Identification of random field for ground stiffness by data assimilation based on surface wave method and sounding tests

Shin-ichi Nishimura[1#], *Yuxiang* Ren[2], *Toshifumi* Shibata[1] and *Takayuki* Shuku[1]

[1]*Okayama University, Fuculty of Envirometal Life and Natural Science, 3-1-1 Tsushima-naka. Kita-ku, Okayama, Japan*
[2]*Okayama University, Graduate School of Environmental and Life Science*
[#]*Corresponding author: theg1786@okayama-u.ac.jp*

ABSTRACT

The surface wave method (SWM) and the screw weight sounding (SWS) are employed as a geophysical exploration method and a sounding test, respectively to identify the spatial distribution of the stiffness of an earth-fill dam in the present study. The ensemble Kalman filter (EnKF) is used as a data assimilation technique. It can estimate the spatial distribution of the Young's modulus as the stiffness of an earth-fill dam by assimilating the travel time to the first arrival of the surface waves. By the ensemble data assimilation, the measured data from the SWM is applied to simultaneously estimate the Young's modulus and evaluate the uncertainties. The SWS results are employed as the prior information to generate the initial ensemble through the sequential Gaussian simulation (sGs). Proposed method has been applied to the actual data of the SWM and the SWS measured at an earth-fill dam site. Consequently, it has been clarified the proposed approach could identify the appropriate random field of Young's modulus.

Keywords: surface wave method; screw weight sounding; data assimilation; ensemble Kalman filter.

1. Introduction

Although boring and standard penetration tests are usually conducted to examine the inside of geo-structures, the interval of the test points must be sparse due to low efficiency and the possibility of damage to the structures. Consequently, weak locations may be overlooked. To deal with these shortcomings, the screw weight sounding test (SWS) (JGS 2015) is employed in this study as the static sounding method for obtaining detailed information on the inside of an earth-fill and the underneath ground. The SWS is advantageous in that it makes short-interval exams possible, which are highly efficient, and it causes little damage to the structures. Additionally, the surface wave method (SWM), which is the one of the geophysical survey methods, is employed to obtain information for a wider area, since it is a much more efficient approach. The SWM has been widely used in investigation. For the SWM, almost traditional inversion processes lead to a deterministic spatial distribution of the S-wave velocity (Jongmans et al. 1993). That means the uncertainty of the estimation cannot be evaluated. To examine a wide area of an earth-fill efficiently and accurately, two methods, namely, the SWS and the SWM, are theoretically combined with a data assimilation method.

Although the sounding results are reliable to estimate the stiffness of the ground, only the information of the point estimate is available. While, the geophysical exploration can present the spatially averaged information of the stiffness. In this study, two kinds of the test results are synthesized by the data assimilation technique, and the technique is applied to the identify the spatial distribution, namely random field of the stiffness. The ensemble Kalman filter (EnKF) (Evensen 2009) is used as a data assimilation technique. It can estimate the spatial distribution of the Young's modulus as the stiffness of an earth-fill dam by assimilating the travel time to the first arrival of the surface waves.

Using the Monte Carlo method, the EnKF can calculate the probability distributions of the parameters with a group of realizations. Compared to other data assimilation methods, such as the particle filter (PF), due to the hypothesis whereby the probability density function (pdf) is Gaussian in the EnKF, a relatively accurate probability distribution can still be obtained when the number of realizations is small. It brings the computational cost of the forecast model applied to each ensemble member to an acceptable level.

The EnKF method is widely used in meteorology and oceanography (Annan et al. 2005, Ueno et al. 2007), and is gradually being applied in mechanical engineering (Akita et al. 2010), petroleum engineering (Liu and Oliver 2005), and geotechnical engineering (Franssen and Kinzelbach 2008). By assimilating the measurements, the EnKF can modify the poorly known parameters and lead to best-guess estimates. There should be a correlation between the measurement data and the parameters to be updated. For different models, appropriate measurement data must be carefully chosen. For instance, in the study of Evensen et al. (2007), the well-log measurements and production rates of oil, gas, and water were employed to estimate the fluid contacts,

porosity, and permeability. Akita et al. (2010) utilized the displacement of a structure to estimate the spring constant. Tao et al. (2020) used EnKF to predict the soil settlement of an airport road. Pandurangan et al. (2015) used the data of tiltmeter to estimate fracture growth parameters. In contrast, there have been relatively few studies which applied the EnKF for the parameters of the soil strength or rigidity. The study of Caballero et al. (2018) involved Young's modulus of ground. In this study, the Young's modulus of an earth-fill dam is identified using the travel time to the first arrival of the surface waves, which has a strong correlation with the stiffness of the ground.

By the ensemble data assimilation, the measured data from the SWM is applied to simultaneously estimate the Young's modulus and evaluate the uncertainties. The SWS results are employed as the prior information to generate the initial ensemble through the sequential Gaussian simulation (sGs) (Deutsch and Journel 1992). In the experiments of assimilation, it has been shown that the reproducibility of the parameter field is enhanced by this initial ensemble generation method, and that the uncertainties of the identified parameters can be reduced by the assimilation.

2. Surface wave method

The measurements obtained by the SWM are utilized for the assimilation. According to the study of Hayashi et al. (2001), Fig.1 shows the arrangement of the geophones and shot points. Regardless of the total length, the geophones in this study are always arranged at 2 m intervals, and the shot points are placed between the geophones and at the ends of them, so that the number of shot points is 1 more than the geophones. During the in-situ tests, a sledgehammer is used to generate surface waves by striking it against the shot points, and the shot-gather obtained by recording the waveforms is used as the measurement data for the corresponding times.

