different in its values in Figs. 2-a and 2-b, but this difference depends on the fact that the probabilities are defined for ε_i=0.5, 1.0, 1.5,.... for Fig. 2-a and for ε_i=0.25, 0.5, 0.75,... for Fig. 2-b. While the dimension of the unit measure ε_o is L^{4i} (c.f. Eq.(22)), $\beta\varepsilon$ is nondimensional, since the area of the cell region g is involved in the quantity ε, ε_o and β (c.f. Eq.(39)).

Because Eq.(44) gives the same value for the same value of $\varepsilon_j+\varepsilon_k+\varepsilon_l$, then Fig. 2 is shown with the parameter $\varepsilon_j+\varepsilon_k+\varepsilon_l$. According to the probability theory it is a matter of course that the probability has its peak value at the value $\bar{\varepsilon} = (\varepsilon_j+\varepsilon_k+\varepsilon_l)/3$. From the figure the sharp distribution is found for the smaller value

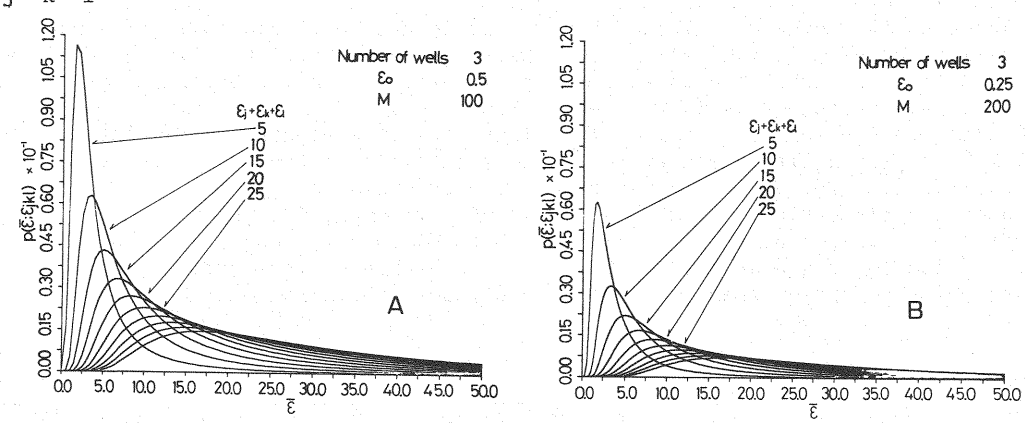

Fig. 2 Probability of Simulated error

of $\varepsilon_j+\varepsilon_k+\varepsilon_l$ and we can judge approximately how much the error of the final simulation for the region concerned is. On the other hand, larger the value $\varepsilon_j+\varepsilon_k+\varepsilon_l$, lower the peak of probability, and then it becomes rather difficult to evaluate even whether the simulation has large or small error.

How the probability $p(\bar{\varepsilon}: \varepsilon_{jl2\cdots k})$ depends on the number of observation wells k, by which the accuracy of simulation is evaluated, is shown in Fig. 3. In the figure, for the easier understanding, the probability is calculated for the case of $\varepsilon_{jl}=\varepsilon_{j2}=\varepsilon_{j3}=\cdots$ $\cdots\varepsilon_{jk}=3.0$. From the figure, it may be concluded that we can judge much more accurately the accuracy (or error) of the final simulation of ground water behaviors, based on the information at increasing number of gauging wells.

