MASTER'S THESIS OF NATURAL SCIENCE

Ensemble Kalman Filter for Transient Pumping Test Data

비정류 양수 시험에의 앙상블 칼만 필터

August 2022

Graduate School of Seoul National University

College of Natural Science

School of Earth and Environmental Sciences

Daeha Lee

Ensemble Kalman Filter for Transient Pumping Test

Advising Professor Kang-Kun Lee

Submitting a master's thesis of School of Earth and Environmental Science

August 2022

Graduate School of Seoul National University

College of Natural Science

School of Earth and Environmental Sciences

Daeha Lee

Confirming the master's thesis written by
Daeha Lee
August 2022

Chair _____(Seal)
Vice Chair _____(Seal)
Examiner _____(Seal)

ABSTRACT

As global climate change affects the hydrosphere, there is an increase in attention to management in groundwater usage. In general, numerical models are used to quantify and predict the effects of groundwater usage. Although hydraulic parameters determine groundwater flow, they have been poorly constrained due to difficulties in direct observation of underground conditions and heterogeneity. Based on the data assimilation with groundwater flow models, hydraulic parameters can be estimated and the method is expected to assist simulations more accurately. Here, EnKF (Ensemble Kalman filter) and NS-EnKF (Normal-Score Ensemble Kalman filter) are applied to the field experiment pumping test at Eumseong site to estimate the hydraulic parameters, such as a hydraulic conductivity, specific yield, and storage coefficient, based on the general parameter estimation tool, curve-fitting. Performances of filters in estimating hydraulic parameters are also tested before the application.

As no perturbation is applied to observation heads in the filter performance test, NS-EnKF displays higher performance on estimating three hydraulic parameters. Also with the smaller RMSE after the estimation compared to the initial parameters, both EnKF and NS-EnKF filters are identified to be valid for the parameter estimation.

Comparing both filters with expectations of hydraulic conductivity with the experiment data, EnKF displays higher performance, and the range of the estimated hydraulic conductivity is verified to be in the physical range. By analyzing and validating the application of both filters with pumping test experiments with curve-

fitting method to derive a precise distribution, more accurate groundwater flow modeling in reality is an expected benefit for this research.

Key words: hydraulic conductivity, specific yield, storage coefficient, pumping test, Ensemble Kalman Filter, Normal-Score Ensemble Kalman Filter

Student number : 2020-22099

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Research Background

As managing groundwater has become a paramount issue to deal with restrictive water resources, a groundwater model is a key factor in handling managements with appropriate prediction. Additionally, predictive groundwater models help groundwater users to understand possible risk in groundwater usage (Moore, 1979; Kumar and Kumar, 2014). However, inadequate modeling result or errors on modeling might mislead decisions. As a potential effect of exploitation is also evaluated with model prediction, misleading information might pose threats to sustainable water production (Sadat-Noori et al., 2020) and be irreversible. To prevent potential disadvantages from wrongheaded decisions, modeling requires a precise prediction. For adequate calculation and prediction with a model, as solving a groundwater flow equation is calculated based on hydraulic parameters, such as the storage coefficient, specific yield and hydraulic conductivity, these deterministic factors should be estimated precisely (Erdal and Cirpka, 2016).

Conventionally, a modeling process has been simplified by dividing sections assuming that each section is composed of homogeneous hydraulic conductivity (Shakoor et al., 2018; Yadav et al., 2020). This simplification has been considered according to the type of terrain or soil type (Kushwaha et al., 2009).

However, if the scale of the interested area becomes smaller, simplifying might neglect differences inside those homogeneous assumed sections and might result in a wrong prediction (Franssen et al., 1999).

Moreover, estimated groundwater parameters might be different from the true state with a curve-fitting method (Moench, 1994), so directly using curve-fitted parameters might cause errors on various numerical groundwater simulations.

Recent studies have focused on estimating hydraulic parameters from a distribution of groundwater level, as hydraulic parameters determine flow and transport models (Tong et al., 2012; Chen and Zhang, 2006). But there are some difficulties prevailing on the inverse modeling. One of the difficulties is that hydrogeological inverse modeling process is an ill-posed problem, which means that hydraulic parameters might not be uniquely decided. Furthermore, true state values are not directly measured, leading to the inaccessibility of validation (Singh et al., 2008), and even the estimation results are sensitive to the noises on measurements of the groundwater level (Yeh, 1986). For these reasons, for available validation to the performance of the methods and to handle ill-posed problems with known initial conditions, inversion methods have been implemented to the virtual models in general, assuming the ground truth distribution (Sun et al., 2009). Moreover, recent studies are mainly interested in estimating hydraulic conductivity and porosity, with the case of confined aquifer or unconfined aquifer with homogeneous specific yield and specific storage. Also some are based on stronger assumption in the estimation in unconfined aquifer cases that specific yield and specific storage are known (Chen and Zhang, 2006; Li et al., 2012b; Xu et al., 2013b; Zovi et al., 2017a; Tong et al., 2017). However, it is unnatural to get hydraulic head observations with no errors contained, which are the virtual cases. Also, specific yield and specific storage terms are also the soil property, which feature heterogeneous property. Considering both, this study processes the real pumping experiment data involving estimation sections of specific yield and storage coefficient after verification of an estimation performance with virtual cases.

Especially among extensive inverse modeling methods, Bayesian based Monte Carlo method, Ensemble Kalman Filter (EnKF) has been drawn an attention to improve modeling results, as it is effective to deal with a non-linearity of state

equation (Hendricks Franssen and Kinzelbach, 2008). To consider the non-Gaussian distribution, EnKF is further developed, and the normal score ensemble Kalman filter (NS-EnKF), which is an ensemble Kalman filter updating with normal score functions, is newly introduced (Li et al., 2012a; Zovi et al., 2017b). In sum, this study concentrates on handling hydraulic parameters, hydraulic conductivity, specific yield, and storage coefficient with EnKF and NS-EnKF for accurate groundwater simulations with a virtual model and a field experiment data of the pumping test.

1.2 Objectives and Scope

The main purpose of this study is to estimate hydraulic parameters including specific yield and storage coefficient of a numerical model that fit drawdowns from a pumping test and to reduce the error that occurs only by using curve-fitted hydraulic parameters. It is also anticipated to utilize estimated results of state parameters to simulate extended numerical simulations, such as groundwater flow simulations for mass transport and heat transport modeling. As most pumping tests in field sites have restrictions on time step, an estimation is based on the time restricted drawdown data.

To be in line with the results of recent studies to successfully estimate hydraulic conductivity, the EnKF is applied for an assimilation process. Iterative-EnKF is implemented for the updating process with the first head drop, and restart-EnKF is implemented for other time steps. This structure is called the combined EnKF filter in this study. Combined filter performance is tested with virtually constrained fields of parameters. Constrained parameters fields are randomly generated with the ordinary kriging (OK) method. After, the EnKF usage is verified with the forward simulation of mean estimated parameter result and the comparison of estimated field parameters with the constrained field parameters. Furthermore, normal score transformation is applied to the section of estimating hydraulic conductivity with the combined filter, and then compared with the estimation results of EnKF, so that both filter performances are confirmed.

Finally, after the validation, the combined filters of EnKF and NS-EnKF are applied to the pumping test field experiment data, and results are compared referring to other studies and with implied expectations in the drawdown data. Yielding reasonable distributions with ensemble assimilation methods is expected

to lead a way to access aquifer properties with better understanding of dynamics and interpreting convoluted field experiments based on pumping test experiments.

2 METHODOLOGY

2.1 Numerical Model

This study utilizes the finite difference method to calculate the groundwater flow equation which is a partial differential equation and obtain water level data per unit time. One of the most representative software, MODFLOW (Modular Three-Dimensional Finite-Difference Groundwater Flow Model) family, for solving the equation numerically (McDonald and Harbaugh, 2003; Harbaugh, 2005; Niswonger et al., 2011), is applied to simulate hydraulic heads in this study.

2.1.1 Computational Tool

The MODFLOW family software (McDonald and Harbaugh, 2003), developed by USGS (United States Geological Survey), is one of groundwater models interacted with surface water. Especially for this study, a transient flow with initial conditions of mixed boundary condition is the main assumption with the simulation of a pumping test. In other words, a constant head Dirichlet boundary and a no-flux Neumann boundary in both sides of a grid, are assumed for the head simulation. Although the program is written in fortran language, this program is assisted with FloPy (Bakker et al., 2016). FloPy is written in python language, and supports pre and post-processing of MODFLOW family groups.

Among various MODFLOW families, this study focuses on the utilization of MODFLOW-NWT (Niswonger et al., 2011), as it deals with smoothed head drops in unconfined aquifers of pressure changes and gravitational effects. In other words, for unconfined aquifers, MODFLOW-NWT provides a smooth transition with confined storage and zero storage (Niswonger, 2014). However, the difference with the calculation in an unconfined aquifer, which calculates in smoothed way, makes the drawdown to be smaller than the head drop calculated only with a storage

coefficient. However, in this way, the simulated drawdown is smaller than the measured drawdown for the experiment conducted in the Eumseong site. Referring to the unconfined aquifer calculation in MODFLOW-2000, which is, when the cell is saturated, an unconfined aquifer head change is calculated same way as a confined aquifer in MODFLOW-2000(Keating and Zyvoloski, 2009), the model is constructed to simulate first drawdown with hydraulic conductivity and storage coefficient, and then with hydraulic conductivity, storage coefficient, and specific yield.

Representatively for the MODFLOW tutorials with FloPy, confined steady-state flow model and unconfined transient flow model are accessible (https://flopy.readthedocs.io/en/3.3.5/tutorials.html#modflow-tutorials). The latter model is compatible with most drawdowns in the pumping tests in the interest area, Eumseong site, as it is an unconfined aquifer. A transient flow model at unconfined aquifer is implemented with the equation of the three-dimensional flow of groundwater of Darcy's law and a partial differential equation which represents hydraulic head distribution, derived from Darcy's law (Langevin et al., 2017). Darcy's mathematical equation describes a flux of the fluid, and for a three-dimensional version, a hydraulic conductivity is considered as a diagonal form with a principal axis:

$$\mathrm{q} = -\mathrm{K}\nabla\mathrm{h} = -\begin{pmatrix} K_{xx} & 0 & 0 \\ 0 & K_{yy} & 0 \\ 0 & 0 & K_{zz} \end{pmatrix}\nabla h \qquad \text{(Eq. 2.1)}$$

where q is a specific discharge of groundwater (L/T), K is a hydraulic conductivity vector (L/T), $\nabla\mathrm{h}$ is a potentiometric head gradient. An assumption is that the principal axes of hydraulic conductivity are parallel to the coordinate axes of x, y, z and describes that a specific discharge of the fluid occurs with a spatial potentiometric head gradient in the direct proportion with hydraulic

conductivity. In the perspective of a representative elementary volume, the Darcy's law leads to the three-dimensional groundwater governing equation as follows:

$$\frac{\partial}{\partial x}\left(T_{xx}\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(T_{yy}\frac{\partial h}{\partial y}\right)+\frac{\partial}{\partial z}\left(T_{zz}\frac{\partial h}{\partial z}\right)+\mathrm{Q}=\mathrm{S}\frac{\partial h}{\partial t} \qquad \text{(Eq. 2.2)}$$

where T is a transmissivity (L^2/T), S is a storage coefficient (dimensional less), Q is a recharge or discharge rate of groundwater inventory (L/T), and t is a time (T). Potentiometric head change with the derivative of time term, $\frac{\partial h}{\partial t}$ is additionally considered with a storage coefficient term, compared to equation 2.1, and also regards the incoming flow Q (L/T). The storage coefficient is considered as a combination of specific yield and specific storage with the thickness of aquifer under the unconfined aquifer, or a specific storage with the thickness of aquifer (=S_sb) for the confined aquifer (Fetter, 2018). A hydraulic conductivity, a storage coefficient for initial drawdown, and a specific yield are heterogeneous parameters, and in this study, these parameter ensembles are generated based on the ordinary kriging method. A detailed explanation is in the Section '2.1.3 Ordinary Kriging'. Although the groundwater flow equation represents the 3D model, a vertical flow is neglected and two-dimensional flow is only considered.

2.1.2 Mesh Construction

Considering to create a model with mixed boundary conditions, not to have critical issues of interpreting the pumping tests with numerical models, the mesh is constructed with a large void space, 5.8 *m* far from both ends of observation wells. In other words, the area of interest has a rare impact on the solution by the assumed boundary conditions (Camporese et al., 2012). The unit of the grid size consists of 0.2 m x 0.2 m, 99 x 99 numbers of cells. The number of layer is one and grid is considered as 2-dimensional flow in this study. The initial water table, constant head boundaries, and aquifer thickness are set as 20 *m* depth from the surface.

Wells are distributed same as the distribution of observation and pumping wells located at the Eumseong experiment site both for the performance test and the real data application. The name of observation and pumping wells is denoted as EM (Eumseong Monitoring). To be sure, EM-6 pumping well is located exactly in the middle of the mesh (Figure 2.2). Wells are in 2 m horizontal and vertical distance, except for EM-1 and EM-2, EM-2 and EM-3, EM-9 and EM-10. These wells are in 4 m horizontal distance.

The field experiment pumping test is conducted in 2020 and is denoted as ES20. The pumping well is located at EM-6 and other wells are used for observing drawdown. Figure 2.1 is the scheme for the ES20 pumping test.