In the data assimilation to analyze the surface wave, the finite element method (FEM) is employed. The FEM allows the boundary conditions of different models to be taken into consideration, and can introduce the heterogeneity of the ground into the models easily. The traditional data analysis method of the SWM uses the dispersion properties of the surface waves to calculate the wave velocity at different depths. Inasmuch as high-frequency waves can be filtered out by a larger size grid (Zerwer 2002) in a finite element simulation, fairly tiny meshes are required to meet the conditions for an analysis of the dispersion. The fact that the model size should be consistent with real tests leads to models with lengths of 30 m to 40 m. For models that perfectly simulate high-frequency waves, the computational cost is too large for practical use. Therefore, only the first arrival time of the surface waves is used as the measurement in this study in order to avoid having to assimilate the frequency of the surface waves, which allows for a model with relatively large grids to be used. The first arrival time of each waveform needs to be determined manually because of the noise included in the real surface wave data. As the noise in realizations can be eliminated with appropriate boundary conditions, the first arrival time can be determined automatically.

3. Screw weight sounding test

The spatial distribution of the strength parameters of decrepit earth-fill dams is discussed, and an identification method for the distribution is proposed. Generally, the identification of spatial correlation of the soil parameters is very difficult, since the number of the sampling points is limited to model the spatial variability. The difficulty can be solved by the sounding tests, since they can be conducted at widely spread points. Although the strength of earth-fill dams is generally predicted from standard penetration test (SPT) *N*-values, screw weight sounding (SWS) tests are employed in this research as a simpler method to obtain the spatial distribution of the *N*-values. The SWS has two advantages, compared with other major sounding methods.

1) The exams with very sort interval is possible, since it is a static souring and does not give damage to old and deteriorated earth-fills. This is very advantageous to identify the correlation lengths.

2) The examinations is possible in the narrow spaces, or in the test sites difficult to access with the large equipment such as the electric cone penetration test (CPT).

The relationship between the SPT *N*-value and the SWS *N*-values is modelled;

$$N_{SWS} = 0.067N_{SW} + 0.002W_{SW} \tag{1}$$

in which N_{SWS} is the *N*-value derived from SWS, N_{SW} is the number of half rotations and W_{SW} is the total weight of the loads.

Next, the N-values are required to be transformed in to Young's modulus. A roughly suitable model which makes the measurements of initial spatial distribution of the Young's modulus get close to the real measurements by the data assimilation could be chosen for convenience. The transformation model used in this study is

$$E = cN \tag{2}$$

where c is a constant transformation coefficient. It is common for c to be 700 or 2800 (e.g. for borehole load tests or horizontal load tests).

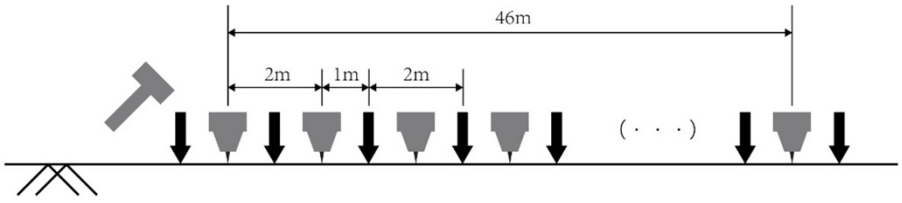

Figure 1. Arrangement of geophones and shot points.

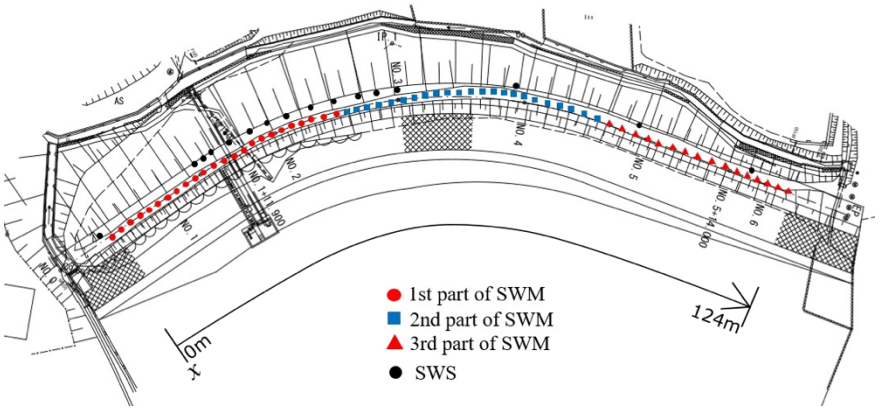

Figure 2. Plan view of SWS and SWM test locations.

4. Identification of random field

4.1. Identification method

The *N*-value distribution derived from the SWS tests is spatially interpolated with the Monte Carlo method (MCM). The random field theory has been used to depict the spatial variability of soil properties. Nishimura et al. (2016) employed the geostatistical method with data from cone penetration tests to create a reasonable *N*-value field. The spatial distribution of the ground in this section is created by the analogous method. The data from the SWS tests, taken from the same site as the SWM tests, are used to establish a statistical model for the *N*-value field (the spatial distribution of SWS *N*-value).

