

Research papers

Stochastic analysis of oscillatory hydraulic tomography



Yu-Li Wang ^a, Tian-Chyi Jim Yeh ^{a,b,*}, Dong Xu ^c, Kuangjia Li ^d, Jet-Chau Wen ^{e,f}, Shao-Yang Huang ^f, Wenke Wang ^g, Yonghong Hao ^b

^a Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ, United States

^b Tianjin Key Laboratory of Water Resources and Environment, Tianjin Normal University, Tianjin, China

^c State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, China

^d Development and Research Center, Ministry of Water Resources, China

^e Department of Safety, Health, and Environmental Engineering, National Yunlin University of Science and Technology, Douliu, Taiwan

^f Research Center for Soil and Water Resources and Natural Disaster Prevention, National Yunlin University of Science and Technology, Douliu, Taiwan

^g Key Laboratory of Subsurface Hydrology and Ecological Effects in Arid Region, Chang'an University, Xi'an, China

ARTICLE INFO

This manuscript was handled by Corrado Corradi, Editor-in-Chief

ABSTRACT

This paper investigates the effectiveness of different oscillatory hydraulic tomography (OHT) frequencies for estimating heterogeneous fields. The analysis first formulates the effects of uncharacterized aquifer responses, T , and S fields as the ensemble mean residual flux and residual storage terms in the stochastic unconditional and conditional mean equations. These terms persist unless the T , S , or head fields are specified everywhere. We then conducted OHT to estimate the T and S fields using Monte Carlo experiments. The experiments show that using the heads in response to periodic pumping with different frequencies or multifrequency, the estimates' performance metrics vary from one realization to others. The mean performance metrics over many realizations are, however, indistinguishable, despite the frequency. We attribute the variation in the performance metrics to the lack of parameter or state variable ergodicity. Lastly, we emphasize the importance of dense monitoring networks and a cost-effective data collection strategy to improve the resolution of characterizing aquifer heterogeneity.

1. Introduction

Conventionally, transmissivity (T) and storage coefficient (S) are estimated using a constant rate pumping or injection test and methods of analysis such as [Thiem \(1906\)](#), [Theis \(1935\)](#), or [Cooper and Jacob \(1946\)](#). Periodic pumping tests (PPT), hydraulic tests with periodically varying pumping rates, have also been proposed to estimate these properties. For example, [Black and Kipp \(1981\)](#) developed methods for estimating diffusivity based on analytical solutions in terms of the pressure attenuation and the phase lag during a sinusoidal point or line excitations. They concluded that the methods are not a substitute for conventional aquifer testing due to time and equipment costs. Still, they are a useful tool to apply in specific applications.

[Butler and McElwee \(1990\)](#) examined the head's sensitivity at pumping well to different known T values at 6 zones, concentrically circling the well (a 1-D radial flow). They stated that variations in the pumping rate increase the drawdown's sensitivities to each layer's property. Without estimating the property, they concluded that the

estimate's uncertainty could be reduced.

[Rasmussen et al. \(2003\)](#) and [Renner and Messar \(2006\)](#) applied the PPT to field aquifers. They found that the estimated properties were hydraulic properties averaged over different volumes of heterogeneous subsurface and consistent with conventional aquifer tests. They emphasized that the approach minimizes investigation-derived wastes (e.g., contaminated groundwater) and introduces a signal different from background perturbations (easy to detect). With small-scale soil column experiments, [Song and Renner \(2008\)](#) reported that the pressure analysis yields a relatively constant permeability for the high porosity sample at different pressures and oscillation periods. They found slightly period dependent permeability for the low porosity samples and the stacked sample. The period dependence diminishes with increasing period.

[Rabinovich et al. \(2013\)](#) theoretically investigated heterogeneous aquifers' effective conductivity in unsteady periodic flow by a stochastic approach. They concluded that the effective conductivity and the ensemble mean head and flux have significant dynamic effects. On the

* Corresponding author.

E-mail address: yeh@hwr.arizona.edu (T.-C.J. Yeh).

other hand, [Guiltinan and Becker \(2015\)](#) used constant rate pumping tests and periodic slug tests to detect wells' hydraulic connectivity in fractured bedrock at a field site. They neglected early time drawdown behaviors and analyzed the effective hydraulic diffusivity by fitting the late time drawdown to Theis curve based on homogeneous aquifers. They suggested that the periodic tests' estimated field is more sensitive to the hydraulic connectivity among well pairs than those from the constant-rate pumping test. The value of estimated hydraulic diffusivity from the periodic tests approaches that of the constant rate test as the oscillation frequency decreases. Similarly, the usefulness of periodic pumping has also been investigated by [Fischer et al. \(2018\)](#), [Schuite et al. \(2017\)](#), and [Lavenue and deMarsily \(2001\)](#) for the karstic and fractured aquifer characterization, by [Sun et al. \(2015\)](#) and [Sun et al. \(2017\)](#) for leakage detection, and by [Rabinovich et al. \(2015\)](#) for imaging unconfined aquifer.

[Ahn and Horne \(2010\)](#) developed analytical solutions for attenuation and phase shift of pressure pulse testing in a 1-D radial-flow through 3 concentric layers of different permeability (with 10 unknown permeability blocks) surrounding the test well. They reported that the attenuation and phase shift provided an 'indicator characteristic' to reveal reservoir heterogeneity. Likewise, based on numerical experiments, [Cardiff et al. \(2013\)](#) reported that including head data from pumping with different frequencies can improve the aquifer hydraulic parameter estimates, even if the pumping is only conducted in one pumping well.

On the other hand, hydraulic tomography (HT) with constant pumping rates has been proposed as a new aquifer characterization method ([Yeh and Liu, 2000](#); [Zhu and Yeh, 2005](#)). It has been proven over the past two decades as a matured technology for high-resolution characterization of saturated and variably saturated porous media (e.g., [Cardiff and Barrash, 2011](#); [Mao et al., 2013b](#); [Berg and Illman, 2011, 2015](#); [Zhao and Illman, 2018](#)) and fractured rocks at fields and laboratories ([Illman et al., 2009](#); [Zha et al., 2016](#)). Further, theoretical and physical explanations of the robustness of tomographic surveys have been provided by [Wen et al. \(2019\)](#), [Sun et al. \(2013\)](#), [Huang et al. \(2011\)](#), and [Yeh et al. \(2014\)](#).

Recently, [Zhou et al. \(2016\)](#) reported that head collected from oscillatory hydraulic tomography (OHT) with multifrequency excitations provides better estimates than those from single-frequency excitations when the number of pumping tests and observation locations is limited. On the contrary, results of the Monte Carlo experiment of aquifer characterization, using river stage variation with different frequencies, by [Wang et al. \(2019\)](#) signify that the estimated T field's ensemble-averaged performance is independent of frequency. They implied that previously reported improvements of estimates using different frequencies based on one realization of heterogeneity are not warranted for all possible realizations. However, the river stage tomography is not the same as OHT, and [Wang et al. \(2019\)](#) only investigated the effect of different single frequencies and did not evaluate the performance of river stage variations of multifrequency.

This paper aims to explain that neglecting heterogeneity introduces the frequency- and scenario-dependence estimated parameters. More importantly, this paper demonstrates that excitations of multifrequency do not overcome difficulties associated with the under-determined nature of the inverse problems and do not necessarily improve the heterogeneity estimates. Finally, the importance of considering flow ergodicity issues for evaluating the robustness of inverse models is stressed.

2. Groundwater flow model

This study assumes that the following equations describe 2-D plan-view groundwater flow induced by periodic pumping at a well during OHT in a heterogeneous confined aquifer.

$$\nabla \cdot [T(\mathbf{x}) \cdot \nabla h(\mathbf{x}, t)] = S(\mathbf{x}) \frac{\partial h(\mathbf{x}, t)}{\partial t} + Q(t) \delta(\mathbf{x} - \mathbf{x}_p) \quad (1)$$

where h represents the head responses (L), T is transmissivity (L_2/T), S is storage coefficient (-), \mathbf{x} is the vector in x and y directions, $\delta(\cdot)$ is Dirac delta function, \mathbf{x}_p denotes the coordinate of pumping well, and T represents time (T). The periodic pumping rate with a given amplitude Q_0 is expressed as

$$Q(t) = Q_0 \exp(2\pi i \omega t) \quad (2)$$

in which ω is the frequency (cycle/T) and i is the imaginary unit. The positive pumping rate represents extraction. The piezometric head's initial condition is uniform, and the boundary condition can be prescribed head or flux boundaries. Expressing the above equations in the frequency domain leads to

$$\nabla \cdot [T(\mathbf{x}) \nabla \cdot \phi] = i\omega S(\mathbf{x}) \phi + Q_0 \quad (3)$$

where ϕ is the phasor (L), which is a complex variable containing the information of head amplitude and phase lag. The boundary condition can be prescribed phasor or flux boundaries.

3. Stochastic analysis of PPT in heterogeneous aquifers

Investigation of the effects of OHT on the estimated T and S in aquifers is most appropriate to formulate the problem in a stochastic framework and consider the periodic pumping at a single well first. Since $T(\mathbf{x})$ and $S(\mathbf{x})$ are spatially varying, and they are difficult to determine at every location of the aquifer, the stochastic analysis conceptualizes T and S as random fields, which can be expressed as

$$T(\mathbf{x}) = \bar{T} + T'(\mathbf{x}), \text{ and } S(\mathbf{x}) = \bar{S} + S'(\mathbf{x}) \quad (4)$$

where \bar{T} and \bar{S} are unconditional means, which are invariant in space and ensemble. Note that the overhead bar stands for the expected value. $T'(\mathbf{x})$ The unconditional perturbations at each \mathbf{x} are characterized by their unconditional covariance functions, which are assumed to be exponential. These covariance functions specify the spatial variance (i.e., the perturbations' uncertainty) and the spatial relationship between perturbations in the ensemble sense.