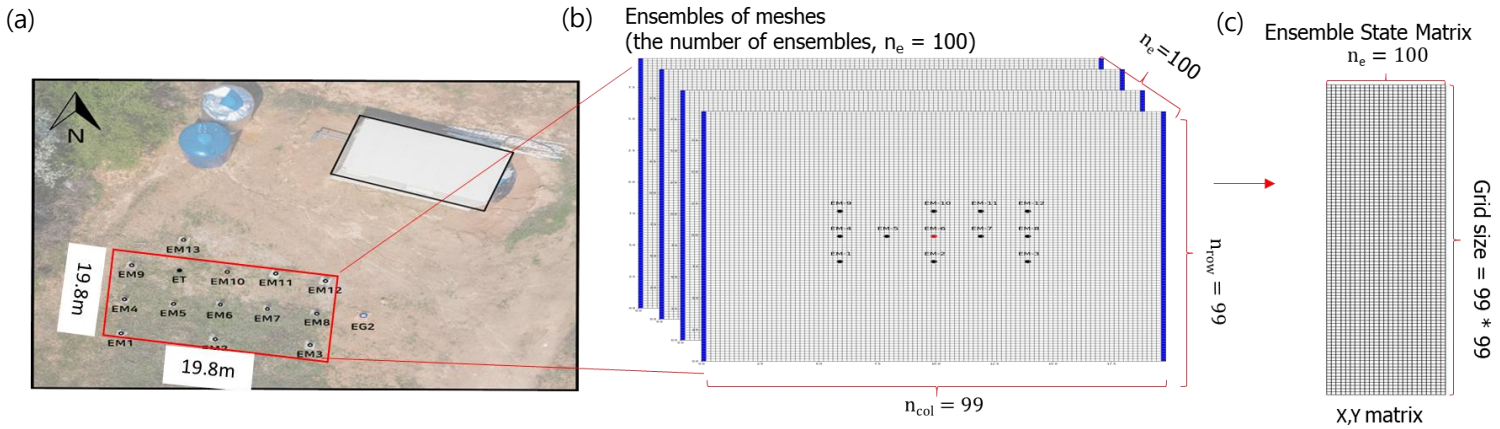

Figure 2. 1 (a) displays the overview of Eumseong site. (b) depicts the mesh construction to model the site. (c) depicts the matrix construction with ensembles members. This is the scheme for the ES20 pumping test model and a matrix construction.

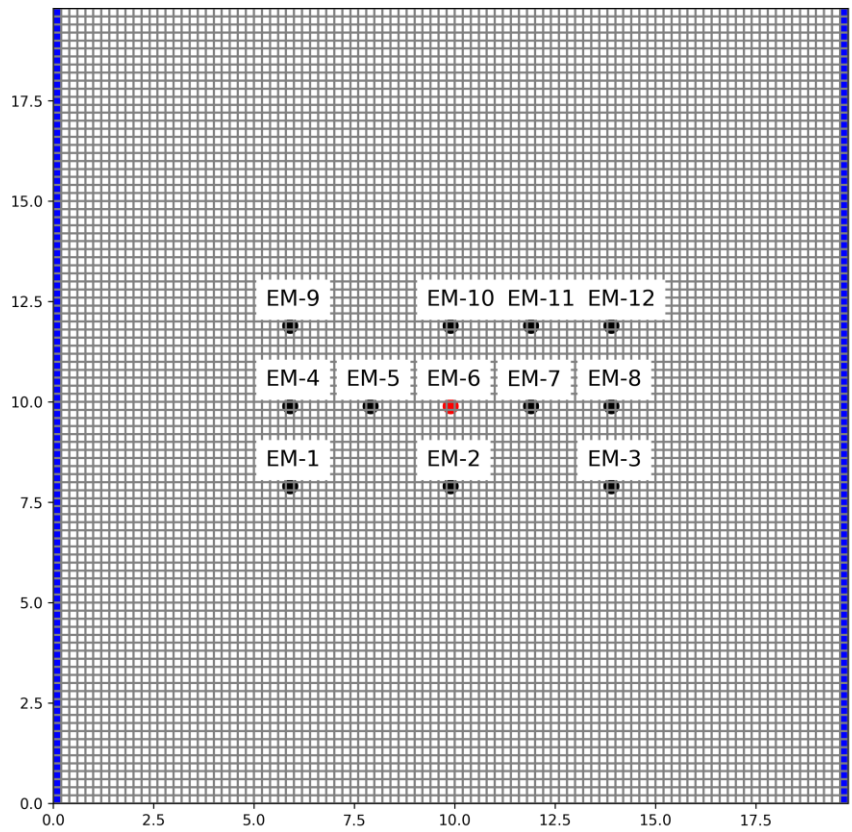

Figure 2. 2 The mesh for a numerical model. Red and black points indicate a pumping well and observation wells. Left and right side blue lines in mesh indicate the constant head boundary with no flux.

2.1.3 Ordinary Kriging

For the two dimensional simulation for the pumping test, hydraulic parameters should be given as grid values. With the field experiment pumping test data, hydraulic head measurement data are spatially sparse with few observation wells. Even with curve-fitted parameters in observation wells, kriging methods are needed for sparse hydraulic parameters to be converted into grid values. Especially, Ordinary Kriging (OK) method is used to compare the result of directly using curve-fitted parameters with a homogeneous assumption for numerical simulations and EnKF applied solutions for numerical simulations. An OK is as follows:

$$Z_{OK}^{*}(x_0) = \sum_{\alpha=1}^{n} w_\alpha Z(x_\alpha) \qquad \text{(Eq. 2.3)}$$

where Z_{OK}^{*} is expected value at neighboring points (x_a), and w_a is weights at sampling points (x_a), and Z(x) is value at sampling point.

OK constrains that the value at the sample point x_0 could be linearly constructed with values at neighboring points x_α with weights w_α (Wackernagel, 2003). Assuming that a random function Z(x) follows a variogram γ(h), to minimize the estimation variance with weight constraints, OK equations are written as a matrix form as follows:

$$\begin{pmatrix} \gamma(x_1 - x_1) & \cdots & \gamma(x_1 - x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n - x_1) & \cdots & \gamma(x_n - x_n) & 1 \end{pmatrix} \begin{pmatrix} w_1^{OK} \\ \vdots \\ w_n^{OK} \end{pmatrix} = \begin{pmatrix} \gamma(x_1 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \end{pmatrix} \qquad \text{(Eq. 2.4)}$$

To test the performance of the EnKF, virtual values of hydraulic conductivity, storage coefficient, and specific yield are randomly selected with the range referring to the standard deviations and mean values of the analytical Neuman curve-fitted parameter ranges. Then to convert random point values into field

values, the OK method is used and parameters are set to be the parameter distribution. Parameter ensembles are generated the same way.

For variogram parameters, to segregate the kriged values from the effect of boundaries, 5 m for the range value is set. Specifications are set as in table 2.1.

Table 2. 1 Variogram parameters that are used in this study for an ordinary kriging method.

Variogram Parameters	Nugget	Range	Sill	Variogram Function
	0	5	1	Gaussian

2.2 Kalman Filter

2.2.1 Linear quadratic estimation (LQE)

Linear quadratic estimation, also called as the Kalman filter, is the estimator for the full-state measurements dealing with dynamics containing the restricted observed measurements as it might be improbable to observe the full-state. Mathematically it is possible to estimate if the observability Gramian guarantees (Brunton and Kutz, 2019). State-space equations in linear form are derived from the nonlinear input-output system equations approximated by the Taylor expansion:

$$\frac{d}{dt}x = f(x,u) \qquad \text{(Eq. 2.3)}$$

$$y = g(x,u) \qquad \text{(Eq. 2.4)}$$

where a full-state term is notated as x, an observed state is notated as y, and a control protocol is notated as u.

For differences in x and u, notated in Δx, Δu, function f and g are expanded with Taylor series in the point of $(\bar{x},\bar{u})$ where $f(\bar{x},\bar{u}) = 0$. For the function g, it is done similarly as for the case of function f:

$$f(\bar{x}+\Delta x,\bar{u}+\Delta u) = f(\bar{x},\bar{u}) + \left.\frac{df}{dx}\right|_{(\bar{x},\bar{u})}\Delta x + \left.\frac{df}{du}\right|_{(\bar{x},\bar{u})}\Delta u + \cdots \qquad \text{(Eq. 2.5)}$$

$$g(\bar{x}+\Delta x,\bar{u}+\Delta u) = g(\bar{x},\bar{u}) + \left.\frac{dg}{dx}\right|_{(\bar{x},\bar{u})}\Delta x + \left.\frac{dg}{du}\right|_{(\bar{x},\bar{u})}\Delta u + \cdots \qquad \text{(Eq. 2.6)}$$

By ignoring high order terms and being included into noise terms representing a disturbance in the system, noted as w_d, and a noise in the measurement, noted as w_n, linearized dynamics are stated with matrices of A, B, C, D:

$$\frac{d}{dt}x = Ax + Bu + w_d \quad \text{(Eq. 2.7)}$$

$$y = Cx + Du + w_n \quad \text{(Eq. 2.8)}$$

where disturbance and noise terms are assumed to follow a zero-mean Gaussian distribution. To simplify the equations, the control protocol is considered as having a linear relationship with a full-state vector. Simplified renewal discrete version of the state space model is as the following equation (Sun et al., 2016):

$$x_{k+1} = M_k x_k + \eta_k \quad \text{(Eq. 2.9)}$$

$$y_k = H_k x_k + \epsilon_k \quad \text{(Eq. 2.10)}$$

where an observation y_k is considered as linearized function of full-state x_k, so that H_k is often called as a linear matrix. M_k is a model matrix that calculates *a posteriori* full-state matrix from *a priori* full-state matrix. Similar to the continuous state-space equations, η_k is an error that a model contains compared to the real full-state changes and ϵ_k is an error of a measurement.

A model propagation in general is with the cost function minimizing an expectation value of the difference between full-state vector and its estimate, *a posteriori* estimation is described with *a priori* estimation and a Kalman gain (Drécourt, 2003):

$$x_k^a = x_k^f + K_k(y_k - H_k x_k^f) \quad \text{(Eq. 2.11)}$$

$$K_k = P_k^f H_k^T [R_k + H_k P_k^f H_k^T]^{-1} \quad \text{(Eq. 2.12)}$$

$$P_k^f = M_{k-1} P_{k-1}^a M_{k-1}^T + Q_{k-1} \quad \text{(Eq. 2.13)}$$

$$P_k^a = (I - K_k H_k) P_k^f \quad \text{(Eq. 2.14)}$$

where upper notation a is for values of *a posteriori* estimation and f is for *a priori* estimation. Also the lower indicates the step number. Specifically, P_k is an error covariance matrix of an estimation of a full state and is updated in each step with equation 2.13. R_k is an error covariance matrix of measurement which is in the state space equation, and is noted as ϵ_k. Q_k is an error covariance matrix of a model, which is in the state space equation and noted as η_k. Both R_k and Q_k are fixed constant through iterations. Including disturbance and noise, the cost is from an ensemble average over various simulations (Brunton and Kutz, 2019). The scheme is drawn to help understand what variables are propagated in Figure 2.3, and these steps are briefly described in Figure 2.4. The cost function equation is as follows:

$$J = \lim_{k \to \infty} \mathrm{E}(x_k - \widehat{x_k})^* (x_k - \widehat{x_k}) \quad \text{(Eq. 2.15)}$$

where a hat notation denotes the estimated term and a superscript of star (*) denotes complex conjugate to be the optimal full-state estimator.

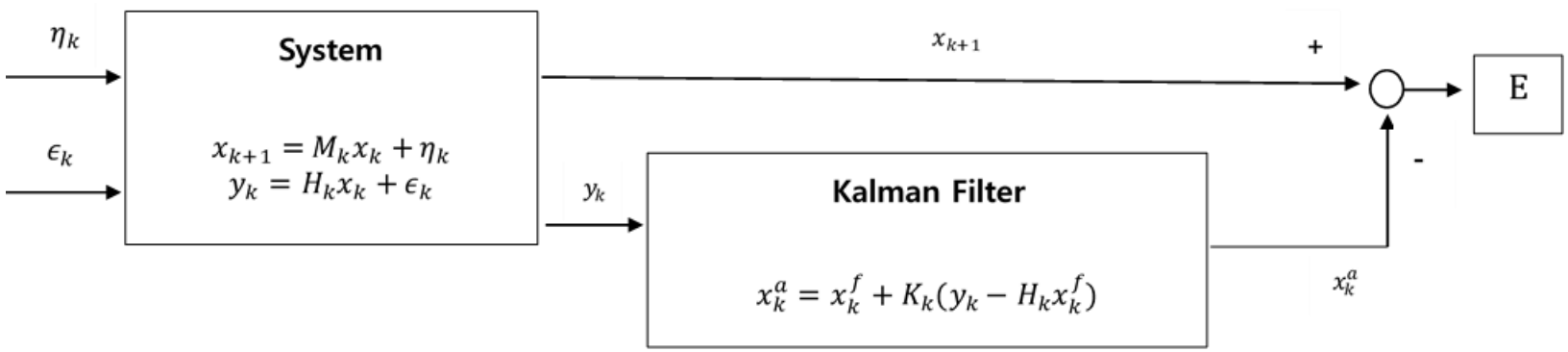

Figure 2. 3 Schematic figure of the Kalman filter for full-state estimation with a simplified system. $\mathbf{E}$ in the figure means expected value for an estimation.