A representative variable for the soil propertieswhich is log*N* in the present study, *s*, is defined by Equation (3) as a function of location $\mathbf{X}=(x, z)$. Variable *s* is assumed to be expressed as the sum of mean value *m* and random variable *U*, which is a normal random variable in this study.

$$s(\mathbf{X}) = m(\mathbf{X}) + U(\mathbf{X}) \tag{3}$$

The random variable function, $s(\mathbf{X})$, is discretized spatially into random vector $\mathbf{s}^t=(s_1,s_2,...,s_M)$, in which s_k is a point estimation value at location $\mathbf{X}=(x_k, z_k)$. The soil parameters, which are obtained from the tests, are defined here as $\mathbf{S}^t=(S_1,S_2,..., S_M)$. Symbol *M* signifies the number of test points. Vector **S** is considered to be the realization of random vector $\mathbf{s}^t=(s_1,s_2,...,s_M)$. If variables $s_1, s_2,...,s_M$ constitute the *M* - variate normal distribution, the probability density function of *s* can then be given by the following equation:

$$f_S(\mathbf{s}) = (2\pi)^{-\frac{M}{2}} |\mathbf{C}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{s}-\mathbf{m})^t \mathbf{C}^{-1}(\mathbf{s}-\mathbf{m})\right\} \tag{4}$$

in which $\mathbf{m}^t=(m_1,m_2,...,m_M)$ is the mean vector of random function $\mathbf{s}^t=(s_1,s_2,...,s_M)$; it is assumed to be the following regression function. In this research, a 3-D statistical model is considered, namely, horizontal coordinate *x*, which is parallel to the embankment axis, vertical coordinate *z*, and another horizontal coordinate, *y*, which is perpendicular to the embankment axis. *x* and *y* are introduced here. The element of the mean vector is described as

$$m_k = a_0 + a_1 x_k + a_2 z_k + a_3 x_k^2 + a_4 z_k^2 + a_5 x_k z_k \tag{5}$$

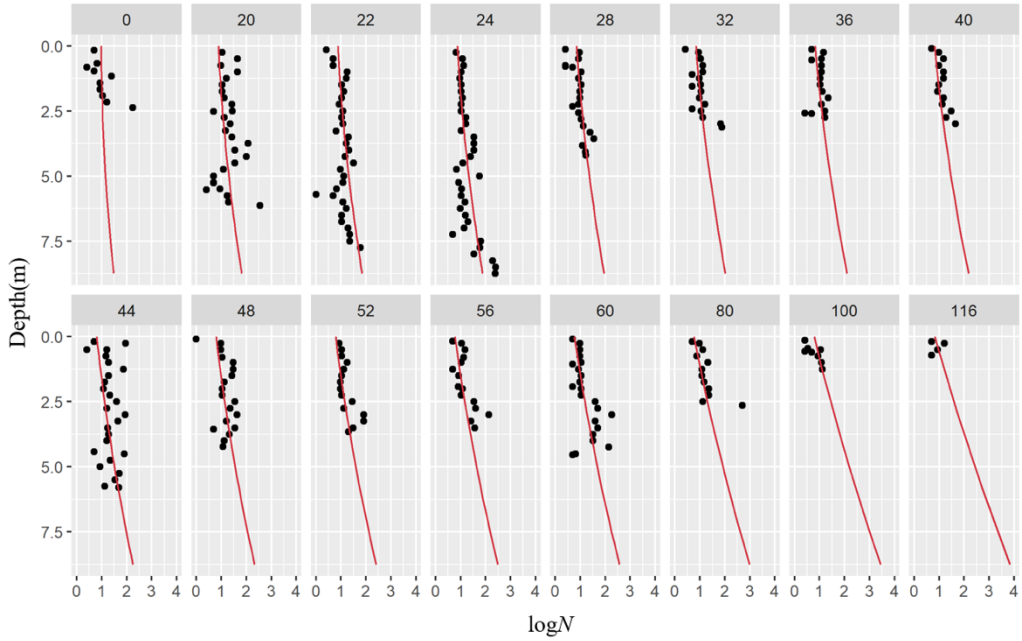

Figure 3. *N*-values derived from SWS and trends.

in which (x_k, z_k) means the coordinate corresponding to the position of parameter s_k, and a_0, a_1, a_2, a_3, a_4, and a_5 are the regression coefficients.

C is the $M \times M$ covariance matrix, which is selected from the following four types in this study:

$$\begin{aligned}
&\mathbf{C} = \left[C_{ij}\right] = \\
&\sigma^2 \exp\left(-\left|x_i - x_j\right| / l_x - \left|z_i - z_j\right| / l_z\right) \quad \text{(a)} \\
&\sigma^2 \exp\left\{-\left(x_i - x_j\right)^2 / l_x^2 - \left(z_i - z_j\right)^2 / l_z^2\right\} \quad \text{(b)} \\
&\sigma^2 \exp\left\{-\sqrt{\left(x_i - x_j\right)^2 / l_x^2 + \left(z_i - z_j\right)^2 / l_z^2}\right\} \quad \text{(c)} \\
&N_e \sigma^2 \exp\left(-\left|x_i - x_j\right| / l_x - \left|z_i - z_j\right| / l_z\right) \quad \text{(d)} \\
&i, j = 1, 2, \cdots, M \\
&\begin{cases} N_e = 1 & (i = j) \\ N_e \le 1 & (i \ne j) \end{cases}
\end{aligned} \tag{6}$$

in which the symbol [C_{ij}] signifies an *i-j* component of the covariance matrix, σ is the standard deviation, and l_x and l_z are the correlation lengths for the *x* and *z* directions, respectively. Parameter N_e is related to the nugget effect. Akaike's Information Criterion, AIC, is defined by Equation (7) considering the logarithmic likelihood.