Accordingly, many possible $\phi(\mathbf{x}, \omega)$ fields exist in this aquifer in response to an oscillatory pumping test with a given Q_0 at a frequency ω . These fields are thus considered as a random field and can be expressed by

$$\phi(\mathbf{x}, \omega) = \bar{\phi}(\mathbf{x}, \omega) + \phi'(\mathbf{x}, \omega) \quad (5)$$

in which $\bar{\phi}(\mathbf{x}, \omega)$ is the unconditional ensemble mean and $\phi'(\mathbf{x}, \omega)$ is the perturbation at \mathbf{x} with a frequency ω .

Substituting Eqs. (4) and (5) into Eq. (3), we have

$$\begin{aligned} \nabla \cdot [(\bar{T} + T'(\mathbf{x})) \nabla \cdot (\bar{\phi}(\mathbf{x}, \omega) + \phi'(\mathbf{x}, \omega))] \\ = i\omega (\bar{S} + S'(\mathbf{x})) (\bar{\phi}(\mathbf{x}, \omega) + \phi'(\mathbf{x}, \omega)) + Q_0 \end{aligned} \quad (6)$$

Instead of all possible $\phi(\mathbf{x})$ fields induced by the pumping with a given frequency, we seek the most likely unconditional mean of $\phi(\mathbf{x})$. Thus, we take the expected value of Eq. (6), and we have the following:

$$E[T'] = E[S'] = E[\phi'] = 0 \text{ and } E[\nabla \cdot (T'(\mathbf{x}) \nabla \phi'(\mathbf{x}, \omega))] = \nabla \cdot E[T'(\mathbf{x}) \nabla \phi'(\mathbf{x}, \omega)] \quad (7)$$

Note that perturbation expectation is zero, and the dominated convergence theorem is $\lim_{x \rightarrow 0} \bar{\nabla f} = \bar{\nabla f}$ (f is any function). Substituting Eqs. (7) into Eq. (6), we have

$$\nabla \cdot (\bar{T} \nabla \bar{\phi}) + \nabla \cdot (\bar{T}' \nabla \phi') = i\omega \bar{S} \bar{\phi} + i\omega \bar{S}' \phi' + Q_0 \quad (8)$$

where again, the overhead bar denotes the expected value (or ensemble average). Notice that the (\mathbf{x}, ω) for ϕ and (\mathbf{x}) for T' and S' are dropped hereafter. This equation is called the ensemble unconditional mean flow

equation. It describes $\bar{\phi}$ field (the most likely ϕ field) in responses to an oscillatory pumping test with a given frequency. This ensemble-mean equation shows that the unconditional \bar{T} , \bar{S} and $\bar{\phi}$ alone do not satisfy the mass balance equation unless two additional terms are included. One of the terms, $\bar{T}'\bar{\nabla}\bar{\phi}'$ or the ensemble unconditional mean residual flux (e.g., Neuman and Orr, 1993; Guadagnini and Neuman, 1999), is the unconditional cross-covariance of variations between $\bar{\nabla}\phi$ and T . The other term ($\bar{S}'\bar{\phi}'$) is the ensemble unconditional mean residual storage or the cross-covariance of variations between S and ϕ . These two terms represent the contributions of unknown perturbations (T' , S' , and ϕ') to the mean flow equation (Gelhar, 1986).

3.1. Unconditional and conditional effective approaches

Next, we examine the impacts of $\bar{T}'\bar{\nabla}\bar{\phi}'$ and $\bar{S}'\bar{\phi}'$ in stochastic modeling approaches on the unconditional or conditional estimates of T and S fields, which yield the most likely unconditional or conditional ϕ field (i.e., ensemble mean field).

3.1.1. Unconditional approach

Suppose $T(x)$ and $S(x)$ of the aquifer are entirely unknown, and we adopt the equivalent homogeneous media (EHM) model to predict the unconditional ensemble mean of ϕ . Then, Eq. (8) can be rewritten as:

$$T_{\text{eff}}\nabla^2\bar{\phi} = i\omega S_{\text{eff}}\bar{\phi} + Q_0 \quad (9)$$

in which $T_{\text{eff}} = \bar{T} + \nabla \cdot (\bar{T}'\bar{\nabla}\bar{\phi}')[\nabla^2\bar{\phi}]^{-1}$ and $S_{\text{eff}} = \bar{S} + \bar{S}'\bar{\phi}'[\bar{\phi}]^{-1}$ are the unconditional effective parameters. Here, the effective parameters are assumed to be invariant in ensemble and space, while some theoretical analyses (e.g., Indelman, 2003) showed that they might vary with radial distance near the pumping well and then approach some constant values at large radial distances. Tartakovsky and Neuman (1998) investigated the effect of temporal nonlocality on one- and three-dimensional mean flows in infinite, statistically homogeneous media. Rabinovich et al. (2013) found that the unconditional effective T derived from the periodic pumping test is frequency-dependent and varying with the variance of T . This means that additional term (i.e., $\nabla \cdot (\bar{T}'\bar{\nabla}\bar{\phi}')[\nabla^2\bar{\phi}]^{-1}$) in the effective parameter varies with the frequency of pumping rates. These effective parameters yield the most likely averaged ϕ field over all possible heterogeneity (the unconditional $\bar{\phi}$ field) under the given periodic pumping discharge. The deviation from its unconditional mean ($\bar{\phi}$)—the uncertainty and spatial variability of ϕ at any location—is quantified by the unconditional variance ϕ .

The application of the effective parameters to a field condition requires the ergodicity assumption. That is, this unconditional mean field $\bar{\phi}$ in an aquifer should reproduce the general pattern of observed ϕ values in the least squared sense but does not honor the observed ϕ value at each observation location (see Wu et al., 2005; Yeh et al., 2015a).

3.1.2. Conditional approach

The conditional approach adopts the highly parameterized conceptual model and requires that the conditional T , S , and mean ϕ fields reproduce measured T , S , and ϕ values at the sampling locations (Yeh et al., 1996). Three possible conditional scenarios are discussed below.

Scenario a) If $T(x)$ and $S(x)$ at every location are given precisely without any uncertainty, Eq. (8) reduces to

$$\nabla \cdot (\bar{T}_c(x)\nabla\bar{\phi}_c) = i\omega\bar{S}_c(x)\bar{\phi}_c + Q_0 \quad (10)$$

The subscript c denotes conditional, and the overhead bar again represents the expectation. In this case, the conditional mean fields are identical to the given realization (i.e., $\bar{T}_c(x) = T(x)$ and $\bar{S}_c(x) = S(x)$). Since T and S are fully specified, the exact ϕ is equal to the fully conditioned mean head field $\bar{\phi}_c$ and can be solved uniquely with the given initial and boundary conditions. Notice that $\bar{T}'\bar{\nabla}\bar{\phi}'$ and $\bar{S}'\bar{\phi}'$ vanish

in this equation.

Scenario b) This scenario represents the case where the ϕ value at every location is known precisely, but T and S fields are entirely unknown. Under this circumstance, the following conditioned mean flow equation depicts the ϕ fields:

$$\begin{aligned} \nabla \left[(\bar{T} + T') \nabla \bar{\phi}_c \right] &= i\omega (\bar{S} + S') \bar{\phi}_c + Q_0 \text{ or} \\ \nabla \left[\bar{T} \nabla \bar{\phi}_c \right] - i\omega \bar{S} \bar{\phi}_c + \nabla \left[T' \nabla \bar{\phi}_c \right] - i\omega S' \bar{\phi}_c &= Q_0 \end{aligned} \quad (11)$$

Notice that $\bar{\phi}_c$ is a fully conditioned mean or the true ϕ field with zero variance (i.e., uncertainty). \bar{T} and \bar{S} are unconditional mean values, which are known and constant in space and the ensemble, while T' and S' are unknown perturbations. In Eq. (11), the first two terms are the known net flux at each location resulting from the known \bar{T} and \bar{S} , and the known $\bar{\phi}_c$. The second two terms denote the unknown net flux and storage perturbations at each location due to the unknown T' and S' , and the known $\bar{\phi}_c$ field.

Although $\bar{T}'\bar{\nabla}\bar{\phi}'$ and $\bar{S}'\bar{\phi}'$ disappear in Eq. (11), estimation of T' and S' remains an under-determined inverse problem due to the unknown spatial distribution of the flux induced by Q_0 . That is, many possible T' and S' pairs lead to the same head field unless T' and S' or the spatial distribution of flux along the boundaries are specified (Mao et al., 2013a; Yeh et al., 2015a, 2015b). Because all the variables associated with frequency ω (i.e., the first two terms in Eq. (11), $\bar{\phi}_c, \omega$) and Q_0 are specified, the system of equations to be solved is independent of frequency and so are the estimated T' and S' .

Scenario c) Suppose ϕ , T , and S are known exactly at some (not all) locations in the aquifer. The terms $\bar{T}'\bar{\nabla}\bar{\phi}'$ and $\bar{S}'\bar{\phi}'$ in Eq. (8) are zero at locations where T and S are specified, and at others are non-zero. The residual head variances likewise are zero at locations where heads are known, but the residual T and S variances at these head locations are non-zero, although they become small (e.g., p.286 in Cheng et al., 2019). Therefore, the ensemble mean equation becomes

$$\nabla \cdot (T_{\text{eff}}(x)\nabla\bar{\phi}_c) = i\omega S_{\text{eff}}(x)\bar{\phi}_c + Q_0 \quad (12)$$

$T_{\text{eff}}(x)$ and $S_{\text{eff}}(x)$ are the conditional effective parameters, which honor the sample values at the sample locations. These parameters under the given flow conditions yield the conditional mean head fields $\bar{\phi}_c$ that reproduce the observed heads at the observation wells (Yeh et al., 1996). These effective parameters are the sum of their unconditional means (i.e., \bar{T} and \bar{S}) and the contributions from the non-zero $\bar{T}'\bar{\nabla}\bar{\phi}'$ and $\bar{S}'\bar{\phi}'$ terms after conditioning with sparse measurements of ϕ , T , and S . Since the variable ϕ' ingrained in these non-zero terms vary with the frequency, the conditional effective parameters are frequency dependent.