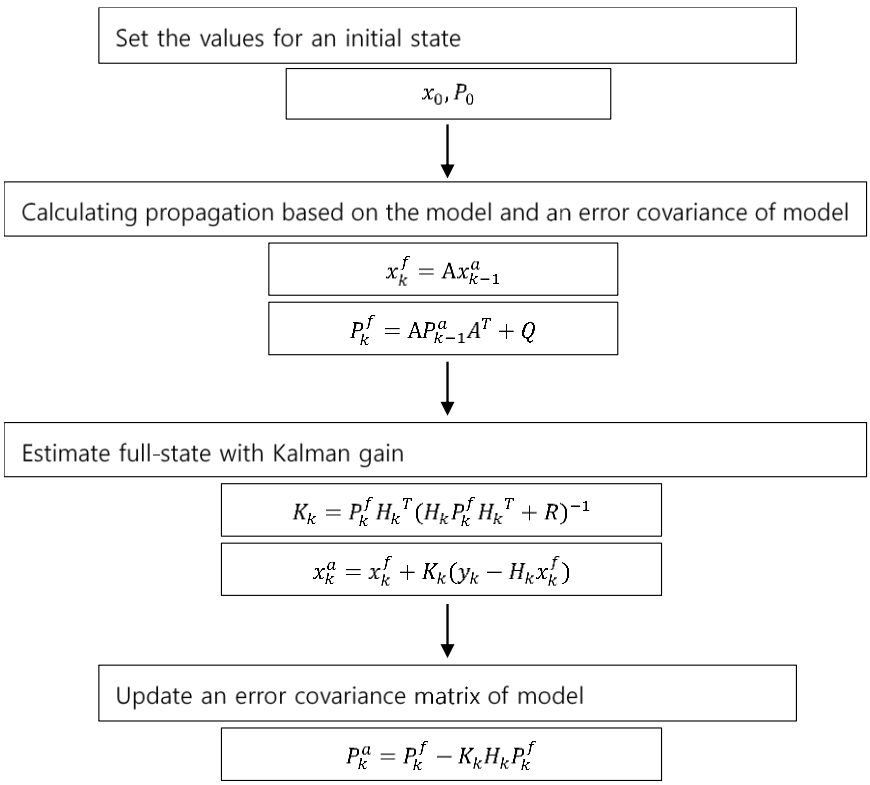

Figure 2. 4 Flowchart of the Kalman Filter. Except for the step setting an initial state, all the steps proceed recurrently.

2.2.2 Ensemble Kalman Filter

The Ensemble Kalman Filter (EnKF) is an estimator with Monte Carlo based prediction of error statistics (Gillijns et al., 2006). In other words, with a guess of an observation, new observations are defined with perturbations that are made up with the Gaussian distribution of a variance with an observation covariance matrix and an ensemble perturbation (Burgers et al., 1998).

In a first step, a state variable matrix is adjointly constructed with measurements and parameters that we want to estimate. For example, for pumping test models, a state variable matrix contains values of hydraulic conductivities and hydraulic heads basically (Li et al., 2012c).

After the construction of a state variable matrix, ensembles are generated with an initial state variable matrix. These ensembles form multi-Gaussian distribution as the EnKF method assumes that variables follow multi-Gaussian distribution in the perspective of ensembles and grids.

For one step ahead in time scale, *posteriori* values are made up from a calculation of numerical modeling in the next step. With new values, state variables are updated and substituted into predicted values, and with the Kalman Filter algorithm, full state estimation is processed.

In this study, unconfined transient pumping test data are used for assimilation. Not only hydraulic conductivities, but also storage coefficients for the first drawdown and specific yields have to be considered. To design assimilation steps, previous study results are considered. EnKF and Normal-Scored EnKF (NS-EnKF) for non-gaussian field are known to be effective tools for estimating hydraulic conductivity with head measurements containing steady states in piezometric head (Zhou et al., 2012; Xu et al., 2013a). Also, steady states for unconfined aquifers, with the information of pumping rates, distances from well, and drawdowns,

Dupuit-Thiem equation can be used to figure out aquifer hydraulic conductivities (Fetter, 2018). If the time step is far from the starting point of the pumping test, then it becomes closer to the steady state, so the assimilation steps are divided into three sections. In the first section with sequential order, with estimated hydraulic conductivities, storage coefficients for the first drawdown are estimated with hydraulic head measurements, and this section is only for the first time step data. In the next section, with the head drops from the results of forward simulation with estimated storage coefficients, specific yields are estimated with head measurements. At the third section, as the third section is the furthest section from the start, with the head drops from the first section, hydraulic conductivities are estimated with head measurements. If we refer to the Dupuit-Thiem equation, only hydraulic conductivity values set spatial distribution of drawdowns, so with the constant field of specific yield and storage coefficient from the mean value of curve-fitted parameters from the pumping test in the Eumseong site, hydraulic conductivity is estimated before these three sections are processed with third section data. This step of estimating hydraulic conductivity is called the conditioning section. To sum up, after the conditioning step for initial estimation for hydraulic conductivities, storage coefficients and specific yields are estimated sequentially, and with estimated specific yields and storage coefficients, hydraulic conductivities are estimated again. For the assurance of the estimation steps with varied parameters, after conditioning and processing three sections, storage coefficients, specific yields, and hydraulic conductivities are estimated again by processing assimilation in three sections again. Total seven sections are constructed for the estimation process.

To solve the problem of only using one time steps for assimilating storage coefficients, the first time step is assimilated as an iterative way. In (Gu and Oliver, 2007), iteration is ended if convergence criteria is satisfied, but in this study, the

first time step is iterated for 30 times, not considering a convergence threshold and to evenly dispense estimation steps.

Also, the known information of the numerical model is a constant head distribution before the pumping starts. Thus the conditioning step needs to restart the simulation from the constant head distribution (Chen et al., 2018). In other words, a restart EnKF method is used for conditioning. Furthermore, in second and third sections of the assimilation, as a pumping model runs with the first drawdown with storage coefficient and hydraulic conductivity, restart EnKF is implemented with the first drawdowns.

Hydraulic conductivity, storage coefficient and specific yield are assumed to be in log normal Gaussian distribution, so the values are converted into the logarithm value with base 10 (Freeze, 1975).

For the error matrix R, an error standard deviation is needed. The Mini-Diver, which is used for measuring pressure in observation wells, for the water column measurement range in 20 m has typical accuracy of $\pm$ 1 cmH_2O. So the standard deviation value is set to be 0.01 m for all three sections. For the conditioning process, considering the three sigma rule and the error of the Baro-Diver, which is used to measure atmospheric pressure, the standard deviation is set to be 0.04 m. This is a slightly higher value compared to the consideration of water level drop during the pumping test, but as the section is only for the rough constriction of initial hydraulic conductivity distributions of ensembles, not to constrain too much with invalid parameters, the value is set.

RMSE is calculated after the estimation process is done. The equation for calculating RMSE is as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

where n is a number of elements for an error calculation, y_i is an estimated or simulated value, and $\hat{y}_i$ is an observation or set value. Results and filter performance are evaluated based on RMSE in this study.

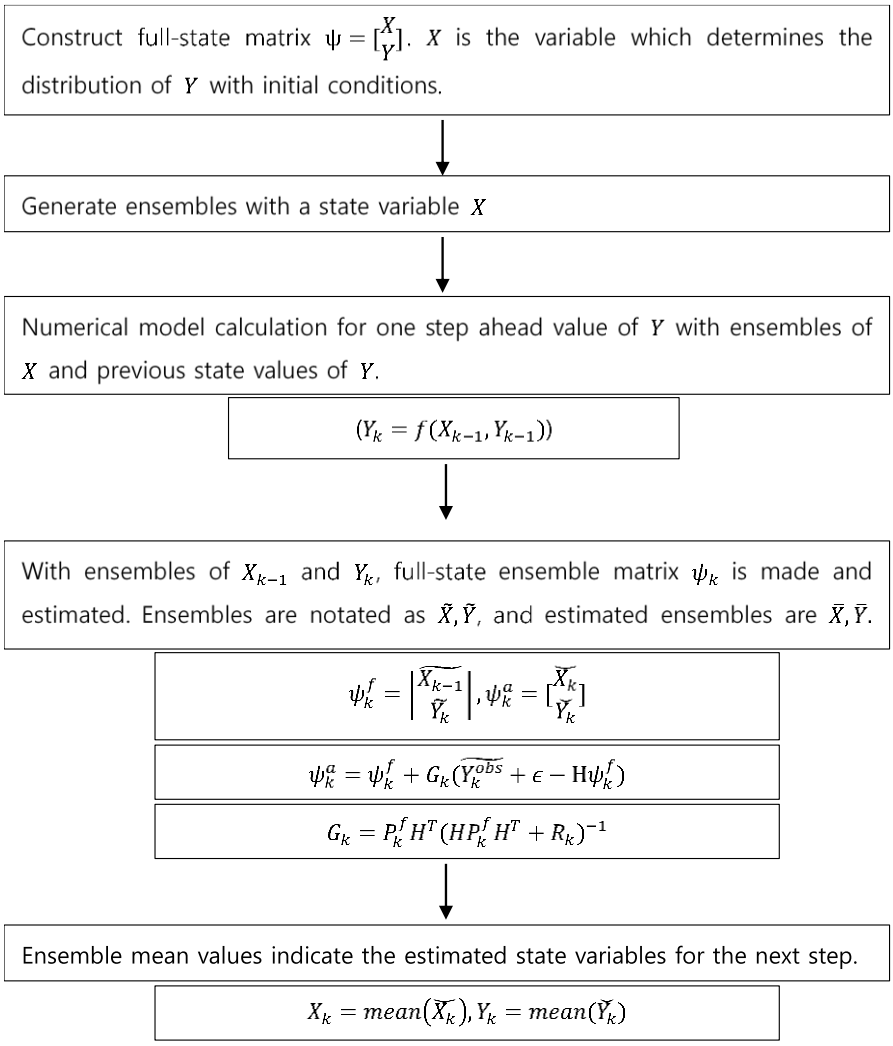

Figure 2. 5 The flowchart of the Ensemble Kalman Filter. The procedure recurrently proceeds.

2.2.3 Normal-Score Ensemble Kalman Filter

Previously introduced filter, the EnKF, is manipulated when the full-state ensemble matrices are generated following the form of a Gaussian distribution. The only difference between the EnKF and the NS-EnKF is that the NS-EnKF introduces a function called "normal-score transformation function" to make ensembles of state variables X and Y form a standard normal Gaussian distribution (Zhou et al., 2011). Two normal-score functions noted as ϕ_1 and ϕ_2 are introduced. With ensemble values, normal-score transformation functions are constructed for all components with local cumulative distribution functions (cdf). For the directly measured state variable *Y*, the function is applied before the updating steps for estimating hydraulic conductivity.

Before the updating process of NS-EnKF, the normal score transformation function is applied to *a priori* forecast matrix to make the state variables of ensembles form the standard normal distribution. After the updating process, to create *a posteriori* assimilation matrix, an inverse transformation procedure of the applied normal score transformation function is conducted. This normal score transformation is based on the cumulative distribution function (cdf) of the fitted ensemble variables with fixed quantile numbers. As this transformation is based on the cdf, it is not applicable for the values that are out of the range of fitted values to process transformations or inverse transformations. Values that are out of the range are transformed into maximum and minimum values, 5.199 and -5.199. In this study, all the state variables of *priori* forecast measurements are used to create the transformation function, so this is not the problem for the procedure. But before the inverse transformation, non-inversed *posteriori* assimilation matrix might contain the values that are over the maximum and minimum values, 5.199 and -5.199. These values are transformed into each end value of the fitted values, so the

normal score transformation applied state variables are bounded into the values that are from the ensembles that are generated before the updating process.

Priori full-state ensemble matrices on each step are transformed and then updated. Then *posteriori* full-state ensemble matrices on each step are calculated with Kalman gain same as in the EnKF. From the estimated, *posteriori* full-state ensemble matrices, after back transformation, the progress proceeds to the next level. It is also done recurrently.

In this study, to preserve prior model structure, normal-score functions for hydraulic parameters are not used, and only used for converting hydraulic heads before the updating process in the estimation sections of hydraulic conductivity distribution (Li et al., 2012a). In other words, $\phi_1(X) = X$ in equation 2.15. Moreover, recent studies are done for estimating hydraulic conductivity based on the piezometric head data (Zhou et al., 2012; Li et al., 2012a). Therefore, the normal-score transformation is applied only on estimating hydraulic conductivities in the third section. Equations are as follows:

$$\overline{\overline{\psi_k^f}} = \begin{bmatrix} \phi_1\left(\widetilde{X_{k-1}}\right) \\ \phi_2\left(\widetilde{Y_k}\right) \end{bmatrix}, \overline{\overline{\psi_k^a}} = \begin{bmatrix} \phi_1(\widetilde{X_k}) \\ \phi_2(\widetilde{Y_k}) \end{bmatrix} \quad \text{(Eq. 2.15)}$$

$$\overline{\overline{\psi_k^a}} = \overline{\overline{\psi_k^f}} + \overline{\overline{G_k}}(\phi_2(\widetilde{Y_k^{obs}}) + \epsilon - H\overline{\overline{\psi_k^f}}) \quad \text{(Eq. 2.16)}$$

$$\overline{\overline{G_k}} = \overline{\overline{P_k^f}} H^T (H\overline{\overline{P_k^f}} H^T + R_k)^{-1} \quad \text{(Eq. 2.17)}$$

$$X_k = mean\left(\widetilde{X_k}\right), Y_k = mean(\widetilde{Y_k}) \quad \text{(Eq. 2.18)}$$

where $\overline{\overline{P^f}}$ denotes a covariance matrix of ψ^f computed by a quantity of $(\psi^f - \overline{\psi^f})$ with $\overline{\psi^f}$, indicating the matrix with repeated vectors containing information of mean values of ensemble for all grid points.