$$\begin{aligned}
&\mathrm{AIC} = -2 \cdot \max\left\{\ln f_S(\mathbf{S})\right\} + 2L = M \ln 2\pi \\
&+ \min\left\{\ln\left|\mathbf{C}\right| + (\mathbf{S} - \mathbf{m})^t \mathbf{C}^{-1} (\mathbf{S} - \mathbf{m})\right\} + 2L
\end{aligned} \tag{7}$$

in which L is the number of unknown parameters included in Equation (4). By minimizing AIC (MAIC), the regression coefficients of the mean function, the number of regression coefficients, the standard deviation, σ, a type of covariance function, the nugget effect parameter, and the correlation lengths are determined. In other words, the determined parameters and the selected covariance function correspond to the minimum AIC.

4.2. Spatial distribution of N-value on studied site

Fig.2 shows the plan view and the SWS test points of the studied site of the earth-fill dam located in Okayama. The test results are depicted in Fig.3. The trend function of log*N*, obtained by the method described in the previous section, is as follows:

$$\begin{aligned}
&m = 0.912 - 0.004x - 0.028z \\
&+ 0.00003x^2 + 0.013z^2 + 0.0019xz
\end{aligned} \tag{8}$$

This trend function, namely, the regression curve drawn for each measurement point, is shown in Fig.3 with the measured *N*-values. The black points are the logarithms of the *N*-values and the red curve is the regression curve given by Equation (8). The number at the top of each plot indicates the location of the measurement point, which is the value of x in Equation (8). It can be seen that the trend function calculated by the MAICE is relatively consistent with the measurements.

The horizontal correlation distance is identified as 15.7 m, while the vertical correlation distance is 1.24 m respectively, in this study, and the covariance function is as follows:.

$$\begin{aligned}
&C = \left[C_{ij}\right] = N_e \sigma^2 \exp\left(-\frac{\left|x_i - x_j\right|}{15.7} - \frac{\left|z_i - z_j\right|}{1.24}\right) \\
&\sigma = 0.383 \\
&N_e = 0.593 \times 0.690 \quad \left(\left|x_i - x_j\right| \ne 0, \left|z_i - z_j\right| \ne 0\right) \\
&N_e = 0.593 \quad \left(\left|x_i - x_j\right| \ne 0, \left|z_i - z_j\right| = 0\right) \\
&N_e = 0.690 \quad \left(\left|x_i - x_j\right| = 0, \left|z_i - z_j\right| \ne 0\right)
\end{aligned} \tag{9}$$

Based on this statistical model, the random fields of *N*-values are generated by the sGs and converted to Young's modulus fields.

5. Ensemble Kalman filter

The inversion process of the EnKF consists of the evolution step and the update step. In the evolution step, a group of numerical models (the ensemble) with different initial states are used to forecast the evolution of a real system to calculate those measurements. Then, the states of the numerical models can be calibrated by the difference between the measurements of the numerical models and those of the real system in the update step. The EnKF is successful in many high-dimensional, nonlinear, and non-Gaussian application (Katzfuss, et al. 2016). In the present study, the finite element method (FEM) is used to solve the numerical models and predict those measurements. There are two kinds of analysis schemes for updating, that is, the stochastic analysis scheme and the deterministic analysis scheme. The stochastic analysis scheme is used in the present study; it includes

$$\hat{x}_{t|t}^{(i)} = \tilde{x}_{t|t-1}^{(i)} + K_t\left(y_t - \tilde{y}_t^{(i)}\right), \quad i = 1, \ldots, N \tag{10}$$

where $\tilde{x}_{t|t-1}^{(i)}$ is a set of forecast state vectors, which consists of all the state variables that needed to be estimated at time step $t-1$ and $\hat{x}_{t|t}^{(i)}$ is a set of updated state vectors at time step t. N is the number of ensemble members.

The state vectors here actually refer to the joint state vectors, which include the m parameters to be estimated (Young's modulus) and the n states to be assimilated (first arrival time of the surface waves). y_t is the measurement vector (n-dimensional), which consists of all the measurements at time step t, and $\tilde{y}_t^{(i)}$ is the measurement vector of the ith ensemble, given by the following linear observation equation:

$$\tilde{y}_t^{(i)} = H_t \tilde{x}_t^{(i)} + w_t^{(i)}, \quad i = 1, \ldots, N \tag{11}$$

where H_t is an observation operator used to extract the predicted measurements from the joint state vector and $w_t^{(i)} \sim N(0, R_t)$ is the perturbation component. In Equation (12), K_t is the Kalman gain which is given by

$$\hat{K}_t = \hat{V}_{t|t-1} H_t^T \left(H_t \hat{V}_{t|t-1} H_t^T + \hat{R}_t \right)^{-1} \tag{12}$$

where $\hat{V}_{t|t-1}$ is the ensemble covariance at time step t; it is represented by

$$\hat{V}_{t|t-1} = \frac{1}{N-1} \sum_{i=1}^{N} x'^{(i)}_{t|t-1} x'^{(i)T}_{t|t-1} \tag{13}$$

$$x'^{(i)}_{t|t-1} = x^{(i)}_{t|t-1} - \frac{1}{N} \sum_{i=1}^{N} x^{(i)}_{t|t-1} \tag{14}$$