Because of the frequency-dependent nature of the conditional effective parameters, some studies (as discussed in the introduction) found that a pumping test with a multifrequency rate could lead to a more detailed characterization of the heterogeneity than that with a constant rate or single frequency. The expectation nature of the two non-zero terms (conditional cross-covariance) in the conditional mean flow equation may explain these findings. As we have remarked in Scenarios a) and b), these terms become small or vanish only if T' , S' , or ϕ' are specified at some locations. That is, without adding additional observation wells or parameter measurements, manipulating pumping rates with different frequencies does not remove these cross-covariance terms but changes their patterns (as reflected in the sensitivity maps in Figs. 6 and 7 of Cardiff et al., 2013). Nevertheless, these terms are still ensemble-averaged relationships between many possible T' , S' , and ϕ' fields. That is, many combinations of these fields can lead to the same cross-covariance pattern. During the inverse modeling using data from multifrequency pumping rates, these cross-covariance terms remain

unresolved, and many possible solutions exist. The previous studies claiming improvements in the estimates are likely due to the simple heterogeneity pattern considered in their studies (e.g., Ahn and Horne, 2010), which only have a limited number of possibilities. Consequently, adding additional frequency information may lead the estimates to a similar but not exact heterogeneous pattern. As the number of the unknowns becomes large and the pattern becomes complex, the possibility of obtaining a close estimate by using a multifrequency would be as slight as using single frequency data.

4. Conditional effective parameter Estimator: SimSLE

To substantiate our points in section 3.1.2 Scenario c, we use a conditional effective parameter estimator to conduct numerical experiments in section 5. The SimSLE (Simultaneous Successive Linear Estimator, Xiang et al. (2009)) is an extension of SLE (Successive Linear Estimator, Yeh et al. (1995, 1996)). They are parameter estimation algorithms built upon the stochastic conditional expectation concept (Priestley, 1979) and the hydrogeology cokriging concept (Kitanidis and Vomvoris, 1983; Dagan, 1985). The algorithm also considers the nonlinear relationship between the state variable (e.g., water level) and the parameters (Kitanidis, 1995; Yeh et al., 1995). In summary, it is a stochastic nonlinear estimator that derives the conditional effective parameter fields with given locally measured parameter values and observed heads, and yields conditional mean and covariance of head fields (as discussed in the section 3.1.2 scenario c and (Gao et al., 2021)).

Recognizing the non-redundant information collected by tomographic surveys (as elucidated in Wen et al., 2019), Yeh and Liu (2000) developed the 3-D steady-state hydraulic tomography and Zhu and Yeh (2005) built the 3-D transient hydraulic tomography based on the SLE to further estimate the conditional effective parameters. Over the past two decades, SimSLE has become a mature inverse model for hydraulic tomography in heterogeneous porous media as well as fractured rocks. Recently, Cheng et al. (2019) proposed a linear cokriging technique for oscillatory hydraulic tomography. A detailed description of SimSLE is available in many publications. Below is a brief description of the algorithm for this study.

Mathematically, SimSLE is similar to the Kalman filter, developed for time-domain stochastic processes. It iteratively determines the conditional effective parameter fields ($T_{\text{eff}}(\mathbf{x})$ and $S_{\text{eff}}(\mathbf{x})$ in section 3.1.2. scenario c) using the linear estimator

$$\hat{\mathbf{f}}^{(r+1)} = \hat{\mathbf{f}}^{(r)} + \mathbf{\epsilon}_{\text{ff}}^{(r)} \mathbf{J}_{\text{fh}}^{(r)} [\mathbf{J}_{\text{fh}}^{(r)T} \mathbf{\epsilon}_{\text{ff}}^{(r)} \mathbf{J}_{\text{fh}}^{(r)} + \mu \theta]^{-1} [\mathbf{h}^* - \mathbf{h}^{(r)}] \quad (13)$$

in which $\hat{\mathbf{f}}^{(r)}$ is an $m \times 1$ vector (m is the total number of elements), consisting of perturbations of the conditionally estimated effective $\ln T$ or $\ln S$, and the superscript r is the iteration index. It represents the estimate at iteration r minus the unconditional mean $\ln T$ or $\ln S$. Notice that \ln denotes the natural logarithm transformation, which is employed to avoid the estimates' negative values. $\mathbf{h}^{(r)}$ ($n \times 1$, n is the total number of observations) is the simulated head in the time or frequency domain, depending on the time- or frequency domain approach. It is simulated by Eq. (12) with the newly estimated parameter fields (i.e., adding $\hat{\mathbf{f}}^{(r)}$ to its unconditional mean values and transforming back to $T_{\text{eff}}^{(r)}(\mathbf{x})$ and $S_{\text{eff}}^{(r)}(\mathbf{x})$). \mathbf{h}^* ($n \times 1$) is the observed head. The term, $\mathbf{J}_{\text{fh}}^{(r)}$ ($m \times n$), is the sensitivity of head at a given location in the aquifer with respect to the change in $\ln T$ or $\ln S$ at any location in the aquifer. The sensitivity matrix is evaluated by the adjoint state approach [e.g., Sykes et al., 1985; Sun and Yeh, 1990] using $T_{\text{eff}}^{(r)}(\mathbf{x})$ and $S_{\text{eff}}^{(r)}(\mathbf{x})$. $\mathbf{\epsilon}_{\text{ff}}^{(r)}$ ($m \times m$) is the covariance of $\ln T$ or $\ln S$. μ is a dynamic stabilizer and $\theta = \text{dia}(\mathbf{J}_{\text{fh}}^{(r)T} \mathbf{\epsilon}_{\text{ff}}^{(r)} \mathbf{J}_{\text{fh}}^{(r)}) + \text{Max}[\text{dia}(\mathbf{J}_{\text{fh}}^{(r)T} \mathbf{\epsilon}_{\text{ff}}^{(r)} \mathbf{J}_{\text{fh}}^{(r)})]$. $\text{Max}[\cdot]$ is the maximum value. When $r = 0$, $\hat{\mathbf{f}}^{(0)}$ is zero and $\mathbf{\epsilon}_{\text{ff}}^{(0)}$ ($m \times m$) is the unconditional covariance. After the first linear estimation, the unconditional covariance becomes the residual covariance of $\hat{\mathbf{f}}^{(r+1)}$ and is

updated by

$$\mathbf{\epsilon}_{\text{ff}}^{(r+1)} = \mathbf{\epsilon}_{\text{ff}}^{(r)} - \mathbf{\epsilon}_{\text{ff}}^{(r)} \mathbf{J}_{\text{fh}}^{(r)} [\mathbf{J}_{\text{fh}}^{(r)T} \mathbf{\epsilon}_{\text{ff}}^{(r)} \mathbf{J}_{\text{fh}}^{(r)} + \mu \theta]^{-1} \mathbf{J}_{\text{fh}}^{(r)T} \mathbf{\epsilon}_{\text{ff}}^{(r)} \quad (14)$$

The diagonal term of the residual covariance matrix (i.e., residual variance) represents the remaining uncertainty of the estimated effective $\ln T$ or $\ln S$ at every location after the head information is included. Albeit the residual variance is an approximation, a small value indicates that the estimated effective $\ln T$ or $\ln S$ likely is close to the true fields. A large value indicates the uncertainty is not resolved. Again, this uncertainty is an ensemble statistic and is not equal to the actual difference between the true and estimate.

The iterative update is terminated if the mean squared error of simulated and observed heads is smaller than a specified tolerance (e.g., 10^{-7}), or the increase of the spatial variance of the estimated parameter becomes insignificant between successive iterations. That is, the algorithm is unable to extract further heterogeneity information contained in the observed heads.

5. Monte Carlo numerical experiments

OHT attempts to solve the under-determined inverse problems (scenario c with conditioning only on the head), which have many possible solutions. Utilizing heads induced by excitations with different frequencies for inverse modeling may result in accidentally superior or poor estimates. Monte Carlo (MC) experiments, therefore, are necessary to verify the OHT's reliability.

The simulation domain of the MC experiments is a 2-D horizontal confined aquifer consisting of 40×40 square elements. Each element is $30 \text{ m} \times 30 \text{ m}$. The aquifer is bounded by a constant head boundary of 30 m . The initial head is 30 m everywhere. In this experiment, we solve the time-domain governing flow equations (Eqs. (1) and (2)) and use 32-time steps to represent a single periodic cycle to ensure the results' accuracy. Two hundred realizations of random fields, with given means of T and S ($0.01 \text{ m}^2/\text{min}$ and 10^{-5}), variances of $\ln T$ and $\ln S$ (2 and 0.5), and correlation lengths ($\lambda_x = \lambda_y = 400 \text{ m}$), are generated using a spectral method (Gutjahr, 1989; Robin et al., 1993). Nine wells are installed in the aquifer (circles in Fig. 1). An OHT using four pumping wells (square in Fig. 1) with periodic pumping rates of four periods ($\omega^{-1} = 15, 20, 50$, and 100 min/cycle) is used in each realization. That is, each realization includes 16 individual pumping tests (4 frequencies times 4 pumping wells).