2.3 Eumseong Data

2.3.1 Pumping experiment

In line with the purpose of this study to apply and estimate for the real pumping test data in the Eumseong site, pumping test data are acquired from the interest area located in Mountain 6-3, Naesan-ri, Daeso-myeon, Eumseong-gun, Chungcheongbuk-do, Republic of Korea (Figure 2.4). In the location, wells are labelled with EG, EM, and ET. There are only two EG wells and one ET well, and are not our interest as all the sensors measuring hydraulic heads are in EM wells. EM wells are fully penetrating wells with a radius of an inch.

The pumping test was conducted at March 17th, 2020 for 610 minutes. The pumping well is located at EM-6 with the pumping rate of 16.56 $\boldsymbol{m^3/day}$. Other 11 wells, from EM-1 to EM-12 are used to observe drawdown data. To eliminate noises occurred with the direct well-effect, observed data at the pumping well are not used. Pressures are measured with the Mini-Divers from Van Essen Instruments, which are installed at the 10 *m* depth from the water table. The pressure is measured every 10 minute during the pumping test. Pressures that Mini-Divers measured are in the unit of $\boldsymbol{cmH_2O}$, and are corrected with the Baro-Divers which measure the atmospheric pressure in the unit of psi, installed at EM-6 and EM-10. All the units are converted into $\boldsymbol{mH_2O}$ assuming the water density to be 1000 $\boldsymbol{kg/m^3}$. A water level is finally calculated by subtracting the drawdown from 20 *m*, assuming that the initial water level in the study area is 20 *m*.

2.3.2 Curve-fitting method with Neuman solution

With the pumping experiment, Neuman solution (Neuman, 1974, 1975), which is estimating hydraulic parameters with a curve-fitting method, is implemented using the software, AQTESOLV. Estimated parameters are a hydraulic conductivity, a storage coefficient for the water pressure changes which is governing the initial drawdown, and a specific yield for the gravitational head drops assuming the homogeneity of the aquifer. However, with the homogeneity assumption, the estimated parameters should be converted into field parameters with mean values, or extrapolation. In this study, to convert the estimated value into the field parameters, mean values for the estimated parameters in each well is calculated to generate a homogeneous aquifer model. With homogeneous field parameters, the numerical model is simulated. First drawdown is simulated based on the mean value of hydraulic conductivity and storage coefficient, and with the first drawdown, afterwards are simulated with mean value of specific yield, storage coefficient, and hydraulic conductivity. In other words, with the homogeneous parameters, the combined model is simulated. Then, the result is compared with the estimated results using EnKF and NS-EnKF.

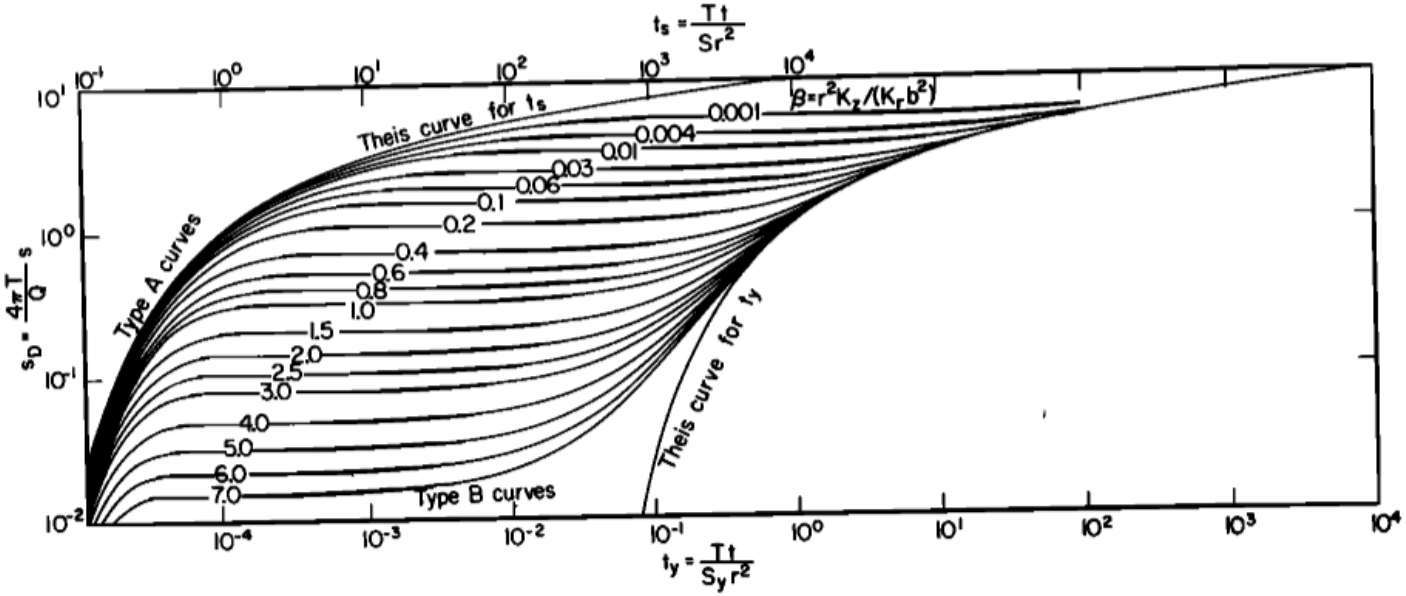

Figure 2. 6 Neuman curves for analyzing pumping test drawdown data in fully penetrating wells in an unconfined aquifer from Neuman (1975).

3 RESULTS AND DISCUSSION

3.1 Curve-fitted parameters

With the pumping test drawdown data, curve-fitting method is applied to estimate parameters roughly. Water table for the initial condition is set as 20 *m*, which is same as the setting of the numerical model, and the well radius, an inch, is used for the input information before curve-fitting. The result of the estimated parameters in each observation well is in following Table 3.1. A hydraulic conductivity is calculated from the transmissivity, divided with 20 *m* water level. With the parameters converted into the logarithm with base of 10, hydraulic conductivity shows quite small standard deviation value of 0.035 *m*. Not even a difference occurring in an order difference implies that the area of interest is quite homogeneous, and is the reason for smoothing out with the method of ordinary kriging. The logarithmic standard deviations of specific yield and storage coefficient are 0.223 *m* and 0.192 *m*, which of values are higher than those of hydraulic conductivity. These standard deviation values and mean values are the base of generating ensembles afterwards.

Table 3. 1 Curve-fitted parameters estimated with Neuman solutions. T, K, Sy, S denote transmissivity, hydraulic conductivity, specific yield, and storage coefficient. All the units are in meter and day.

ES20	Neuman curve-fitted parameters						
Well	T	K	Sy	S	log K	log Sy	log S
EM-1	22.120	1.106	0.055	0.005	0.044	-1.263	-2.291
EM-2	22.260	1.113	0.118	0.014	0.046	-0.927	-1.867
EM-3	20.920	1.046	0.129	0.016	0.020	-0.891	-1.791
EM-4	23.360	1.168	0.065	0.006	0.067	-1.187	-2.254
EM-5	19.480	0.974	0.206	0.015	-0.011	-0.687	-1.837
EM-7	21.400	1.070	0.229	0.014	0.029	-0.639	-1.859
EM-8	22.910	1.146	0.164	0.012	0.059	-0.785	-1.934
EM-9	23.160	1.158	0.062	0.005	0.064	-1.210	-2.298
EM-10	21.710	1.086	0.148	0.009	0.036	-0.831	-2.045
EM-11	26.400	1.320	0.062	0.007	0.121	-1.209	-2.138
EM-12	24.500	1.225	0.110	0.008	0.088	-0.961	-2.118
Mean	22.565	1.128	0.122	0.010	0.051	-0.963	-2.039
Std	1.848	0.092	0.060	0.004	0.035	0.223	0.192

3.2 Evaluation of Model Performance

For the application of EnKF and NS-EnKF to the field experiment, there is no ground-truth for a hydraulic conductivity that could be directly measured. So, before the method application to the pumping test data of ES20, a verification of the method for estimating parameters in the virtual model is needed.

To evaluate a performance of the EnKF, a random field with logarithm values of hydraulic conductivity in 11 observation points satisfying Gaussian distribution referring to the mean and standard deviation from the estimated parameters of Neuman curve-fitting were generated and then extrapolated with OK. This hydraulic conductivity distribution is assumed to be the ground-truth data, which is called the virtual field answer in this study. The ground-truth distribution of specific yield and storage coefficients are also generated the same way. With known hydraulic parameters and initial boundary conditions, hydraulic head data are simulated in specific points of time by the combined forward model. These head data are presumed to be the observation measurements. In other words, discrete time series data are extracted in the locations of observation EM wells. In the next step, ensembles are generated randomly as same as generating the virtual field, and then estimated hydraulic parameters are updated by ensemble filters, EnKF and NS-EnKF. After the end of the process, estimated hydraulic parameters are compared to the virtual hydraulic parameters to check whether the mean estimated hydraulic parameters are similar to the virtual hydraulic parameters. If they are similar, then the filter performance for estimating hydraulic parameters is verified.

The error standard deviations of measurement for EnKF and NS-EnKF are set to be the same as the measurement error standard deviation values from the Eumseong pumping site of Mini-Diver 0.01 m. For the conditioning part, error

standard deviation values are set as 0.04 *m*, not to estimate too precisely with wrong specific yield and storage coefficient values. For the NS-EnKF method, as hydraulic heads of ensembles are converted into the standard normal distribution, for the parts that use normal score transformation functions, this error measurement standard deviation values are set as 1 *m*. For both methods, ensemble numbers are one hundred.

With the performance tests, NS-EnKF shows higher performance compared to the results of EnKF with RMSE of hydraulic parameters as written in Table 3. 2. RMSE for the well points and field distribution are calculated, and for every parameter, using a normal-score transformation function in the updating process of a hydraulic conductivity distribution reduces the error. Figure 3. 1 plots the distribution of hydraulic parameters, which are initial, estimated, and answer sequentially. Comparing the estimated distribution with initial distribution, high and low trends are clearly captured in specific yield and storage coefficient parameters. Only with the plot, (b) seems to be slightly different with (c) below EM-1 and EM-3, above EM-9 and EM-10 in Figure 3. 1, estimated hydraulic conductivity distribution RMSE is calculated to be clearly smaller than the initial distribution in Table 3. 2. Also with normal-score transformation, an estimation of hydraulic conductivity is advanced in plot (b) of Figure 3. 4. In other words, both methods reduce the RMSE errors from the initial distribution of hydraulic parameters. Furthermore, the simulated drawdowns with the mean of estimated parameters show negligible difference with the simulated drawdowns with the virtual field in Figure 3. 2 and 3. 4. Error reduction and accurate drawdown simulation imply both methods to be valid for estimating hydraulic parameters, especially hydraulic conductivity, storage coefficient, and specific yield.

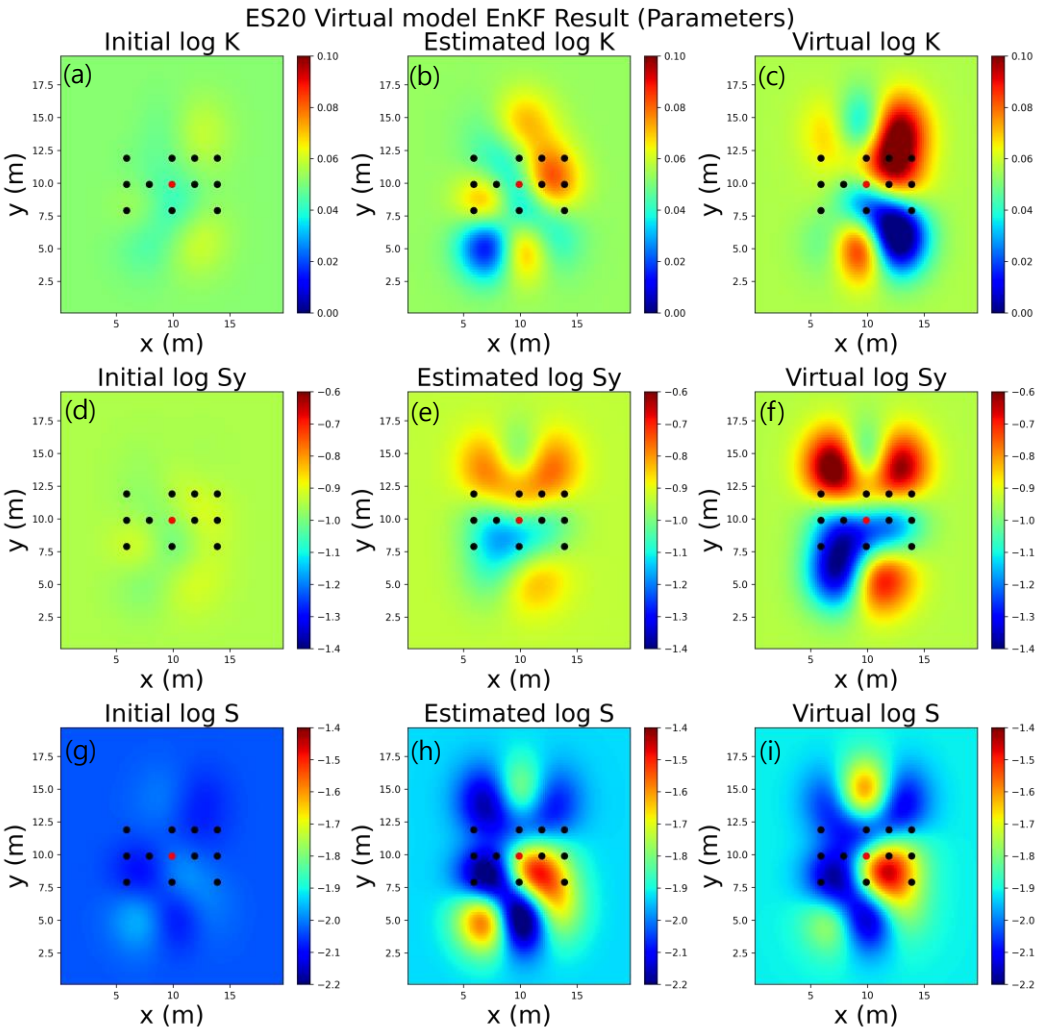

Figure 3. 1 Application of EnKF to ES20 virtual model and estimated parameters. Each row indicates the mean value of hydraulic conductivity (K), specific yield (Sy), and storage coefficient (S). Left columns (a, d, g), middle (b, e, h), and right (c, f, i) columns display initial parameters for EnKF, estimated parameters, and correct answer generated with ordinary kriging.