The $\hat{R}_t$ in Equation (12) is the covariance matrix of perturbation, given by

$$\hat{R}_t = \frac{1}{N-1} \sum_{i=1}^{N} w_t^{(i)} w_t^{(i)T} \tag{15}$$

When the number of measurements is larger than the number of ensemble members (i.e., $n > N$) the inverse matrix in Equation (12) cannot be calculated. Accordingly, the generalized inverse is applied, even when $n < N$. The utilization of the generalized inverse is elaborated in the study of Evensen (2009). The joint state vector is given by

$$x_t = \left[E_t, a_t \right]^T \tag{16}$$

where vector E_t consists of all Young's moduli and a_t consists of the measurements of the first arrival time of the surface waves at time t. According to Equation (10), not only Young's moduli, but also the first arrival time of the surface waves, are calibrated in the updating process. However, only the updated parameters are propagated, and the state is recalculated from the simulation, which is presented by the function f in Equation (17), namely, the FEM at each time step.

$$a_t = f\left(E_t \right) \tag{17}$$

In the updating process, the EnKF may lead to unrealistic geological parameters, for instance, a negative Young's modulus. Consequently, the logarithm (base 10) of Young's modulus is updated, and then all the updated Young's moduli which will be used in the simulation are positive. The original measurements give rise to an error covariance matrix, which excessively modifies the logarithm of Young's modulus. In other words, compared with the logarithmic state vectors, the increments in the measurements are too large. Accordingly, an attenuation factor β, greater than zero and less than one, is applied to reduce the error covariance matrix, which is determined by the measurements and the increments in measurements. A rational size of β depends on the numerical model (i.e., Equation (17)). It can be determined by trial and error so that the increments in the state vectors in each updating process are sufficiently small. The joint state variable being updated is

$$x_t = \left[\log E_t, \beta a_t \right] \tag{18}$$

For the time evolution, the parameters of the physical system can be considered as time-invariant because the data acquisition in the SWM can be accomplished in a short time. Therefore, the evolution equation of Young's modulus is given by

$$E^{(i)}_{t|t-1} = E^{(i)}_{t-1|t-1} \tag{19}$$

and the state at time $t + 1$ is calculated by a numerical simulation as well.

$$a^{(i)}_{t|t-1} = \hat{a}^{(i)}_{t-1|t-1} \tag{20}$$

where $\hat{a}$ is the vector of the first arrival time given by the simulation.

6. Application result for studied site

6.1. Analytical model

In this section, a model to simulate SWM tests for the studied site, is utilized. Corresponding to the arrangement of the geophones, there are 17 measurement points. There are 18 shot points in total, and the shot points are struck sequentially. Each hit is regarded as a time step, in which the EnKF should be updated. The height of the grid used for the analysis is 1 m, with a total height of 9 m. Within a radius of 1 m of a shot point, the length of the grid is set to be very small (0.05 m) in the horizontal direction, and the rest of the grid is 1 m in length. It is necessary to resort to this mesh division for removing the parasitic oscillation in the waves. The FEM model is shown in Fig.4.

Based on trial and error, this grid size ensures a relatively acceptable simulation and minimizes the computational cost. As such, the grid of the FEM model is rebuilt at each time step; however, a static grid is necessary for applying the EnKF updating. Therefore, the

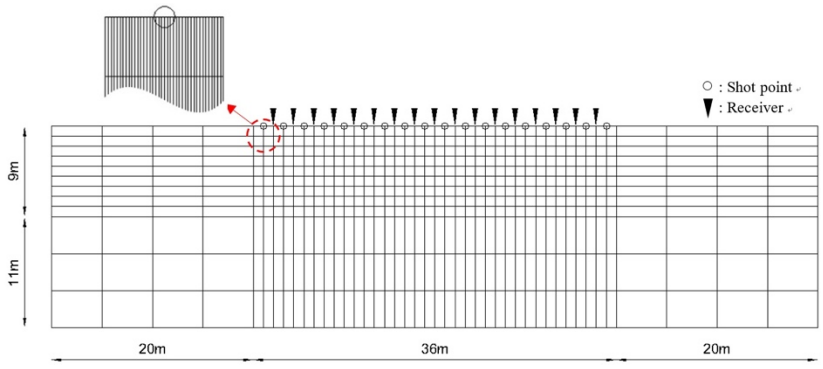

Figure 4. Example of mesh design of FEM model.

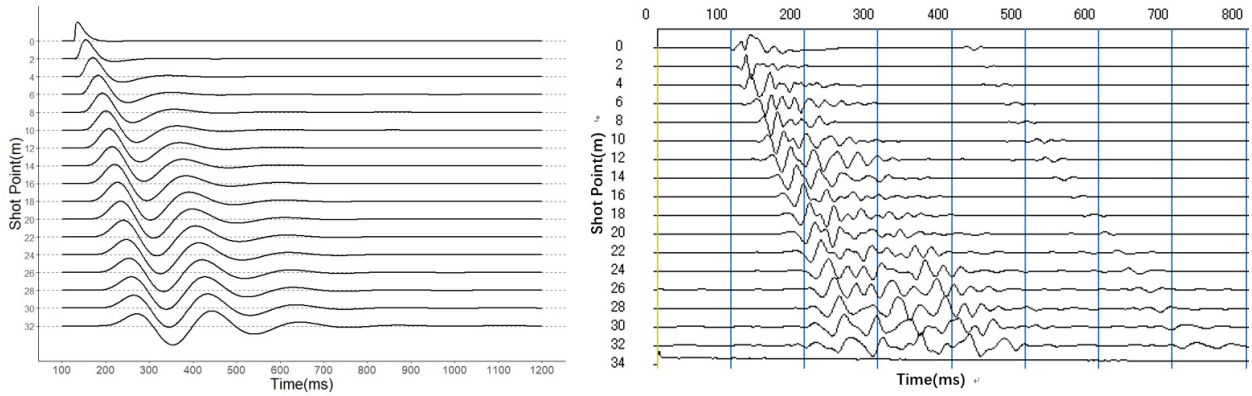

Figure 5. Example of simulation of surface wave propagation (left) and real measurement (right)