Afterward, the simulated head responses (32 timesteps of the last cycle) at monitoring wells after the flow reaches quasi-steady are utilized for the inversion. This analysis is conducted in the time domain. The analysis based on the frequency domain is presented in the later section. The prior information for SimSLE consists of the same means, variances, and correlation lengths as those used to generate the random field. We use the same geostatistical parameters as the generated

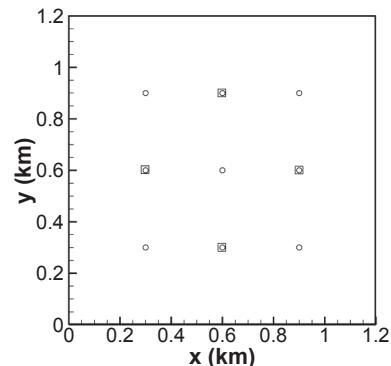


Fig. 1. Synthetic aquifer. Circles are the observation wells and the squares represent pumping wells. Four boundaries are the constant head.

random fields to focus on different pumping frequencies' effects on the estimate. The computation is implemented in a variably saturated flow finite element model VSAFT2 (Yeh et al., 1993), available at <http://tian.hwr.arizona.edu/downloads>, to estimate T and S with observed heads under different frequencies and multifrequency. We use the coefficient of determination (R^2) and the mean squared error (L_2 norm) as the performance metrics to evaluate each realization's estimates.

T Estimates. Fig. 2a displays the mean and standard deviation of R^2 and L_2 between the reference and estimated T fields resulting from 200 realizations based on different single frequencies and those of the combination of all frequencies. The left-side axis of the plot is for R^2 and the right-side axis is for L_2 .

As displayed in the figure, large standard deviations of R^2 for different frequencies suggest that the OHT survey's performance could vary significantly from one realization to another. Interestingly, the mean and standard deviation of R^2 and L_2 of the 200 realizations are almost identical despite the frequency ω . In the multifrequency case, the mean R^2 is 0.66, with a standard deviation of 0.11; the mean L_2 is 0.030, with a standard deviation of 0.016. These values are similar to those cases with a single frequency. This result indicates that using multiple frequencies in OHT does not improve the estimates averaged over the 200 realizations.

S Estimates. Performance metrics of estimated S fields in Fig. 2b show that the mean and standard deviation of R^2 and L_2 are nearly identical regardless of the pumping rate frequency. OHT using the multifrequency yields 0.57 for the mean R^2 with 0.12 for the standard

deviation. The mean L_2 is 2.6×10^{-5} , and the standard deviation is 6.9×10^{-6} . Again, using the multifrequency in OHT does not improve the S estimates.

A Two-Sided T-Test. To further support our conclusion above, we carry out a two-sided t -test with 95% confidence interval to certify if the difference in mean values between the multiple and single frequencies is significant or not. Table 1 lists the t -values corresponding to the results in Fig. 2a and b. If the t value is greater than 1.97 or less than -1.97, the difference is significant. In the table, the values labeled by * indicate the difference is significant (i.e., the multifrequency improves the estimate). The values without * indicate the difference is not significant (i.e., the multifrequency does not improve the estimate). The t -values of L_2 and R in Table 1 show that T 's estimate from multifrequency data does not significantly differ from those from the single frequency. In contrast, the estimate of S may slightly differ from those from the single frequency.

Number of Realizations. A sufficient number of realizations is a must to ensure the representativeness of MC experiment's results. Fig. 3a and b show the means and standard deviations of R^2 and L_2 as a function of the number of realizations. They confirm that 200 realizations are sufficiently large to support our findings. In other words, the MC results certify that in many trials, manipulating pumping frequency does not reduce the conditional residual flux ($T' \nabla \phi'$) and residual storage ($S' \phi'$), as indicated by the nearly constant means and large standard deviations of R^2 and L_2 in Fig. 2. Specifically, had the estimates been improved (close to the correct field), the residual flux and storage terms should have diminished so should the means and standard deviations. As discussed in section 3.1.2, we emphasize that additional non-redundant information from additional pumping or observation wells (Wen et al., 2019; Wang et al., 2019) guarantees the estimate's improvement.

6. Single realization

We also present the results based on a single realization below. The reference T and S fields and the wells for this case are displayed in Fig. 4a and b.

Estimated T . The estimated T fields using the OHT with the pumping rate of the periods (15, 20, 50, 100, and the combination of all the periods) are displayed in Fig. 5a, b, c, d, and e, respectively. In Fig. 5f, g, h, i, and j, we show the scatter plots of the reference field against the estimated, based on OHT with different periods. Also included in these figures are the regression lines' slope and intercept, the correlation between the estimated and reference $\ln T$ (R^2), and the mean squared error of $\ln T$ (L_2).

According to these figures, OHT using different single or multifrequency excitations captures the general pattern of heterogeneity. Excitations of different frequencies improve the estimates at some locations and worsen them at others. There is no clear winner or loser for characterizing the detailed parameter variations.

Intuitively, a low-frequency (long period) excitation could cover a greater area than the high frequencies; high-frequency (short period) excitations could map details near the well field. Consequently, a multifrequency (a mixture of different periods) excitation, combining various frequencies' merits, could improve the overall estimate. Nevertheless, the scatter plots of this realization suggest otherwise. In particular, the slope, intercept, R^2 , and L_2 values indicate that the

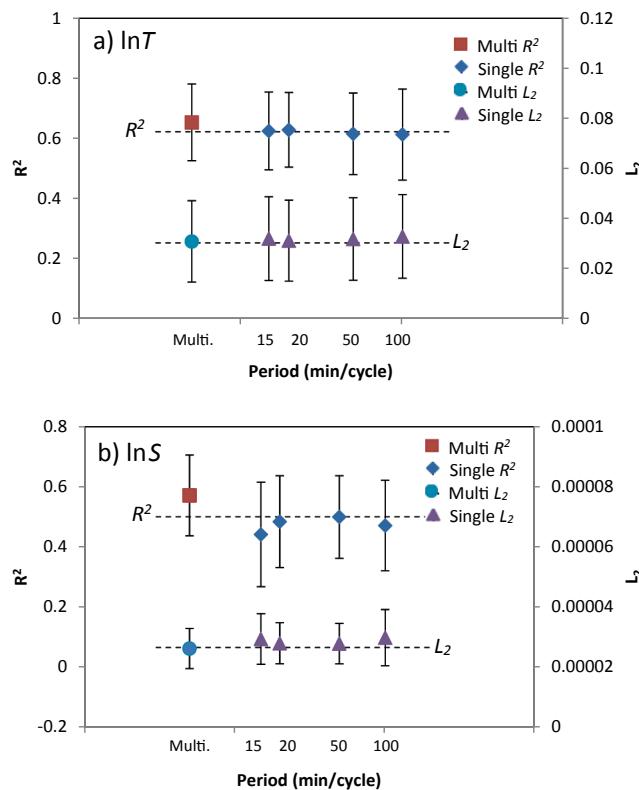


Fig. 2. The mean performance metrics (R^2 and L_2) of estimated $\ln T$ and $\ln S$ using pumping rates with different frequencies (or periods, $\omega^{-1} = 15, 20, 50$, and 100 (min/cycle)) and with multiple frequencies (i.e., all of the different frequencies). The vertical bar represents the mean metrics \pm one standard deviation of the metrics. Mean R^2 of results based on different frequencies are labeled with blue diamonds and that of the multiple frequencies is a red square. Mean L_2 of results based on different frequencies are labeled with purple triangles and that of multiple frequencies is a blue circle. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1

t values for the performance metrics for $\ln T$ and $\ln S$ estimates using excitations with different frequencies. * indicates significant (The performance of multifrequency is different than the single frequency).

Period	15	20	50	100
$L_2 T$ estimates	0.94	0.26	0.84	1.71
$R T$ estimates	0.70	0.61	0.91	0.98
$L_2 S$ estimates	5.97*	3.70*	3.48*	6.39*
$R S$ estimates	2.61*	1.79	1.49	2.04*

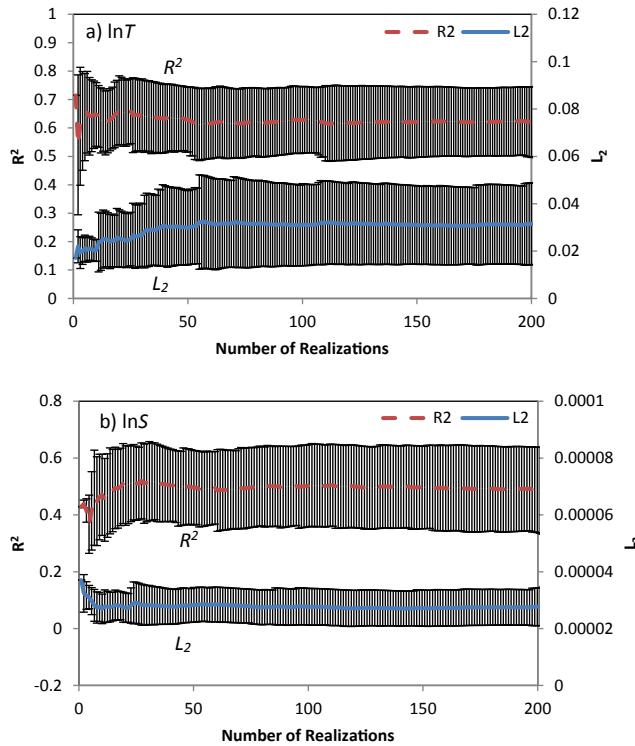


Fig. 3. The mean performance metrics (R^2 and L_2 , the red dashed line and the blue line, respectively) of estimated $\ln T$ and $\ln S$ as a function of the number of realization for the case where the period of the pumping rate = 20 (min/cycle). The vertical bar represents the mean of the performance metrics \pm one standard deviation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

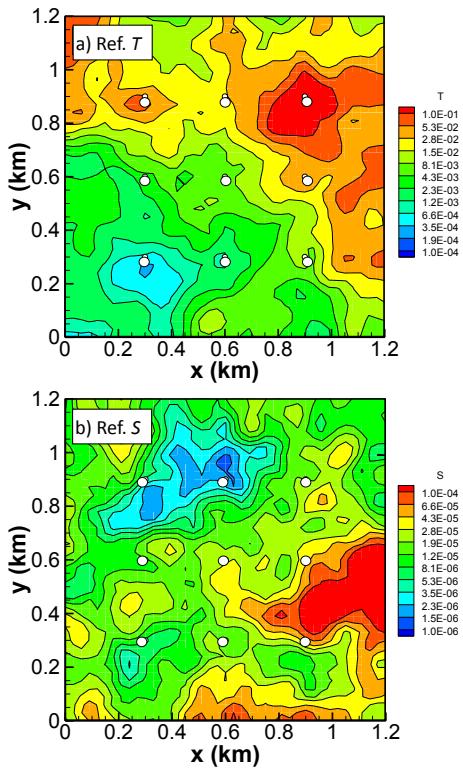


Fig. 4. The reference hydraulic transmissivity T (m^2/min) and storage coefficient S (-) fields of one realization. White circles are the observation wells.

estimate based on a period of 15 (a high frequency) is least satisfactory, and based on a period of 20 is the best. The scatter plots also indicate that the estimate based on all frequencies (multifrequency, Fig. 5j) is worse than that in Fig. 5i based on a long period of 100 (a low frequency). The results of this realization contradict the intuition. However, they corroborate the large standard deviations in R^2 and L_2 values in Fig. 2. They further demand MC experiments to assess the reliability of the OHT estimates for ill-defined inverse problems.