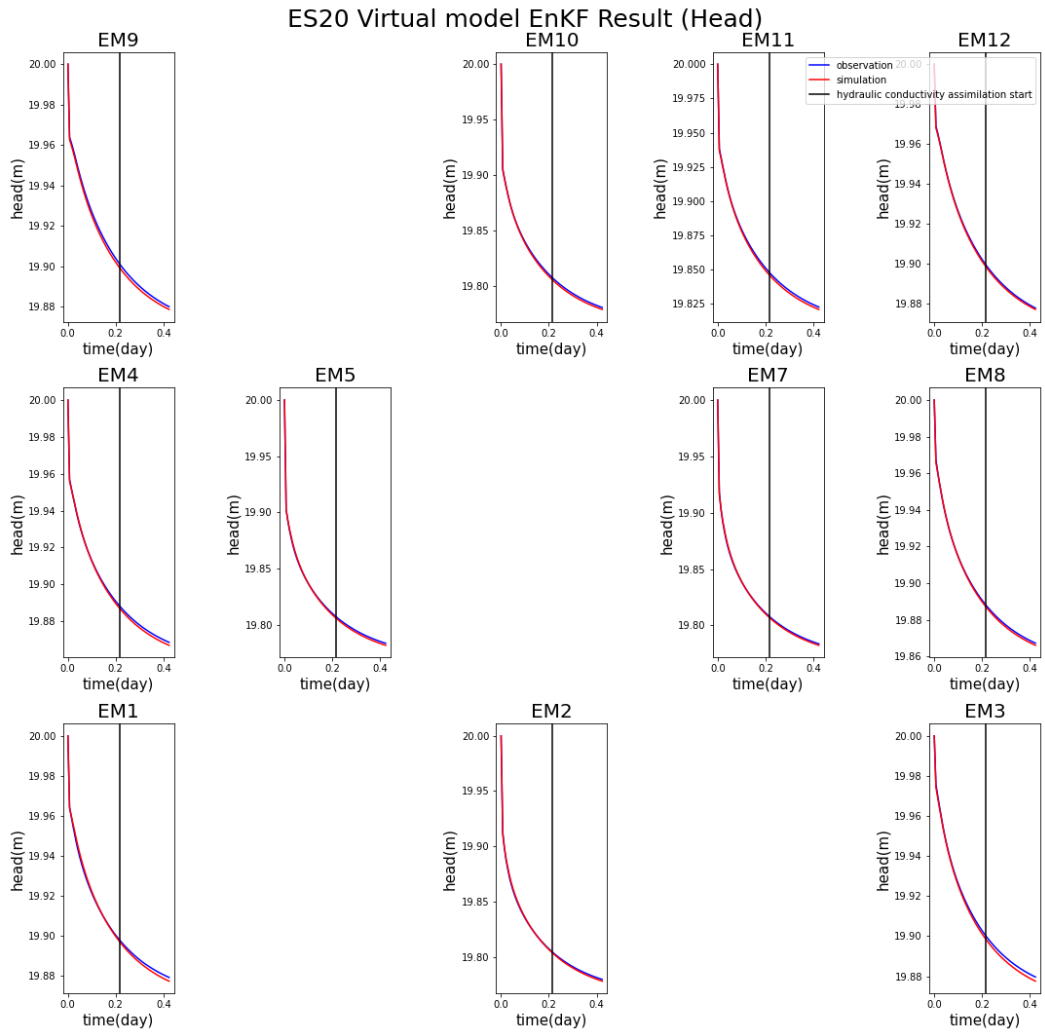

Figure 3. 2 ES20 virtual model forward modeled heads with the mean values of estimated ensembles of EnKF. Blue lines indicate the correct answer considered as the observation and red lines indicate forward modeling results. Vertical black lines in the middle of the plots indicate a starting point for conditioning.

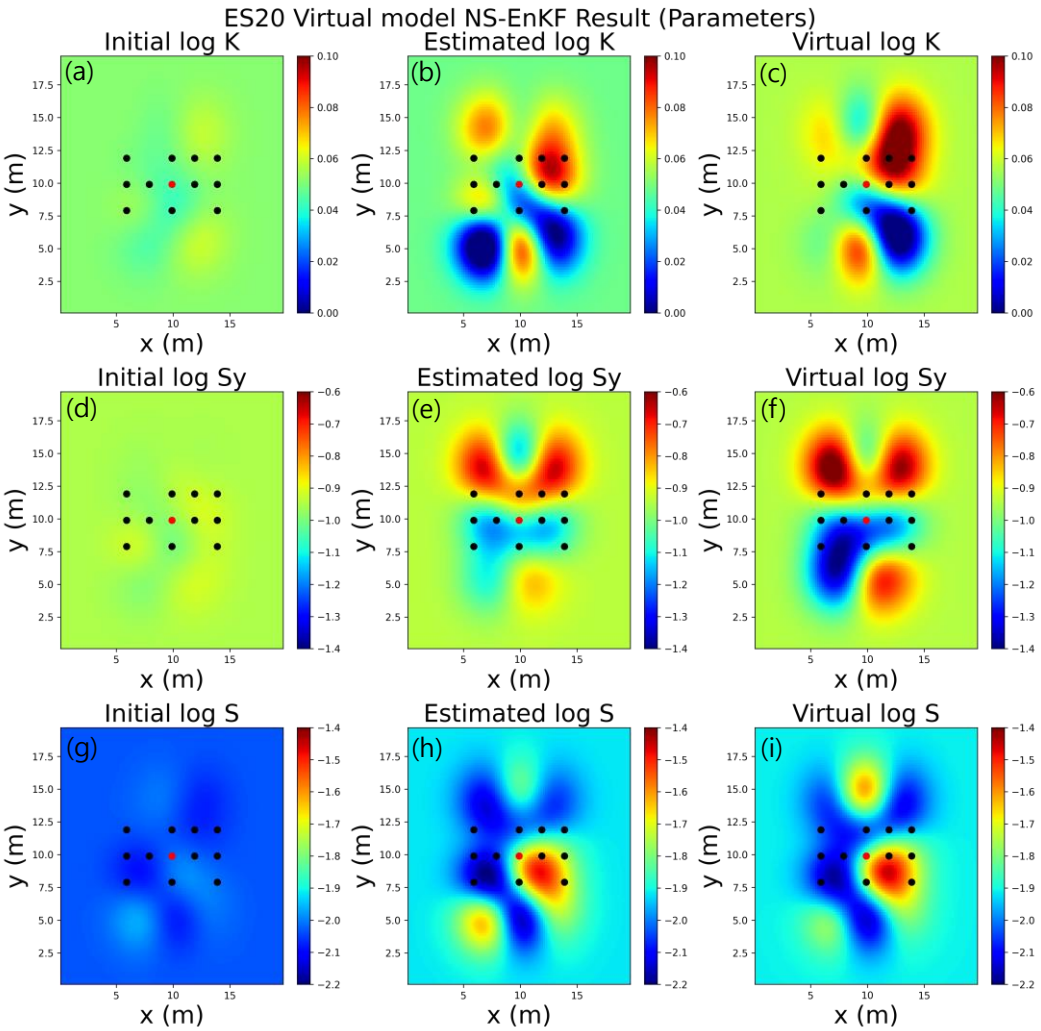

Figure 3. 3 Application of NS-EnKF to ES20 virtual model and estimated parameters. Each row indicates the mean value of hydraulic conductivity (K), specific yield(Sy), and storage coefficient (S). Left columns (a, d, g), middle (b, e, h), and right (c, f, i) columns display initial parameters for EnKF, estimated parameters, and correct answer generated with ordinary kriging.

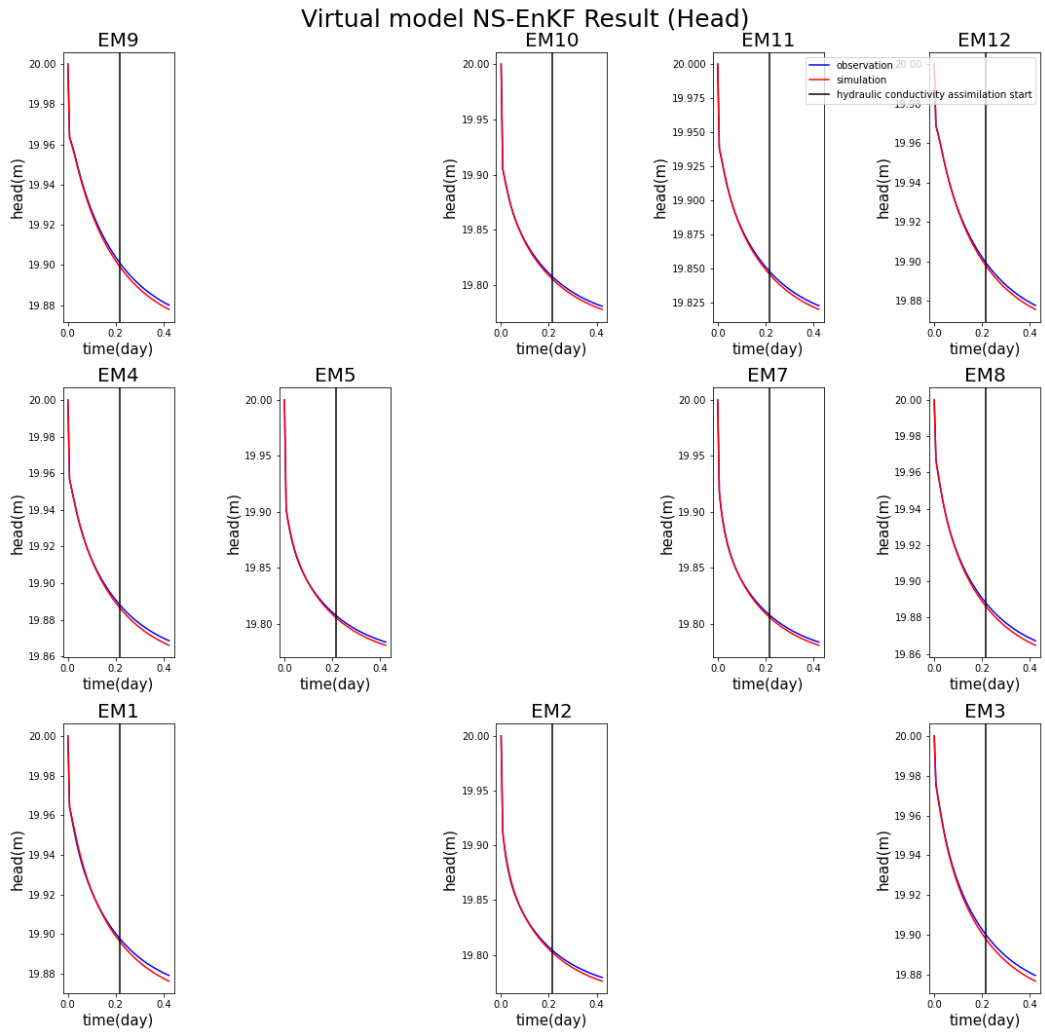

Figure 3. 4 ES20 virtual model forward modeled heads with the mean values of estimated ensembles of NS-EnKF. Blue lines indicate the correct answer considered as the observation and red lines indicate forward modeling results. Vertical black lines in the middle of the plots indicate a starting point for conditioning.

Table 3. 2 RMSE for well point and field values of each hydraulic parameter for EnKF and NS-EnKF for the virtual models and variances of Neuman solution. Differences are calculated with the mean values of parameters of ensembles and the virtual field.

	Well points		Field	
Parameters	EnKF	NS-EnKF	EnKF	NS-EnKF
Initial log K	0.026	0.026	0.016	0.016
Initial log Sy	0.181	0.181	0.127	0.127
Initial log S	0.214	0.214	0.153	0.153
Estimated log K	0.017	0.009	0.015	0.014
Estimated log Sy	0.091	0.071	0.075	0.062
Estimated log S	0.036	0.018	0.054	0.045

3.3 Application of EnKF and NS-EnKF to the field experiment data

The result of the performance test in Section 3.2 verifies that EnKF and NS-EnKF methods are valid to be used for estimating hydraulic parameters. With the pumping test conducted in the Eumseong site, both methods are implemented to estimate heterogeneity in hydraulic parameters. As there is no ground-truth data for parameters, results of log-conductivity fields from EnKF and NS-EnKF were compared with the expectation analyzed based only on the drawdown data. Furthermore, forward model results are compared with measured water levels and estimated hydraulic conductivity range is compared whether the values are in physically acceptable range.