grid created for the sGs during the formation of the initial ensemble is used in the following updating, and the parameters of the elements for the FEM model in the next time step are determined by applying the updated grid. The Young's modulus of every element in the FEM model actually equals the average Young's modulus of every element in the updated grid, which falls within the search square centered at the location of each element. The side length of the search square is 1 m in this experiment.

In a homogeneous medium, shear wave velocity V_s is given by

$$V_s = \sqrt{\frac{E}{2(1+\nu)\rho}} \tag{21}$$

where ν is Poisson's ratio, E is Young's modulus, and ρ is the mass density. Compared with Young's modulus, which may change in a large range, the ranges of Poisson's ratio (0.3 ~ 0.4 for sandy soil) and mass density (19 kN/m^3 for most of the sandy soil) are relatively small. Therefore, the geological parameters, other than Young's modulus, are assumed to be known in order to limit the degrees of freedom of the unknowns. Poisson's ratio is set to be 0.35, and the unit weight is set to be and the unit weight is set to be 19 kN/m^3. The time increment for one step of the FEM is 0.001 s.

Fig.5 gives an example of a simulation of the surface wave propagation. It can be seen that the characteristics of the waveform roughly match those of a real surface wave, but the calculated frequency of the wave is lower than the measured surface wave due to the grid size, as has already been mentioned. However, the first arrival time of the waves is less influenced by the grid size or the frequency.

The attenuation factor β of the measurements in Equation (18) is set to be 0.005, and the distribution of the perturbation of the realizations, w_t, is $N(0,0.12)$. In addition, the residual sum of the squares (*RSS*) between the ensemble mean of the measurement and the reference measurement is used as the criterion for determining whether or not the model response has been reproduced.

$$RSS = \sum_{i=1}^{n_t} \left(y_{t,i} - \hat{y}_{t,i} \right)^2 \tag{21}$$

where n_t is the number of measurements at time point t, and $y_{t,i}$ and $\hat{y}_{t,i}$ are the reference measurement vector and ensemble mean of the measurement vector, respectively.

6.2. Analytical results

The SWM tests were implemented on the studied site, consisting of three parts as shown in Fig.2. The length of the first two parts is 46 m, and each part is arranged with 24 geophones and 25 shot points. The length of the third part is 32 m. For the relationship between the sites of the SWS tests and those of the SWM tests, the first part of the waveform data obtained from the SWM tests corresponds to the section from 0 to 46 m (1st part) in Fig.2. The second part corresponds to the section from 46 to 92 m (2nd part), and the third part corresponds to the section from 92 to 124 m (3rd part). The plan view of the SWM tests is shown in Fig.2 and an example of the waveform is shown in Fig.5. Realizations are applied with the same perturbations (i.e., w_t in Equation (11)). To improve the accuracy of the estimates, the maximum penetration depth in the SWS tests is utilized to distinguish the dam and the bedground, which is treated as prior information for the EnKF. The portion below this maximum penetration depth is considered to be hard bedground.

The Young's moduli of those grid cells would not be updated by the EnKF directly, but characterized by the grid cells. In this section, the Young's modulus of the bedground part is fixed to be five times the average of the dam part. According to the maximum penetration depth, shown in Fig.3, the height of the bedground part increases significantly with the increase in horizontal coordinate x. The transformation coefficient c is assumed

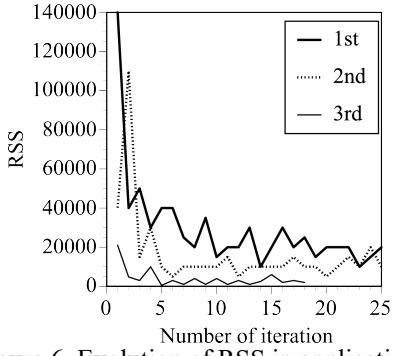

Figure 6. Evolution of RSS in application.

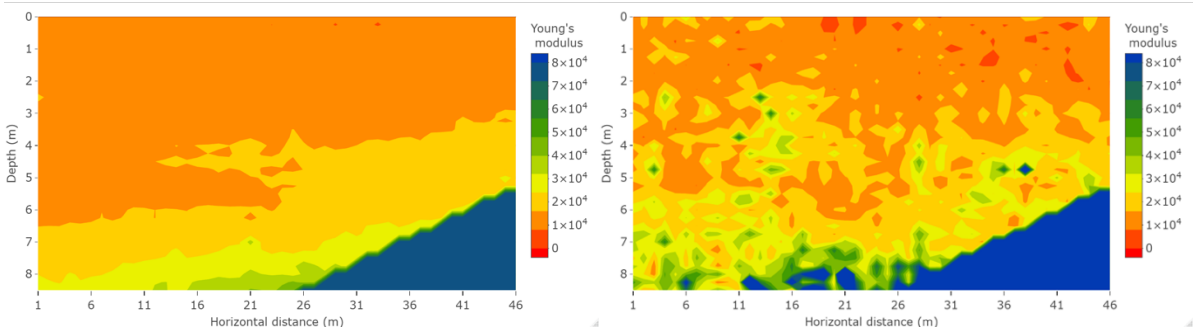

Figure 7. Young's moduli in 1st part at initial step (left) and final step (right).