Estimated S . Plots similar to those in Fig. 5 are presented in Fig. 6 a through j for S estimates. Akin to the results in Fig. 5, OHTs with different frequencies create different local-scale anomalies but derive a similar general spatial distribution. Fig. 6f shows that excitation with a period of 15 yields the least acceptable estimate. The slope, intercept, R^2 , and L_2 values in Fig. 6j from multifrequency excitation are superior to those with different single frequencies. These results seem to follow our intuition but contradict the previous findings based on the T estimates. This contradiction may be because T and S behave differently and interact with each other. Such conflicting T and S estimates' behaviors highlight the uncertainty in results from OHT with excitations of different frequencies in a single realization. The MC simulations thus are deemed to be appropriate.

7. Effects of domain size and Time-Domain solution

To show that the results in the above examples are independent of the domain size, the correlation scale, and the solution technique (i.e., time- or frequency-domain solution), we present an additional MC example, solved by the frequency-domain equations (eq. (3)).

The new aquifer is $98 \text{ m} \times 98 \text{ m}$, consisting of 70×70 rectangular elements; the grid spacing varies, as shown in Fig. 7. The OHT uses nine wells (Fig. 7), and the horizontal and vertical spacings between the two wells are 5 m. The spacing is 9 times smaller than the size of the entire domain. The width of the element is 0.5 m near the well and 2 m in the remote region. A zero-amplitude boundary bounds the aquifer. For this MC simulation, six hundred realizations of a random T field with a mean of $T = 10^{-4} \text{ m}^2/\text{s}$ and a variance of 1 in $\ln T$ and correlation lengths are $\lambda_x = \lambda_y = 5 \text{ m}$. The S field is homogeneous with a value of 10^{-4} , as in previous OHT analyses by others.

The OHT surveys employ three sequential pumping tests at three wells (square in Fig. 7) with periodic pumping rates of four periods ($\omega^{-1} = 400, 1600, 3600$, and 10800 s/cycle). As a result, 12 individual pumping tests (4 frequencies times 3 pumping wells) are conducted for each realization. The numerical accuracy limits the selection of the lower periods. The radius of influence corresponding to each period is 20, 40, 60, and 104 m, based on the formula (i.e., the distance of signal propagation is proportional to $\sqrt{T/\omega}$) in Cardiff and Barrash (2015). During the test at a selected well, the other eight wells' responses are collected for inverse modeling.

We then use VSAFT2 to simulate the OHT surveys by solving the frequency domain's governing equation (Eq. (3)). The aggregation-based algebraic multigrid method (Notay, 2010, 2012; Notay and Napov, 2012). Similarly, we derive the sensitivity using the adjoint method in the frequency domain. Afterward, SimSLE in VSAFT2 estimates the T field over the entire domain. We use the same means, variances, and correlation lengths as those used to generate the random field as the prior information of SimSLE to focus on the effect of different pumping frequencies on the estimate. Following previous studies that claim improvements due to multifrequency tests, we treat the S field as known and estimate the T field only. Also, we evaluate T estimates in the near-field (the area within 15 m from the center of the domain, or 3 times of the correlation length) and the entire domain.

Near-Field Estimates. We summarize in Fig. 8 the mean and standard deviation of R^2 and L_2 of the near-field T estimates of the MC experiments at different frequencies. Once again, we observe large standard deviations of R^2 and L_2 for different frequencies, indicative of

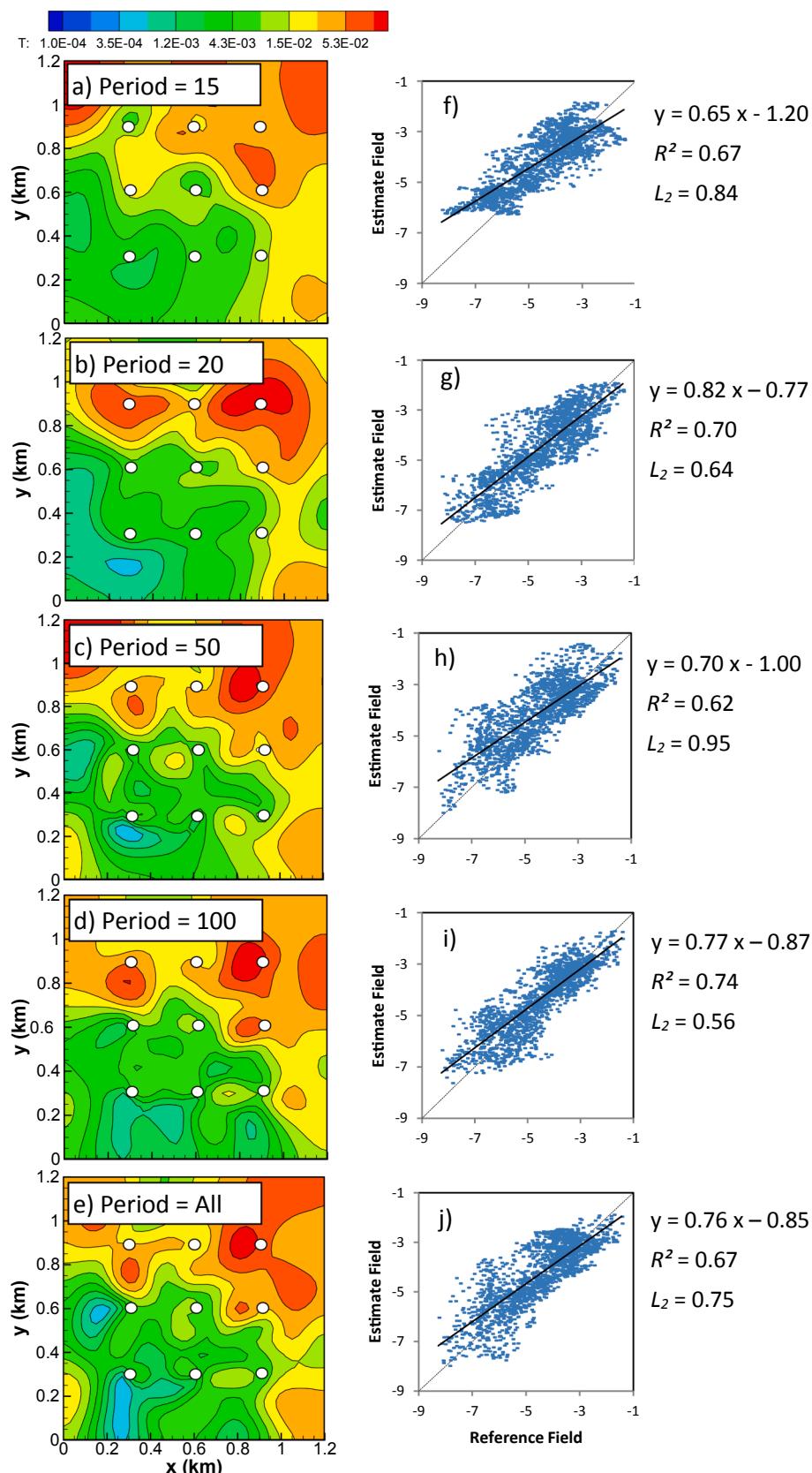


Fig. 5. a-e) are the estimated $\ln T$ (m^2/min) fields using different and multiple frequencies. The white circles are the observation wells. f-j) are histograms of the estimates (the blue vertical bars) and the reference field (the black line). The linear regression equation, R^2 , and L_2 are the performance metrics of the estimated and reference $\ln T$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