Paramount problems in the application of EnKF and NS-EnKF methods are that no exact scope for hydraulic parameters exists, and even those parameters could change in time. Also, measuring hydraulic heads contains mechanical errors and there might be numerical errors in converting pressure units from psi to mH_2O. From these reasons, the measurement error covariance matrix R value is selected slightly higher than the Mini-Diver error range in conditioning parts as explained in Section 2.2 and 3.2.

Another problem is that the updating process might yield non-physical values. Especially specific yield value for soil has a maximum value of 0.4 (Heath, 1983). Not to generate non-physical parameters, specific yield values higher than the value of 0.4 are converted into 0.4 after the updating processes for the specific yield in the second section.

To check whether estimated parameter values are physically acceptable, maximum and minimum estimated hydraulic conductivity in the mesh grid is calculated. For the visualization of trends, the estimated hydraulic parameters are plotted (Figure 3.3 and 3.5). Also, with the estimated ensemble mean values,

forward models are simulated to compare with the pumping test data. For the parameters, RMSE with the curve-fitted parameter values are calculated to quantify the difference of estimated parameters.

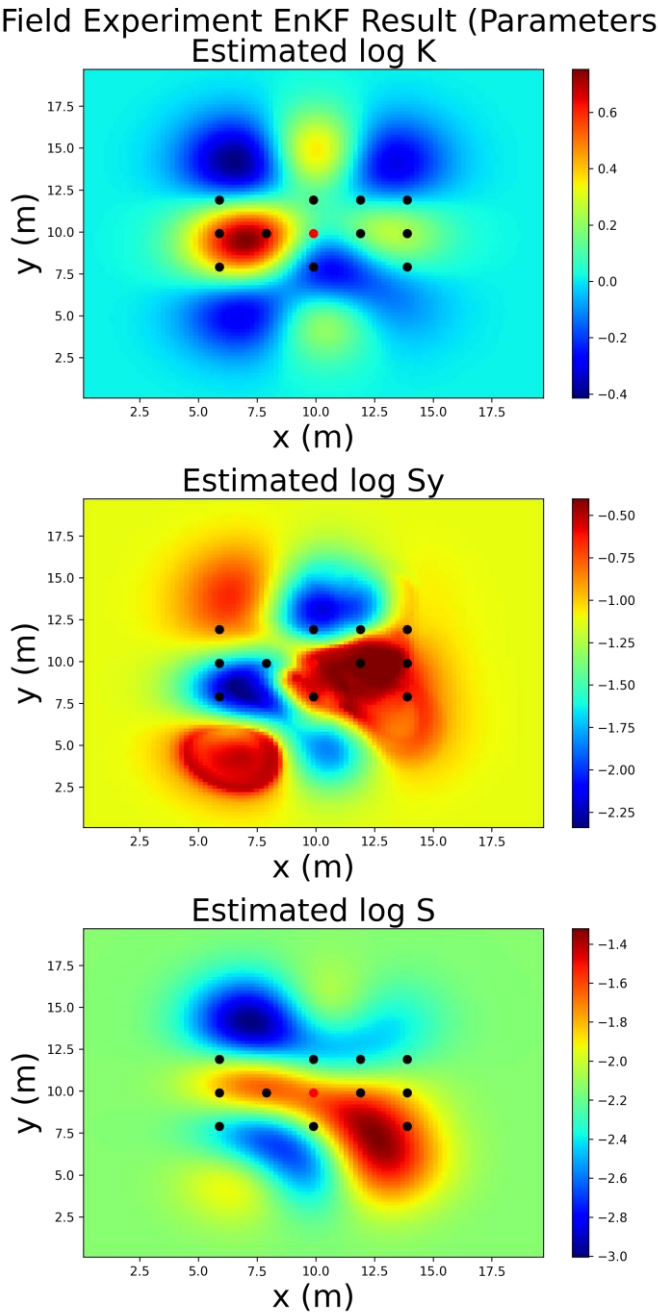

Figure 3. 5 Estimated hydraulic parameters with the EnKF method for the pumping test conducted in 2020.

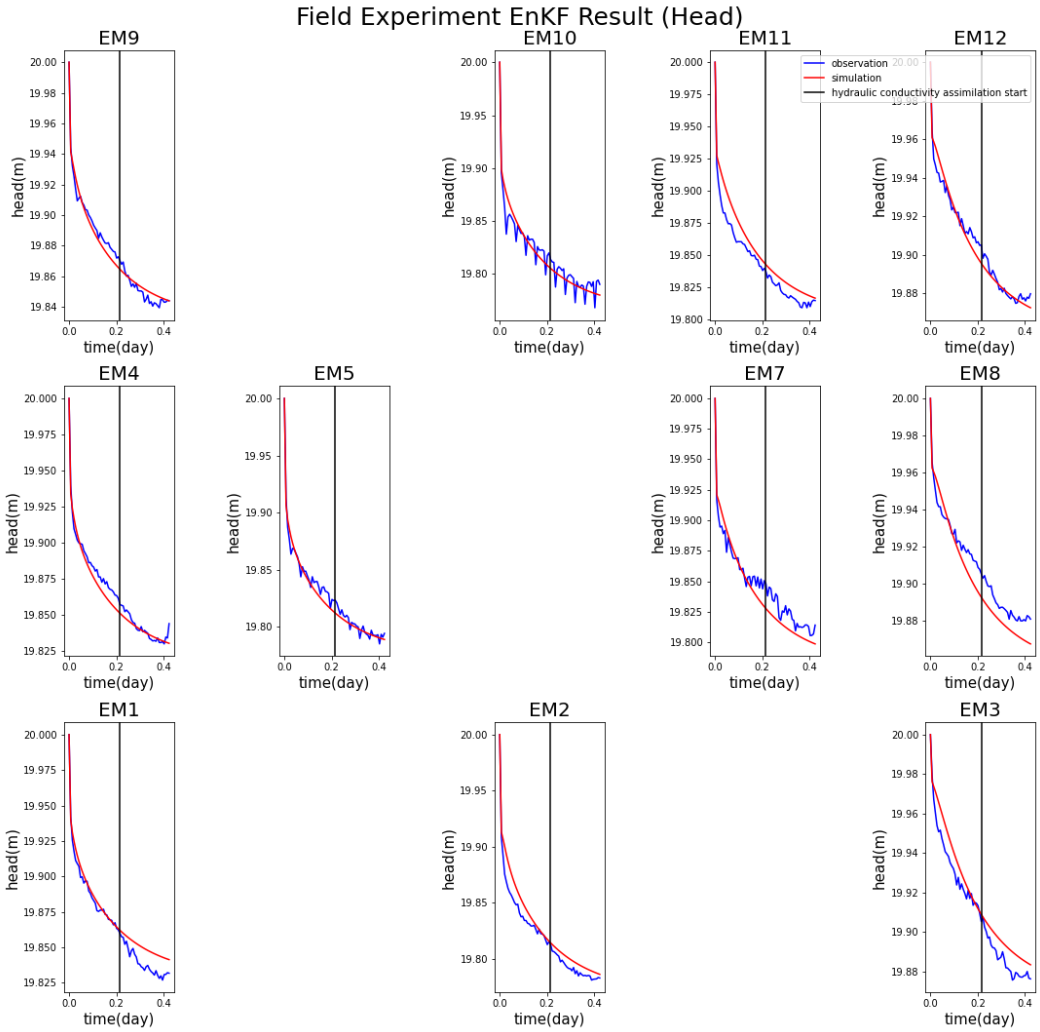

Figure 3. 6 Forward modeling head results based on the estimated parameters of 2020 pumping test data with the EnKF method.

Table 3. 3 Estimated hydraulic parameters with EnKF for pumping data. K is hydraulic conductivity, Sy is specific yield, and S is storage coefficient. Spatial units are meter and temporal units are day. All the numbers are noted to the third significant number.

Well	Estimated log K	Estimated log Sy	Estimated log S
EM-1	0.359	-2.048	-2.364
EM-2	-0.199	-0.622	-1.993
EM-3	0.037	-0.648	-1.581
EM-4	0.581	-1.659	-1.857
EM-5	0.639	-1.703	-1.665
EM-7	0.150	-0.403	-1.708
EM-8	0.239	-0.541	-1.887
EM-9	0.050	-0.871	-2.178
EM-10	0.114	-1.860	-2.366
EM-11	0.014	-1.475	-2.364
EM-12	-0.020	-0.921	-2.249
Mean	0.179	-1.159	-2.019
Std	0.257	0.598	0.299
RMSE with Neuman	0.290	0.555	0.205

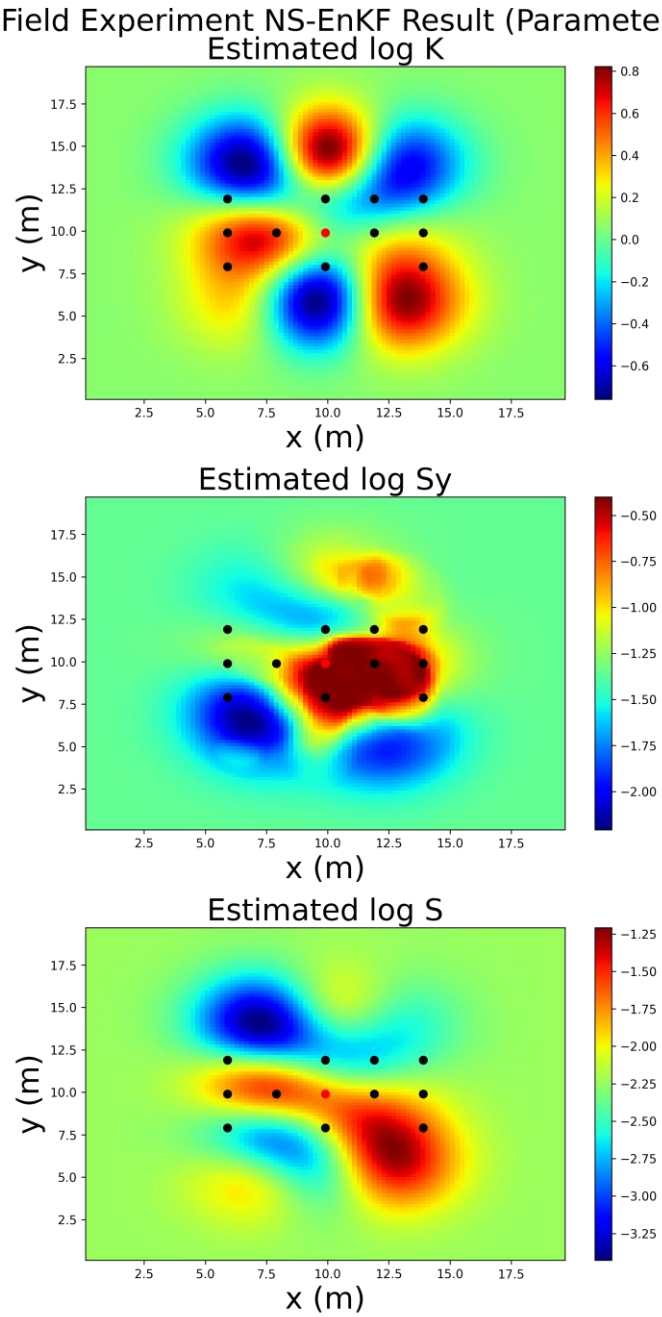

Figure 3. 7 Estimated hydraulic parameters with the NS-EnKF method for the pumping test conducted in 2020.

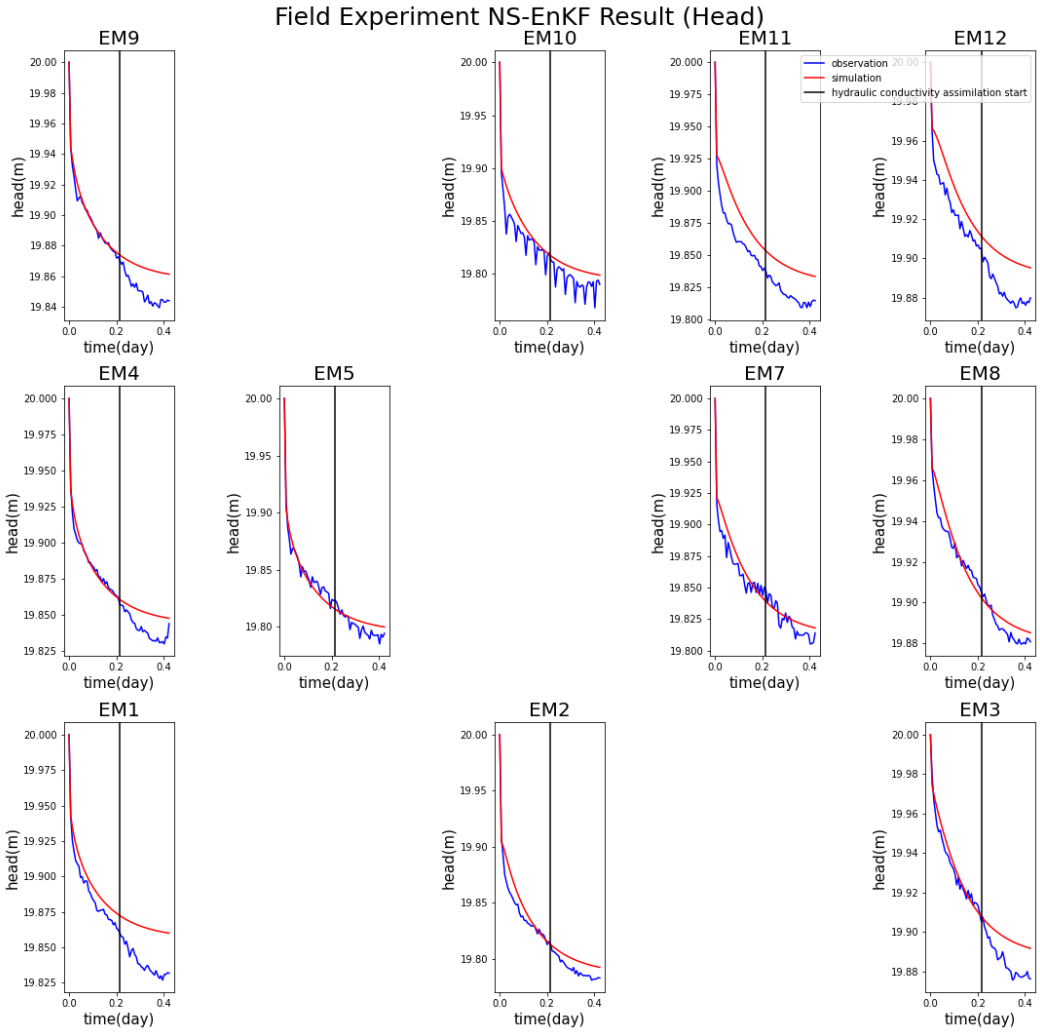

Figure 3. 8 Forward modeling head results based on the estimated parameters of 2020 pumping test data with the NS-EnKF method.