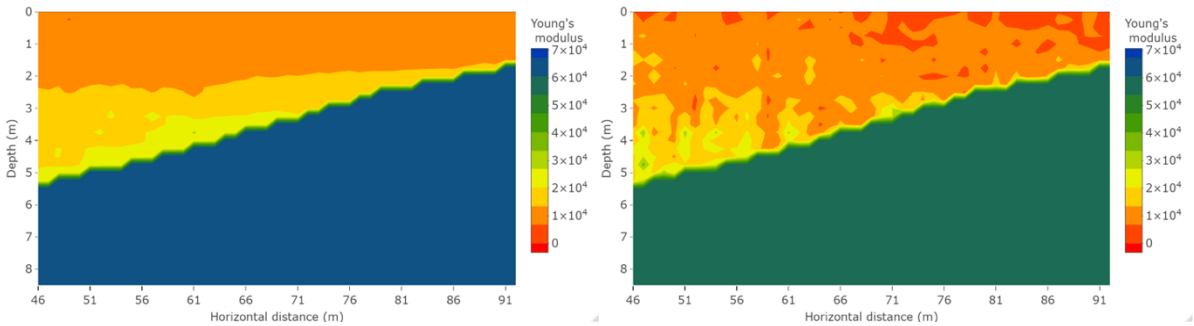

Figure 8. Young's moduli in 2nd part at initial step (left) and final step (right).

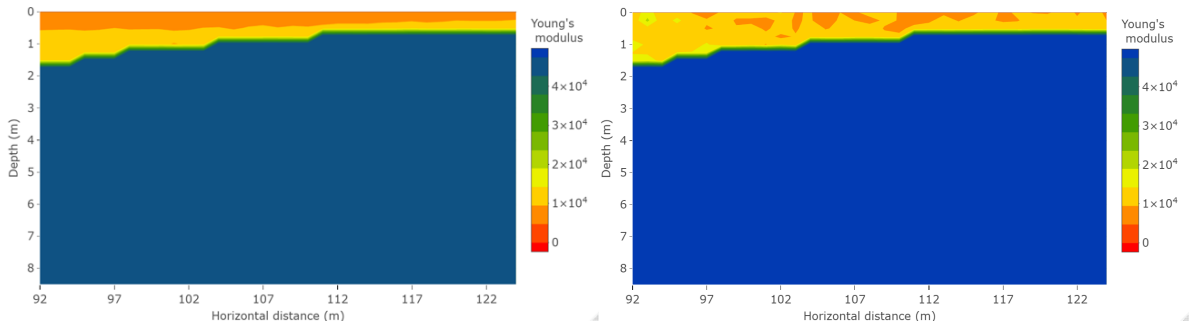

Figure 9. Young's moduli in 3rd part at initial step (left) and final step (right)

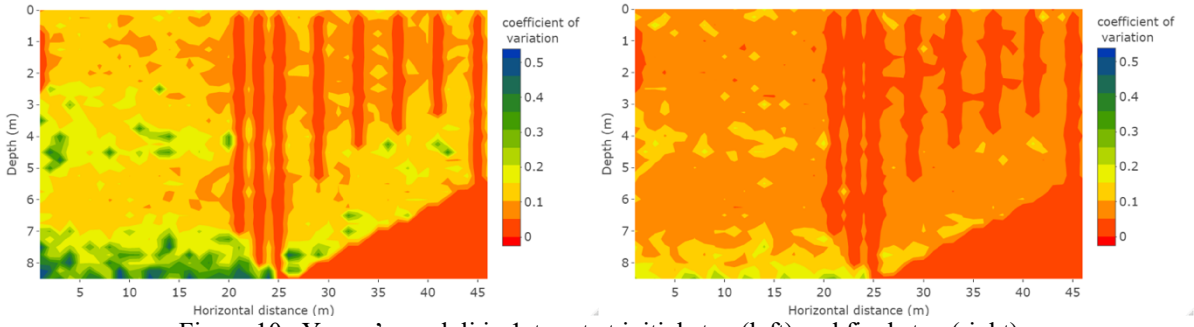

Figure 10. Young's moduli in 1st part at initial step (left) and final step (right)

to equal 2800 in presented simulations. The changes in *RSS* of the three parts are shown in Fig.6. The values of *RSS* for the three parts decrease quickly at the early stage of the assimilation.

Figs.7–9 show the spatial distributions of the Young's moduli updated by the EnKF in the first, second, and third parts, respectively. Only the Young's moduli at the initial and final time steps are shown in the figures. In the assimilation of the measured data, the spatial distribution is rapidly calibrated, and there is no longer a very significant difference between the spatial distributions at the 6th and final time steps. These changes in the spatial distributions of the Young's moduli correspond to the changes in the *RSS*. The large blue portions in Figs.7–9 represent the bedground. Not only the initial ensemble and the measurements, which are assimilated, but also prior information on the stratum, affect the estimated spatial distribution very strongly.