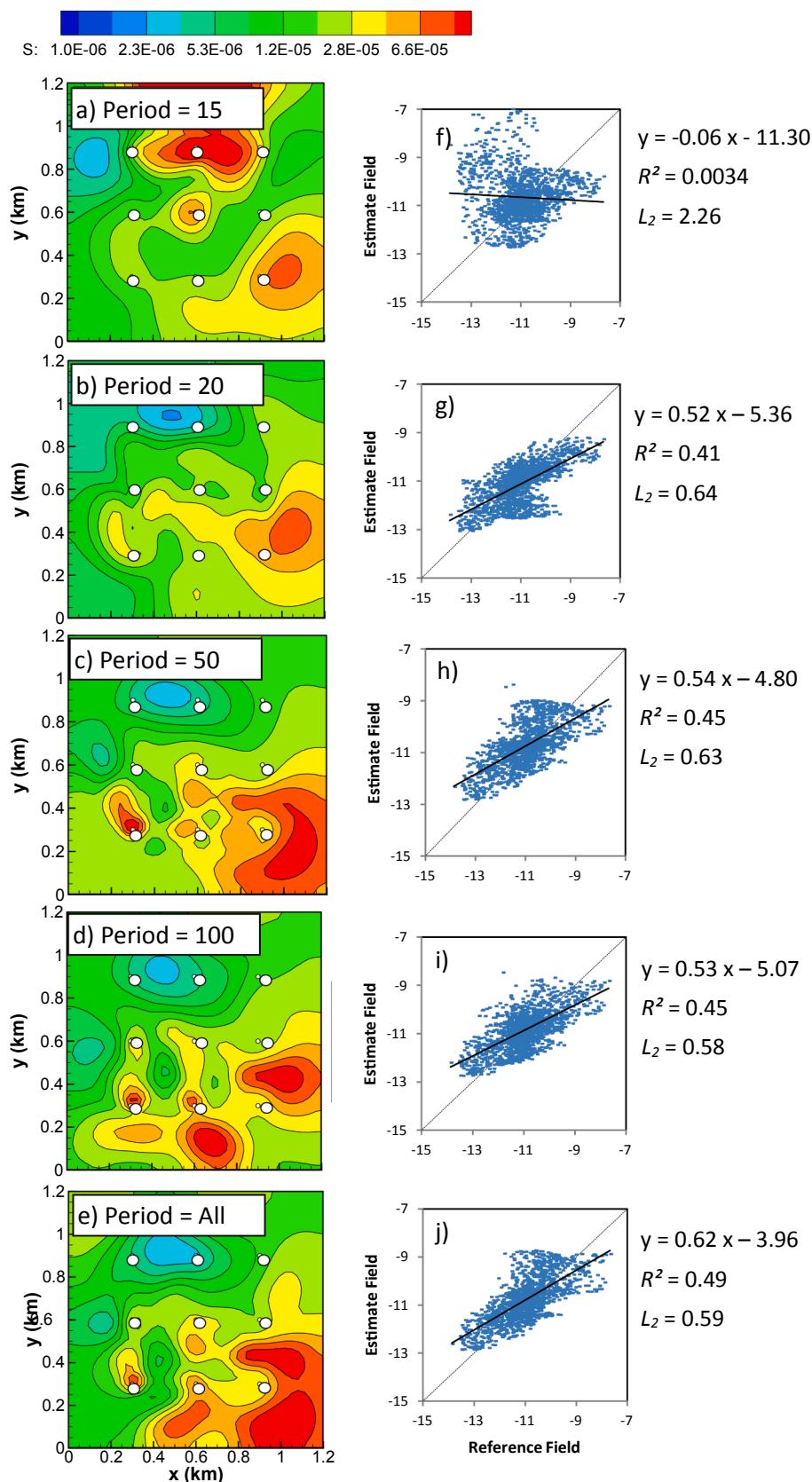


Fig. 6. a-e) are the estimated lnS (–) fields using different and multiple frequencies. The white circles are the observation wells. f-j) are histograms of the estimates (the blue vertical bars) and the reference field (the black line). The linear regression equation, R^2 , and L_2 are the performance metrics of the estimated and reference lnS. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

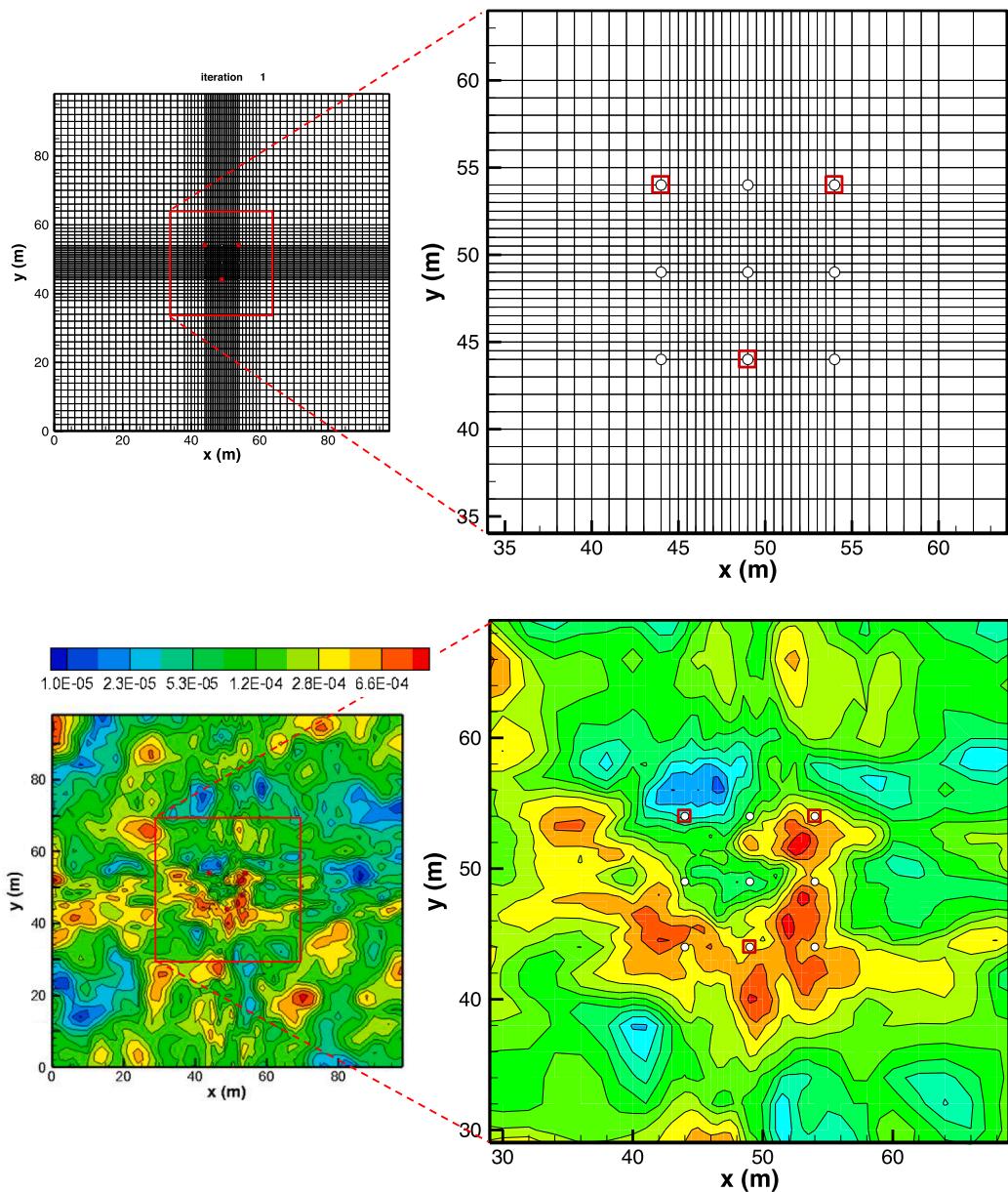


Fig. 7. Synthetic aquifer with large simulation domain. White circles are the observation wells and the squares represent pumping wells. Four boundaries are the constant head. The reference hydraulic transmissivity T (m^2/s) field of one realization.

large variability of the OHT performance between different realizations. The mean and standard deviation of R^2 and L_2 remain almost identical despite the pumping rate frequency. Based on t -test (Table 2), we find that the differences in mean R and L_2 values between the estimates using different single frequencies and the combination of all frequencies are statistically insignificant (i.e., no consistency in the test). These results suggest that using multiple frequencies in OHT does not necessarily improve the estimation on average.

For a single realization, the near-field's estimates are displayed in Fig. 9a to e. They are for periods 400, 1600, 3600, 10800, and the combination, respectively. Their scatter plots are illustrated in Fig. 9f through j. Visual comparisons of Fig. 9a through e with the reference field in Fig. 7b reveal that estimates using different single frequencies and multifrequency capture the reference field's general features. However, the slope, intercept, R^2 and L_2 values of the scatter plots indicate that the estimates based on period 1600 and all periods are the best, although the former's slope and intercept are better than the latter.

Also, notice that from period 10,800 is the worst. Again, the estimate using multifrequency (Fig. 9j) is not superior to others in this single realization.

Entire Domain Estimates. Examining the MC estimates over the entire domain, we find that the mean and standard deviation of R^2 and L_2 of the estimates using the multifrequency pumping rate seems slightly better than the single frequency test (Fig. 10). Nevertheless, a t -test (Table 3) indicates that the differences in mean R values between the estimates using different single frequencies, and the combination of all frequencies are statistically insignificant (i.e., no consistency in the test). However, the differences in mean L_2 values are statistically significant. These results suggest that the averaged improvements in the estimates using multiple frequency OHT over many possible cases is not significant.

Comparing Figs. 10 and 8 (near-field estimates), we notice that the values of R^2 of all frequencies have dropped to around 0.2. Meanwhile, the values of L_2 have increased to 0.7. That is, performance metrics, R^2

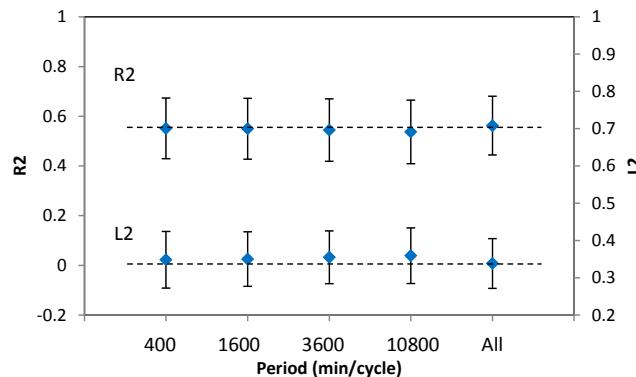


Fig. 8. The mean performance metrics (R^2 and L^2) of estimated $\ln T$ using pumping rates with different frequencies (or periods, $1 \omega^{-1} = 400, 1600, 3600$, and $10,800$ [s/cycle]) and with multiple frequencies (i.e., all of the different frequencies) in the large simulation domain. The vertical bar represents the mean metrics \pm one standard deviation of the metrics.

Table 2

t values for the performance metrics for $\ln T$ estimates using excitations with different frequencies. * indicates significant (The performance of multifrequency is different than the single frequency).