Table 3. 4 Estimated hydraulic parameters with EnKF and Neuman solution of 2021 pumping test. K is hydraulic conductivity, Sy is specific yield, and S is storage coefficient. Spatial units are meter and temporal units are day. All the numbers are noted to the second significant number. Differences are calculated by subtracting the Neuman curve-estimated values from the filter estimated values.

Well	Estimated log K	Estimated log Sy	Estimated log S
EM-1	0.497	-1.944	-2.480
EM-2	-0.256	-0.413	-2.027
EM-3	0.566	-0.790	-1.591
EM-4	0.528	-1.329	-1.833
EM-5	0.558	-1.050	-1.606
EM-7	0.070	-0.405	-1.872
EM-8	0.187	-0.456	-2.020
EM-9	-0.152	-1.333	-2.307
EM-10	-0.023	-1.502	-2.586
EM-11	-0.342	-1.163	-2.609
EM-12	-0.308	-0.979	-2.430
Mean	0.121	-1.033	-2.124
Std	0.366	0.492	0.377
RMSE with Neuman	0.378	0.372	0.294

Table 3. 5 Estimated hydraulic conductivity range in the mesh grid. Parenthesis under each value is changed into the unit value in cm/sec.

	Hydraulic Conductivity (K, m/day)	
	EnKF	NS-EnKF
Maximum	5.662 (6.554×10^{-3})	6.653 (7.700×10^{-3})
Minimum	0.386 (4.472×10^{-4})	0.174 (2.016×10^{-4})

The Eumseong experiment site is known to composed of a soil with silt, clay, and sand (Joun and Lee, 2020). The estimated hydraulic conductivity is converted into the unit of cm/s, and the range is in the range of silty sand as in Figure 3. 9 (Freeze and Cherry, 1979) as both EnKF and NS-EnKF estimated hydraulic conductivity to be maximum order of -3 and minimum -4 in Table 3. 5. In other words, the estimated parameter values are not non-physical and coincides with the previous study.

Moreover, in Table 3. 3 and 3.4, estimated parameters show high differences compared to the curve-fitted parameter range, as estimated parameters are out of the range of a curve-fitted range.

Only with the head data distribution in the site, dividing sections based on the pumping well, drawdowns in the upper side and bottom side are shown to be symmetric, so the distribution of hydraulic conductivity is expected to be symmetric. This expectation agrees with the result of a hydraulic conductivity distribution estimated with EnKF in Figure 3. 5. NS-EnKF result in Figure 3. 7 collides with the expectations, and also forward simulation results in Figure 3. 8 show higher differences compared to the EnKF result in Figure 3. 6. In sum, EnKF shows higher performance for the experiment data compared to NS-EnKF, and both methods estimate parameters in physical range and are valid.

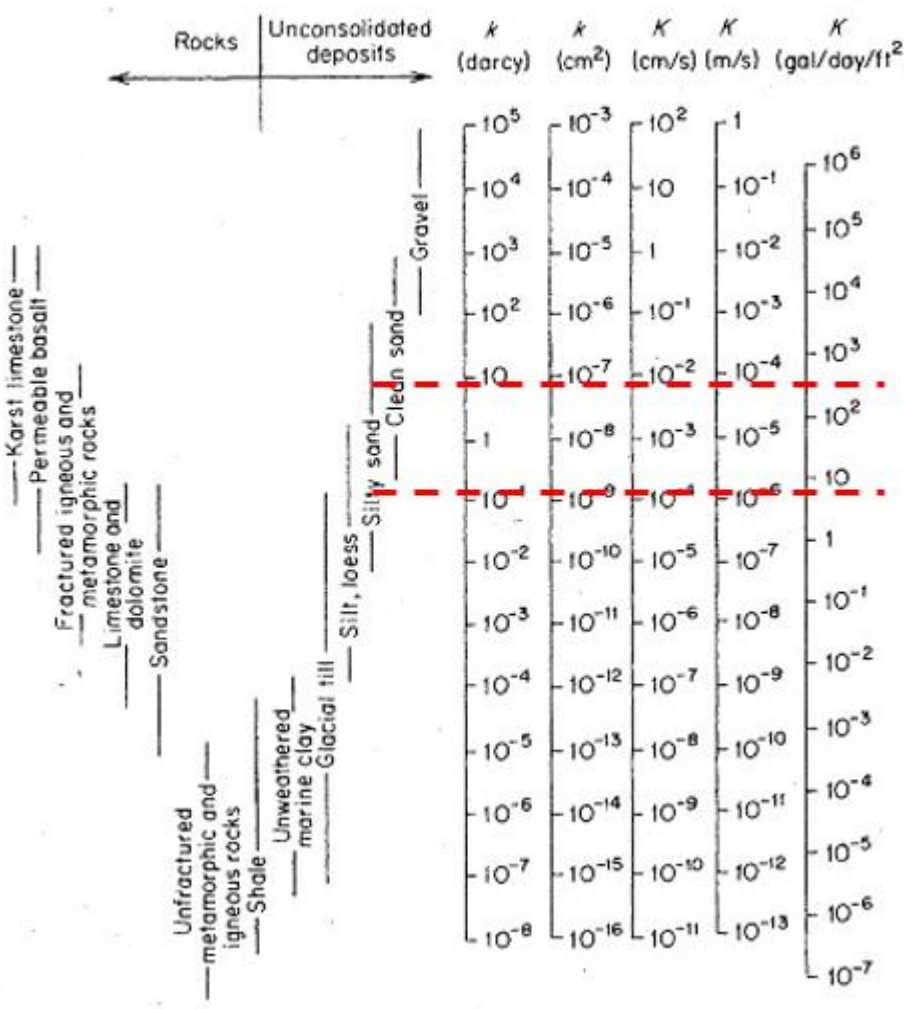

Figure 3. 9 Hydraulic conductivity ranges of various soil types from Freeze and Cherry (1979). Red dotted lines plot the estimated hydraulic conductivity range, and are in the range of silty sand.

After, Neuman curve-fitted parameters are averaged with the homogeneous assumption of the aquifer, and then modeling results are compared with the forward simulation results with the estimated parameters with EnKF and NS-EnKF to visualize the improvements of combining the filters with curve-fitted parameters.

In Table 3. 6, both EnKF and NS-EnKF estimated forward simulations have smaller mean RMSE compared to the curve-fitted forward simulation. Homogeneity assumed curve-fitted forward simulation shows the smallest RMSE with EM-11 data compared to EnKF and NS-EnKF and can be visualized in Figure 3. 10. However, the curve-fitted parameter with EM-11 observed drawdown has the largest difference in hydraulic conductivity compared to the mean value in Table 3. 1, so this implies that curve-fitting method parameters are imperfect. Both filters improve the estimation of hydraulic parameters obtained from curve-fitting method, and contribute more accurate numerical simulations compared to only using curve-fitted parameters with a homogeneity assumption. Referring to the results of EnKF applied pumping test analysis, not only restricted to the pumping test application, but also for experiments which have been preceded with homogeneous curve-fitted parameters are expected to be enhanced with both EnKF and NS-EnKF methods.

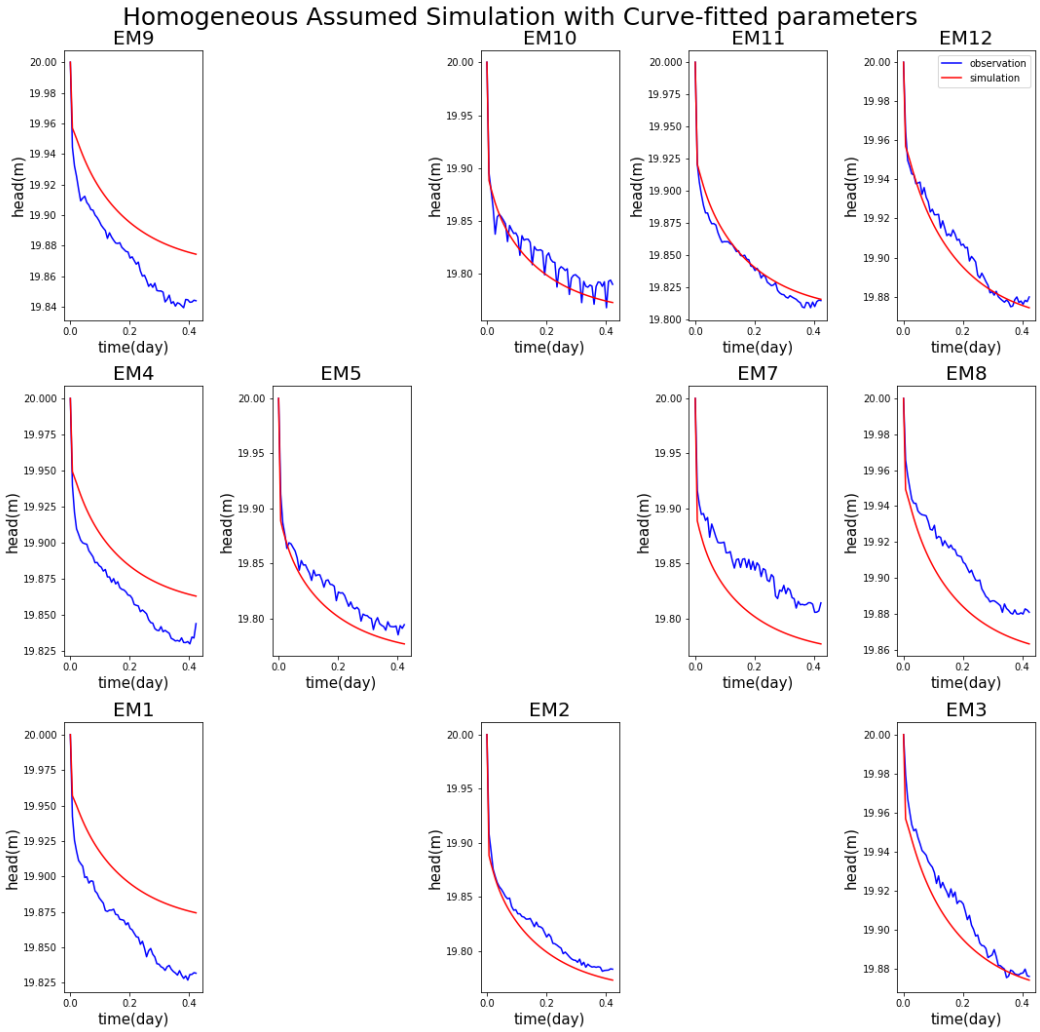

Figure 3. 10 Forward simulation with mean values of hydraulic conductivity, specific storage, and storage coefficient which are estimated with Neuman curve-fitted parameters with 11 observation wells. Red lines indicate the simulation results with a homogeneity. Blue lines indicate observed water level.

Table 3. 6 RMSE of the hydraulic head difference of forward simulation and observed data

Forward simulation RMSE			
Well	Curve-fitted	EnKF	NS-EnKF
EM-1	0.038	0.009	0.026
EM-2	0.011	0.008	0.013
EM-3	0.011	0.008	0.011
EM-4	0.027	0.008	0.015
EM-5	0.017	0.008	0.009
EM-6			
EM-7	0.037	0.013	0.008
EM-8	0.019	0.010	0.005
EM-9	0.028	0.006	0.017
EM-10	0.014	0.010	0.015
EM-11	0.007	0.009	0.022
EM-12	0.006	0.004	0.014
Mean	0.019	0.008	0.014

4 Conclusion

In this study, hydraulic parameters are estimated with Ensemble Kalman Filter and Normal-score Ensemble Kalman filter. Filters present better groundwater head distributions compared to simulation results with simple curve-fitted parameters. Both display a good agreement with reproduction of water level, compared to those of values with Neuman curve-fitted method.

In the comparison of the performances between EnKF and NS-EnKF without measurement errors in head, NS-EnKF displays better performance in estimating hydraulic parameters constructed with ordinary kriging. On the other hand, an application of EnKF to the field experiment with head measurement errors displays better performances than NS-EnKF. Even though both estimated hydraulic conductivity distributions are in the range of silty sand, EnKF coincides with the expectations of the distribution of hydraulic conductivity and is preferred.