Fig.10 shows the spatial distribution of the coefficients of variation in the first part. It is actually the ensemble spread at the initial and final time steps. The spatial distribution of the coefficients of variation represents the uncertainties of Young's modulus. As such, the probability distribution of the Young's modulus on each grid cell is described by the ensemble spread and the ensemble mean. Aside from the bedrock, there are some places in Fig.10 where the coefficient of variation is 0. These places are actually the locations of the SWS tests, and the parameters on these corresponding grid cells are treated as true values when random fields are generated

using sGs. As for the other regions, it can be seen that the uncertainties decrease substantially after the assimilation of the EnKF.

7. Conclusions

(1) The proposed method employed the measurement data from SWM tests conducted on an earth-fill dam located in Okayama, Japan. As prior information, the data from the SWS tests were used to distinguish the bedground and the dam, and to introduce the spatial variability of the Young's moduli.

(2) The first arrival time of artificially excited surface waves was assimilated to estimate the Young's modulus of an earth-fill dam model. With updating, the *RSS* between the ensemble members and the actual measurements obviously decreased in most cases, and the spatial distribution of the Young's moduli could be identified.

(3) In the present study, it has been found that the EnKF is able to evaluate the uncertainties of the parameters. With the assimilation, the coefficient of variation was seen to have substantially decreased. The ability of the EnKF to reduce the uncertainties of the parameters has been verified.

Acknowledgements

This work was partly supported by JSPS KAKENHI Grant Number 20H00442.

References

Akita, T., Takaki, R., and Shima ,E. 2010. "Model parameter estimation by using the ensemble Kalman filter-application to nonlinear complex structure system with large deformation." Trans. JSCES. 2010: 20100021. https://doi.org/10.11421/jsces.2010.20100021

Annan, J., Hargreaves, J., Edwards, N., and Marsh, R. 2005. "Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. " Ocean Modell. 8(1-2):135-154. https://doi.org/10.1016/j.ocemod.2003.12.004

Caballero, E., Rochinha, FA., Borges, M., and Murad, MA. 2018. "An enhanced ensemble Kalman filter scheme incorporating model error in sequential coupling between flow and geomechanics." Int J Numer Anal Methods Geomech., 43(2), 482-500. https://doi.org/10.1002/nag.2872

Deutsch, C. V. and Journel, A. G. 1992. *Geostatistical Software Library and User's Guide*, Oxford University Press.

Evensen G. 2009. *Data Assimilation*. Springer-Verlag GmbH.

Evensen, G., Hove, J., Meisingset, H., Reiso, E., Seim, KS., and Espelid ø. 2007. "Using the EnKF for assisted history matching of a north sea reservoir model." Proceedings of SPE Reservoir Simulation Conference, SPE-106184-MS. https://doi.org/10.2118/106184-MS

Franssen, HJH. and Kinzelbach, W. 2008. "Real-time groundwater flow modeling with the ensemble Kalman filter: joint estimation of states and parameters and the filter inbreeding problem." Water Resour Res. 44(9). https://doi.org/10.1029/2007wr006505

Hayashi, K., Suzuki, H., and Sito, H. 2001. "Surface wave method using artificial sources." Development and application to civil engineering investigations. Oyo Tech Rep. 21, 9-39.

Japanese Geotechnical Society. 2015. J*GS Standards, Geotechnical and Geoenvironmental Investigation Methods*, JIS A 1221:2020. JGS.

Jongmans, D., and Demanet, D. 1993. "The importance of surface waves in vibration study and the use of Rayleigh waves for estimating the dynamic characteristics of soils." Eng Geol., 34(1-2): 105-113. https://doi.org/10.1016/0013-7952(93)90046-f

Katzfuss, M., Stroud, JR., and Wikle, CK. 2016, "Understanding the ensemble kalman filter." Am Stat. 70(4):, 350-357. https://doi.org/10.1080/00031305.2016.1141709

Liu, N., and Oliver, DS. 2005. "Ensemble Kalman filter for automatic history matching of geologic facies." J Petroleum Sci Eng. 47(3-4): 147-161. https://doi.org/10.1016/j.petrol.2005.03.006

Nishimura, S., Shibata, T., and Shuku, T. 2016. "Diagnosis of earth-fill dams by synthesized approach of sounding and surface wave method", *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 10(4), 312-319, DOI: 10.1080/17499518.2016.1197406

Pandurangan, V., Chen, Z., and Jeffrey, RG. 2015. "Mapping hydraulic fractures from tiltmeter data using the ensemble Kalman filter." Int J Numer Anal Methods Geomech., 40(4). 546-567. https://doi.org/10.1002/nag.2415

Tao, Y., Sun, H., and Cai, Y. 2020. "Predicting soil settlement with quantified uncertainties by using ensemble Kalman filtering." Eng Geol., 276, 105753. https://doi.org/10.1016/j.enggeo.2020.105753

Ueno, G., Higuchi, T., Kagimoto, T., and Hirose, N. 2007. "Application of the ensemble Kalman filter and smoother to a coupled atmosphere-ocean model." SOLA. 3: 5-8. https://doi.org/10.2151/sola.2007-002

Zerwer, A., Cascante, G., Hutchinson, J. 2002. "Parameter estimation in finite element simulations of Rayleigh waves." J Geotech Geoenviron Eng., 128(3), 250-261. https://doi.org/10.1061/(asce)1090-0241(2002)128:3(250)