Period	400	1600	3600	10,800
L_2	1.64	5.96*	11.46*	15.58*
R	-0.0044	0.544	1.26	1.76

and L_2 values, are affected by the domain size since they are the statistics over the entire domain, in which the uncertain estimates at far-field dominate the overall statistics. However, these values' differences between different frequencies remain similar for the near-field and the entire domain.

The results based on a single realization are discussed below. The reference T field is displayed in Fig. 7. The estimated T fields using the head in the frequency domain and the pumping rate with different periods (400, 1600, 3600, 10800, and the combination of all the periods) are illustrated in Fig. 11a,b,c,d, and e, respectively. The reference field's scatterplots and the estimates, using different periods, are plotted in Fig. 11f, g, h, i, and j.

Also included in these figures are the performance metrics (i.e., R^2 , L_2 , and slope and intercept of the linear relationship between the estimated and reference $\ln T$). These performance metrics indicate that the estimate from period 400 is the least acceptable, and from all periods is the best. It has slightly better values of performance metrics than the cases with periods 1600 and 3600. From these plots, it is also clear that the excitations with a period of 400 cannot resolve heterogeneity beyond the near-field. This finding is based on the horizontal line at $\ln T$ value of -9 on the estimated field axis. This line indicates that the estimates remain as the initial mean $\ln T$ —no improvement due to a rapid decade of the excitations' strength. As the period increases, this horizontal line becomes smeared, indicating more heterogeneity at far-field is detected, although imprecisely. Again, the domain size affects performance metrics, which are statistics for the entire domain.

The small and large domain OHT analyses suggest that the results from a single realization (or experiment) could be inconclusive. Besides, the effect of boundary conditions does not influence the estimated parameter values near the well field. This result is attributed to the fact that both forward and inverse simulations employ the same boundary conditions. For the effects of unknown boundary conditions on HT, we refer it to Liu et al. (2020), Daranond et al. (2020) and Sun et al. (2013).

Discussion. A revisit of the well-known first-order analysis of the head variation in the time domain at a location (Gelhar, 1993; Wu et al., 2005; Sun et al., 2013) should elucidate the results mentioned above.

Note that the first-order analysis has been proved valid for the variance of $\ln T$ even greater than 1 (highly nonlinear problems) and has been widely adopted (e.g., Gelhar, 1993). Specifically, under given boundary and initial conditions and known stress, simulated head perturbation (deviation from the mean head) at any location in a heterogeneous aquifer can be approximated as a sum of the product of the head sensitivity to and the magnitude of the heterogeneity (T perturbation around its mean) at every aquifer location. Consequently, a large T perturbation at a far distance where the sensitivity is low could significantly impact the head at a given location. Similarly, a small T perturbation at a high sensitivity area may have little impact on the head at that location. In other words, knowing the sensitivity distribution alone (without knowing T perturbation distribution), one cannot determine the head perturbation at a location.

Inverse modeling seeks a T field that can produce the head perturbation at observation wells. The explanation above suggests that even if the sensitivity distribution due to excitation of a given frequency is known, an infinite number of T perturbation fields remain, yielding the same head perturbation at an observation location. As illustrated in Figs. 6 and 7 in Cardiff et al. (2013), different frequencies' sensitivity maps are different but maintain similar patterns. A multifrequency sensitivity map would maintain a similar shape but be dominated by the low frequency, even though some locations' sensitivity values may differ. Specifically, the sensitivity maps of different frequencies are highly correlated.

Inverting a multifrequency test would include additional head observations at different frequencies at the same wells, but these observations are highly correlated. In other words, the multifrequency test generates a system of nonlinear and mutually dependent (or correlated) equations. Because of their mutual dependence, many possible heterogeneous T fields still could lead to the head values at the observation location. Minor differences in heads and sensitivity values due to numerical or measurement errors may produce additional T anomalies comparing with a single frequency OHT. These anomalies may be acceptable or unacceptable (see Figs. 6 and 9): uncertain estimates. Such uncertain (fluctuating) results are also vivid in RMSE values for cases 1

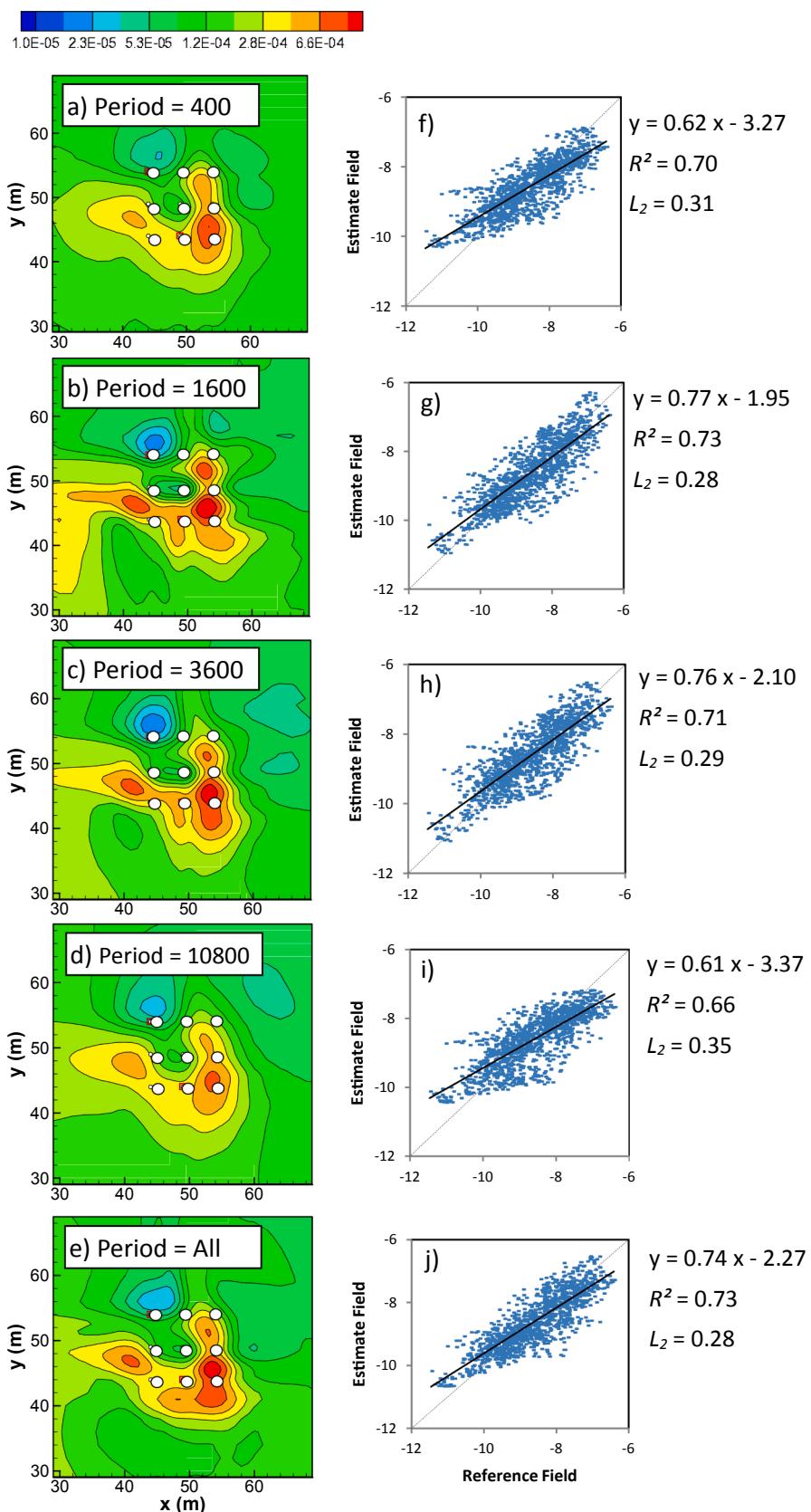


Fig. 9. a-e) are the estimated $\ln T$ (m^2/s) fields using different and multiple frequencies in the large simulation domain. The white circles are the observation wells. f-j) are histograms of the estimates (the blue vertical bars) and the reference field (the black line). The linear regression equation, R^2 , and L_2 are the performance metrics of the estimated and reference $\ln T$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

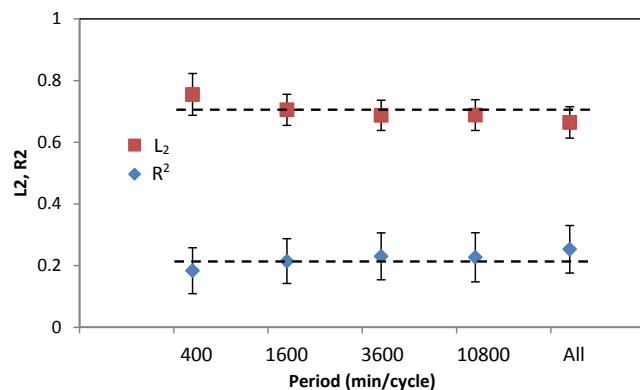


Fig. 10. The mean performance metrics (R^2 and L_2) of estimated lnT using pumping rates with different frequencies (or periods, $1 \omega^{-1} = 400, 1600, 3600$, and $10,800$ [s/cycle]) and with multiple frequencies (i.e., all of the different frequencies) in the large simulation domain. The vertical bar represents the mean metrics \pm one standard deviation of the metrics.

Table 3

t values for the performance metrics for lnT estimates using excitations with different frequencies. * indicates significant (The performance of multifrequency is different than the single frequency).

Period	400	1600	3600	10,800
L_2	37.5*	19.69*	11.27*	11.57*
R	2.33*	1.26	0.75	0.86

through 13 in Table 1 of Zhou et al. (2016). Due to this uncertainty, MC simulation is appropriate to ensure that the multifrequency estimates' improvement holds over many possible heterogeneous aquifers (not just a coincidence). Our MC simulation results and statistical tests indicate that a multifrequency estimate is statistically similar to single low-frequency estimates on average.