For the field experiment pumping test, applying the EnKF for numerical models displays better performance in reproduction of drawdown distributions than only using the curve-fitted parameters for numerical modeling. However, estimation results from both methods are in physical range and are valid for improving forward simulation with the field experiments. Not only the pumping test, but also a variety of groundwater experiments is conducted in the field site with the usage of hydraulic parameters, such as a solute transport experiment with the usage of a hydraulic conductivity parameter. Both filters are expected to improve the intricate groundwater simulations with a precise consideration of hydraulic parameters.

5 Reference

Bakker, M., Post, V. ., Langevin, C. D., Hughes, J. D., White, J. T., Starn, J. J., and Fienen, M. N.: Scripting MODFLOW model development using Python and FloPy, 54, 733–739, 2016.

Brunton, S. L. and Kutz, J. N.: Data-driven science and engineering: Machine learning, dynamical systems, and control, Cambridge University Press, 2019.

Burgers, G., Van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, Mon. Weather Rev., 126, 1719–1724, https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2, 1998.

Camporese, M., Crestani, E., and Salandin, P.: Assessment of local hydraulic parameters by Enkf data assimilation in real aquifers: a case study in downtown Padova (Italy), 2012.

Chen, Y. and Zhang, D.: Data assimilation for transient flow in geologic formations via ensemble Kalman filter, Adv. Water Resour., 29, 1107–1122, https://doi.org/10.1016/j.advwatres.2005.09.007, 2006.

Chen, Z., Gómez-Hernández, J. J., Xu, T., and Zanini, A.: Joint identification of contaminant source and aquifer geometry in a sandbox experiment with the restart ensemble Kalman filter, J. Hydrol., 564, 1074–1084, https://doi.org/10.1016/j.jhydrol.2018.07.073, 2018.

Drécourt, J.-P.: Kalman filtering in hydrological modelling, 43, 2003.

Erdal, D. and Cirpka, O. A.: Joint inference of groundwater–recharge and hydraulic–conductivity fields from head data using the ensemble Kalman filter, Hydrol. Earth Syst. Sci., 20, 555–569, https://doi.org/10.5194/hess-20-555-2016, 2016.

Fetter, C. W.: Applied hydrogeology, Waveland Press, 2018.

Franssen, H. J. H., Gómez-Hernández, J. J., Capilla, J. E., and Sahuquillo, A.: Joint simulation of transmissivity and storativity fields conditional to steady-state and transient hydraulic head data, Adv. Water Resour., 23, 1–13, https://doi.org/10.1016/S0309-1708(99)00006-8, 1999.

Freeze, R. A.: A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media, Water Resour. Res., 11, 725–741, https://doi.org/10.1029/WR011i005p00725, 1975.

Freeze, R. A. and Cherry, J. A.: GroundWater, Prentice-hall, 1979.

Gillijns, S., Mendoza, O. B., Chandrasekar, J., De Moor, B. L. R., Bernstein, D. S., and Ridley, A.: What is the ensemble Kalman filter and how well does it work?, Proc. Am. Control Conf., 2006, 4448–4453, https://doi.org/10.1109/acc.2006.1657419, 2006.

Gu, Y. and Oliver, D. S.: An Iterative Ensemble Kalman Filter for Multiphase Fluid Flow Data Assimilation, SPE J., 12, 438–446, https://doi.org/10.2118/108438-PA, 2007.

Harbaugh, A. W.: MODFLOW-2005, the US Geological Survey modular ground-water model: the ground-water flow process, US Department of the Interior, US Geological Survey Reston, VA, USA, 2005.

Heath, R. C.: Basic ground-water hydrology, Journal of Organizational Behavior, 691–706 pp., https://doi.org/10.3133/wsp2220, 1983.

Hendricks Franssen, H. J. and Kinzelbach, W.: 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, 1–21, https://doi.org/10.1029/2007WR006505, 2008.

Joun, W. T. and Lee, K. K.: Reproducing natural variations in CO2 concentration in vadose zone wells with observed atmospheric pressure and groundwater data, J. Environ. Manage., 266, 110568, https://doi.org/10.1016/j.jenvman.2020.110568, 2020.

Keating, E. and Zyvoloski, G.: A stable and efficient numerical algorithm for unconfined aquifer analysis, Ground Water, 47, 569–579, https://doi.org/10.1111/j.1745-6584.2009.00555.x, 2009.

Kumar, G. N. P. and Kumar, P. A.: Development of Groundwater Flow Model Using Visual MODFLOW, 2, 649–656, 2014.

Kushwaha, R. K., Pandit, M. K., and Goyal, R.: MODFLOW based groundwater resource evaluation and prediction in Mendha sub-basin, NE Rajasthan, J. Geol. Soc. India, 74, 449–458, https://doi.org/10.1007/s12594-009-0154-1, 2009.

Langevin, C. D., Hughes, J. D., Banta, E. R., Niswonger, R. G., Panday, S., and Provost, A. M.: Documentation for the MODFLOW 6 Groundwater Flow Model: U.S. Geological Survey Techniques and Methods, B. 6, Model. Tech., 197, 2017.

Li, L., Zhou, H., Hendricks Franssen, H. J., and Gómez-Hernández, J. J.: Groundwater flow inverse modeling in non-MultiGaussian media: Performance assessment of the normal-score Ensemble Kalman Filter, Hydrol. Earth Syst. Sci., 16, 573–590, https://doi.org/10.5194/hess-16-573-2012, 2012a.

Li, L., Zhou, H., Gómez-Hernández, J. J., and Hendricks Franssen, H. J.: Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via ensemble Kalman filter, J. Hydrol., 428–429, 152–169, https://doi.org/10.1016/j.jhydrol.2012.01.037, 2012b.

McDonald, M. G. and Harbaugh, A. W.: The history of MODFLOW., Ground Water, 41, 280–283, https://doi.org/10.1111/j.1745-6584.2003.tb02591.x, 2003.

Moench, A. F.: Specific Yield as Determined by Type-Curve Analysis of Aquifer-Test Data, Ground Water, 32, 949–957, https://doi.org/10.1111/j.1745-6584.1994.tb00934.x, 1994.

Moore, J. E.: Contribution of groundwater modeling to planning, Dev. Water Sci., 12, 121–128, https://doi.org/10.1016/S0167-5648(09)70013-9, 1979.

Neuman, S. P.: Effect of partial penetration on flow in unconfined aquifers considering delayed gravity response, Water Resour. Res., 10, 303–312, 1974.

Neuman, S. P.: Analysis of pumping test data from anisotropic unconfined aquifers considering delayed gravity response, Water Resour. Res., 11, 329–342, https://doi.org/10.1029/WR011i002p00329, 1975.

Niswonger, R.: MODFLOW-NWT: a Newton Formulation for MODFLOW-2005, https://doi.org/10.5066/F7XP7317, 2014.

Niswonger, R. G., Panday, S., and Ibaraki, M.: MODFLOW-NWT, a Newton formulation for MODFLOW-2005, US Geol. Surv. Tech. Methods, 6, 44, 2011.

Sadat-Noori, M., Glamore, W., and Khojasteh, D.: Groundwater level prediction using genetic programming: the importance of precipitation data and weather station location on model accuracy, Environ. Earth Sci., 79, 1–10, https://doi.org/10.1007/s12665-019-8776-0, 2020.

Shakoor, A., Arshad, M., Ahmad, R., Khan, Z. M., Qamar, U., Farid, H. U., Sultan, M., and Ahmad, F.: Development of groundwater flow model (Modflow) to simulate the escalating groundwater pumping in the Punjab, Pakistan, Pakistan J. Agric. Sci., 55, 635–644, https://doi.org/10.21162/PAKJAS/18.4909, 2018.

Singh, A., Minsker, B. S., and Valocchi, A. J.: An interactive multi-objective optimization framework for groundwater inverse modeling, Adv. Water Resour., 31, 1269–1283, https://doi.org/10.1016/j.advwatres.2008.05.005, 2008.

Sun, A. Y., Morris, A., and Mohanty, S.: Comparison of deterministic ensemble Kalman filters for assimilating hydrogeological data, Adv. Water Resour., 32, 280–292, https://doi.org/10.1016/j.advwatres.2008.11.006, 2009.

Sun, L., Seidou, O., Nistor, I., and Liu, K.: Review of the Kalman-type hydrological data assimilation, Hydrol. Sci. J., 61, 2348–2366, https://doi.org/10.1080/02626667.2015.1127376, 2016.

Tong, J., Hu, B. X., and Yang, J.: Assimilating transient groundwater flow data via a localized ensemble Kalman filter to calibrate a heterogeneous conductivity field, Stoch. Environ. Res. Risk Assess., 26, 467–478, https://doi.org/10.1007/s00477-011-0534-0, 2012.

Tong, J. X., Hu, B. X., and Yang, J. Z.: Comparison between gradient based UCODE_2005 and the ensemble Kalman Filter for transient groundwater flow inverse modeling, Sci. China Earth Sci., 60, 899–909, https://doi.org/10.1007/s11430-015-0235-1, 2017.

Wackernagel, H.: Multivariate Geostatistics, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-662-05294-5, 2003.

Xu, T., Jaime Gómez-Hernández, J., Li, L., and Zhou, H.: Parallelized ensemble Kalman filter for hydraulic conductivity characterization, Comput. Geosci., 52, 42–49, https://doi.org/10.1016/j.cageo.2012.10.007, 2013a.

Xu, T., Jaime Gómez-Hernández, J., Zhou, H., and Li, L.: The power of transient piezometric head data in inverse modeling: An application of the localized normal-score EnKF with covariance inflation in a heterogenous bimodal hydraulic conductivity field, Adv. Water Resour., 54, 100–118, https://doi.org/10.1016/j.advwatres.2013.01.006, 2013b.

Yadav, B., Gupta, P. K., Patidar, N., and Himanshu, S. K.: Ensemble modelling framework for groundwater level prediction in urban areas of India, Sci. Total Environ., 712, 135539, https://doi.org/10.1016/j.scitotenv.2019.135539, 2020.

Yeh, W. W.: Review of Parameter Identification Procedures in Groundwater Hydrology: The Inverse Problem, Water Resour. Res., 22, 95–108, https://doi.org/10.1029/WR022i002p00095, 1986.

Zhou, H., Gómez-Hernández, J. J., Hendricks Franssen, H. J., and Li, L.: An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering, Adv. Water Resour., 34, 844–864, https://doi.org/10.1016/j.advwatres.2011.04.014, 2011.

Zhou, H., Li, L., Franssen, H. J. H., and Gómez-Hernández, J. J.: Pattern Recognition in a Bimodal Aquifer Using the Normal-Score Ensemble Kalman Filter, Math. Geosci., 44, 169–185, https://doi.org/10.1007/s11004-011-9372-3, 2012.

Zovi, F., Camporese, M., Hendricks Franssen, H. J., Huisman, J. A., and Salandin, P.: Identification of high-permeability subsurface structures with multiple point geostatistics and normal score ensemble Kalman filter, J. Hydrol., 548, 208–224, https://doi.org/10.1016/j.jhydrol.2017.02.056, 2017a.

국문 초록

세계적인 기후 변화에 따라 지하수 사용의 관리에의 관심이 높아지고 있다. 일반적으로 지하수 사용에 있어서의 영향을 정량적으로 평가하고 예측하기 위하여 지하수 모델링이 사용된다. 대수층 파라미터들은 직접적인 관측에서의 어려움이 존재하며 불균질성이 존재하기에 지하수의 흐름을 결정함에도 불구하고 약한 제약 조건을 건다. 데이터 동화 방법으로 이 파라미터들을 추정할 수 있으며 더욱 정밀한 시뮬레이션을 가능하게 한다. 본 연구에서는 자료 동화 방법으로 앙상블 칼만 필터(EnKF)와 표준 정규화 앙상블 칼만 필터(NS-EnKF) 방법들을 활용하여 음성 현장 양수 시험 자료에 적용하여 수리전도도, 비산출량, 저류 계수를 추정하고자 한다. 필터들의 성능은 또한 현장 자료에 적용되기 이전에 검증 과정을 거쳤다.

필터들의 성능을 검증하는 데에 있어서 관측 자료에의 섭동은 존재하지 않기에 NS-EnKF 가 세 가지의 파라미터를 추정하는 데에 있어 더욱 높은 성능을 나타내었다. 또한 초기 앙상블들의 평균값들과 비교하였을 때 추정된 파라미터들의 평균 제곱근 편차가 더욱 낮아지는 것을 확인하며 필터를 사용한 파라미터의 추정이 유효함을 확인하였다.

현장 자료에 두 가지의 필터들을 적용함에 따라 수리전도도의 분포에 있어서의 예측되는 분포의 형태, 추정된 파라미터를 바탕으로 한 시뮬레이션과 관측 값의 비교를 통하여 EnKF 가 현장 자료에 있어서 더욱 높은 성능을 보여주는 것을 확인할 수 있었다. 또한, 추정된 수리 전도도의 값의 범위가 토양의 구성에 있어서의 물리적인 범위 내에 존재함에 따라 현실적인 추정임을 검증하였다. 본 연구는 두 가지의 필터를 활용하고 비교 분석, 검증하는 과정을 바탕으로 두 가지 방법을 기존의 파라미터 추정 방법과 결합하게 된다면 더욱 정밀하며 현실적인 지하수 모델링이 가능함을 시사한다.

주요어 : 수리전도도, 비산출량, 저류 계수, 앙상블 칼만 필터, 표준 정규화 앙상블 칼만 필터

학번 : 2020-22099