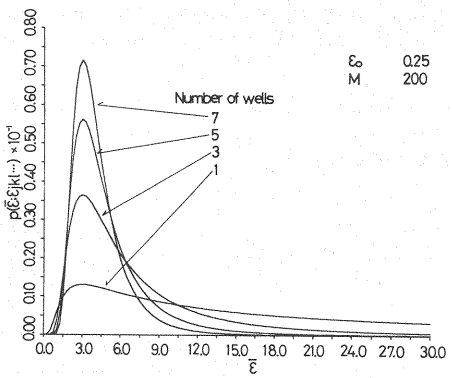

Fig. 3 Probability of Simulated error

In the evaluation of simulation result by the approach discussed above, there still remains very important problem whether we should judge a certain final simulation as "very good" or "bad" by the probability here introduced. In the real problem this is the most important one, but the standard of this judgement should be essentially left to the other problem. In other words, the judgement must be made concerning with the original object of the simulation. The accuracy required for the simulation result should be decided depending on the aims as to whether the simulation is carried out in order to be served for the something like a master planning, to discuss the detailed flow behaviors concerning with an actual construction works, moreover in physical meanings, to discuss only the potential distributions and so forth. Therefore, the error probability introduced here should be applied, considering whether the simulation comes up to the expectations required.

CONCLUSIONS

The methodology how we should evaluate the simulation of ground water behavior has been discussed, and the conclusions obtained through the present paper are summarized as follows.

(1) Through the variational formulation of ground water behaviors, it has been clarified, that in what form of mathematical structure the errors involved in the parameters, boundary and initial conditions affect on the total error of the simulation result. Moreover, it has been demonstrated that the identification process based on the square error of the potential corresponds to that in terms of the energy dimension.

(2) Based on the mathematical and physical structure mentioned above, the probability has been derived which can be served to evaluate the accuracy of final simulation as a whole with help of the information on the potential errors at some observation-wells in the region concerned.

This research work is still at the first step to evaluate the simulation result. The probability demonstrated here is to estimate the simulation error only at a certain time and not evaluate it in the time evolution. These problems are now under the discussion. However, what we want to emphasize here is that the extention of this research may become a clue to discuss the objective evaluation of simulation technique and also the better field-observation-system concerning with the aims of the investigation of ground water behaviors.

If we use enough large number M, the probability given by Eq.(44) may be considered as independent on M.

REFERENCES

1. Takagi, F.: A study on the behavior of basin water by means of the variational technique, Proc. of Japan Society of Civil Engineers, No. 185, pp. 71-81, 1971.
2. Takagi, F.: Physical significances of the variational principle for the behaviour of basinwide water, Proc. of Japan Society of Civil Engineers, No. 227, pp. 27-32, 1974 (in Japanese).
3. Takagi, F. and M. Matsubayashi: Problems on the application of the variational principle with respect to the hydrologic behavior of basinwide water, Proc. of Japan Society of Civil Engineers, No. 274, pp. 31-40, 1978 (in Japanese)
4. Neuman, S.P. and Yakowitz, S.: A statistical approach to the inverse problem of aquifer hydrology, 1. Theory, Water Resources Research, Vol. 15, No. 4, pp. 845-860, 1979
5. Neuman, S.P., G.E.Fogg and E.A.Jacobson: A statistical approach to the inverse problem of aquifer hydrology, 2 Case study, Water Resources Research, Vol. 16, No. 1, pp. 33-58, 1980
6. Neuman, S.P.: A statistical approach to the inverse problem of aquifer hydrology, 3. Improved solution method and added perspective, Water Resources Research, Vol. 16, No. 2, pp. 331-346, 1980
7. For example, Landau, L.D. and E.M.Lifshitz: Course of theoretical physics, Vol. 5, Statistical Physics, Pergamon Press, 1959.

APPENDIX - NOTATION

The following notations are used in this paper:

- b = thickness of the aquifer;
- C = boundary around the region G;
- E = error in total region G;
- f = $\Sigma \exp(-\beta\varepsilon_i)$
- g = area of cell region (=G/N)
- G = total area concerned and its area;
- h = piezometric head of ground water;
- k = permeability coefficient, or number of observational wells;
- L_g = local potential;
- n_i = number of cell region g in which error ε_i occurs;
- N = number of cell regions in total region G;
- r = vertical inflow flux (intensity) to the aquifer;
- s = counterclockwise distance along C;
- S = storage coefficient;
- t = time;
- T = transmissibility (=kb);
- x_i = horizontal coordinates, x_1 and x_2;
- β = parameter expressing distributions of simulation error;
- ε_i = error in cell region g
- $\bar{\varepsilon}$ = mean error of the simulation (=E/N).