In contrast to multifrequency pumping rates, adding observation wells at new locations creates new independent equations for inverse modeling. If the number of equations still is less than the number of unknowns, the inverse problem remains ill-defined. But these independent equations reduce the possible number of solutions and, in turn, uncertainty. Likewise, inverse modeling starting with prior information close to the true field could yield a good estimate, even though it does not add independent equations. But such a prior would vary with the true field. For these reasons, MC simulation with the same prior is appropriate for assessing the multifrequency test's improvements.

8. Ergodicity

This section revisits the ergodicity issue to explain the large standard deviations of R^2 and L_2 in the Monte Carlo simulation. This revisit is necessary because many studies have disregarded the ergodicity assumption embedded in the inverse modeling of ill-defined problems. Wang et al. (2019) raised this issue and demonstrated that a large number of spatial observations are necessary to fulfill the ergodicity assumption, as illustrated by the narrower standard deviations of performance metrics in Fig. 10 of their paper. However, the ergodicity issue in their study is somewhat different from what we encountered in the OHT. As a result, this issue deserves a further revisit.

The large standard deviations of R^2 and L_2 in Monte Carlo simulation (Figs. 2, 8, and 10) indicate that estimates of OHT with pumping rates of different frequencies or multifrequency could vary significantly from one realization to another. We attribute these deviations to the unfulfilled ergodicity assumption. Ergodicity is the foundation for the application of stochastic theories to a single realization of a random field. It is commonly recognized as the requirement that the aquifer must be larger

than many times of the ensemble correlation scale. With such a large aquifer, the spatial mean, variance, and correlation structures of parameter heterogeneity in one realization are identical to those in the ensemble statistics. However, the ergodicity for the state variable fields (e.g., head and concentration) has rarely been explored. As pointed out in p.48–49 of Yeh et al. (2015b), the state variable ergodicity requires that the state variable at the observation well samples sufficient heterogeneity (i.e., obtain representative samples of heterogeneity, REV).

To illustrate the necessary conditions for satisfying the state variable ergodicity, we use the head variance (or standard deviation from the mean head) in the observation well of a cross-hole test in a 2-D infinite aquifer as an example. First, one must recognize that the head at any location theoretically is impacted by all heterogeneity in the domain once pumping starts due to the pressure propagation's diffusive nature. However, the influence of heterogeneity from different locations on the head is not equal in time and space. For example, previous cross-correlation analysis indicates that the head at an observation well, at early times of a cross-hole test, is governed by heterogeneity between pumping and observation wells. However, heterogeneity at both sides of the observation and pumping wells dictate its behavior at the late time. This fact has been explored by Wu et al. (2005), Sun et al. (2013), and Wen et al. (2019) for 2-D aquifers and Mao et al. (2013b) for 3-D unconfined aquifers.)

While the influence of heterogeneity from different times and locations on the head is not equal, the variance of a random parameter field in the ensemble sense is the same everywhere. Also, the correlation pattern at a given time is identical regardless of the paired pumping and observation wells' locations. For this reason, the head variance at the observation well evaluated on the mean parameter field is independent of the pair's location in the aquifer.

In the case of one realization, the parameter perturbations ($T'(x)$ and $S'(x)$, as opposed to their ensemble variances, vary at different locations. Thus, the heterogeneity in the well pair's vicinity dictates the head perturbation at the observation well. To ensure that the head perturbation reflects the effects of the parameters' ensemble variability, one must simulate a cross-hole test for a long time in a large domain (many times of the correlation scale of the parameter). Otherwise, many cross-hole tests must be performed at many locations. In these cases, the head experiences the same spatial statistics as the ensemble one (i.e., state variable ergodicity is met). In the cases where a small domain is used, Monte Carlo simulation is a must. It derives many possible heads at the observation well during a cross-hole test at fixed locations, using many random field realizations in this domain.

This study examines each element's conditional parameter estimate in a highly parameterized model aquifer, given some observed heads. Since the problem is under-determined, many possible estimates exist. Like the head variance discussed previously, the estimates are dictated by the local heterogeneity near the observation and pumping wells. Therefore, the performance metrics (e.g., R^2 and L_2) of the estimates could vary with the OHT well locations or from one realization to another, as demonstrated by the single realization's result in the previous sections. To obtain the metrics' representative values, one needs a dense well field, which permits the observed heads in the inverse model to experience sufficient heterogeneity to satisfy the state variable ergodicity (Fig. 11 in Wang et al., 2019). Otherwise, conducting a Monte Carlo simulation of OHT at a given well field and evaluating the average OHT performance over many realizations of this aquifer's heterogeneity is necessary. This approach ensures the meaningful statistical evaluation of the inverse modeling results.

Many published works have overlooked this ergodicity issue and claimed that using a sparse network in one heterogeneous field, OHT or PPT increases the estimate resolution. These claims are subjective since neither is the above ergodicity issue recognized nor has the theoretical proof been provided.

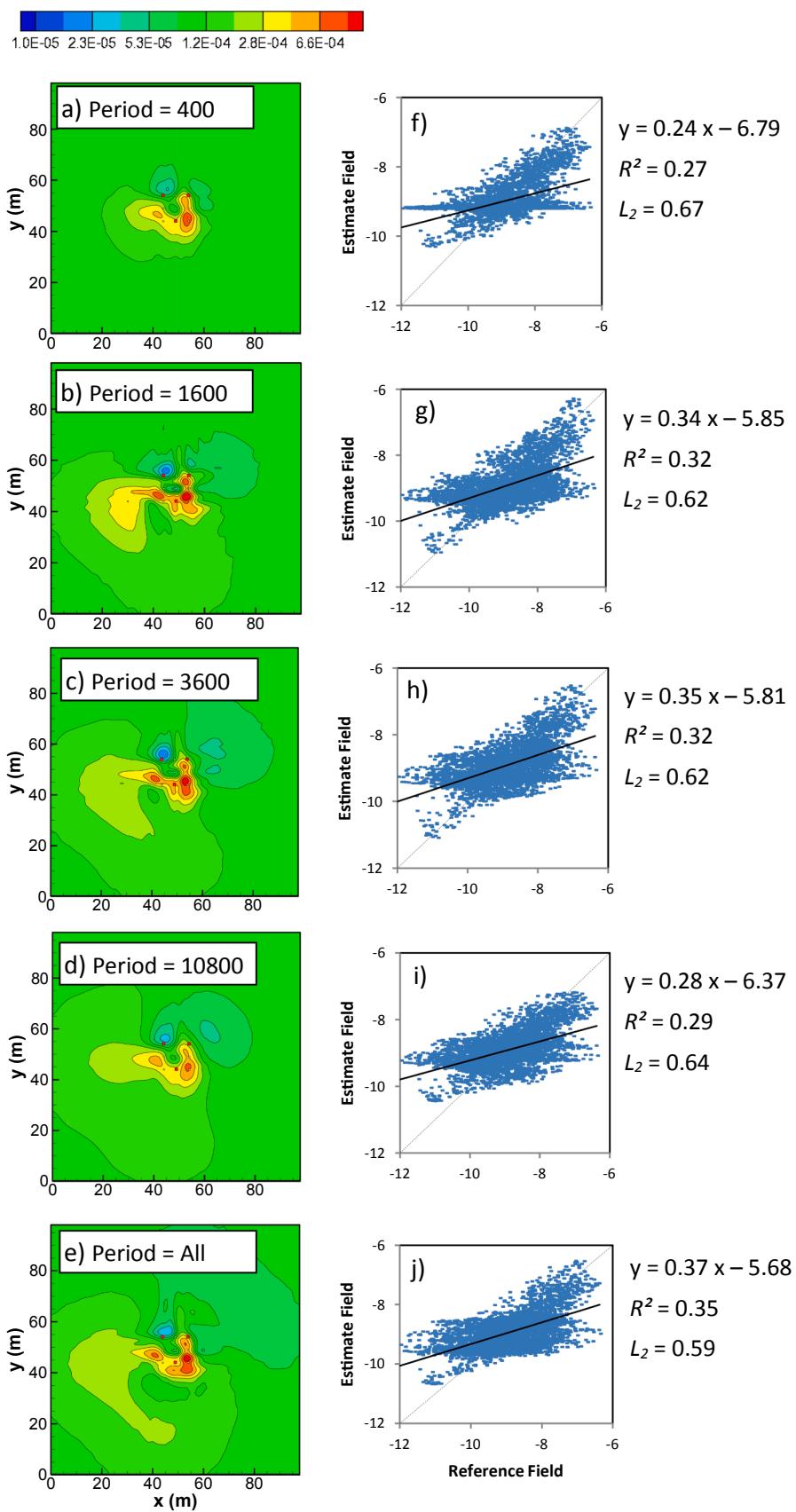


Fig. 11. a)-e) are the estimated $\ln T$ (m^2/s) fields using different and multiple frequencies in the large simulation domain.

9. Conclusion

This paper conducts a stochastic analysis to elucidate the effectiveness of delineating aquifer hydraulic heterogeneity using a periodic pumping test or an oscillatory hydraulic tomography. The analysis shows that unknown heterogeneity, heads, and fluxes produce the ensemble residual flux and storage terms in the ensemble mean equations. These terms lead to frequency dependence, scenario dependence, and dynamic effects of effective hydraulic parameters. Further, this analysis shows that manipulating the frequency of the pumping rate does not eliminate these terms. These terms vanish only if T and S distributions or water level and flux fields are measured everywhere.

Results of Monte Carlo simulation confirm the analysis that the multifrequency PPT or OHT does not increase the resolution of aquifer characterization. The robustness of multifrequency tests proclaimed by the previous studies misunderstood the following: 1) Change in sensitivity value alone does not warrant estimates' improvements. That is, according to some sensitivity analysis, multifrequency excitations may appear to bring forth some additional information. Still, without additional observation wells, the heterogeneity's location remains unknown. 2) The ergodicity assumption ingrained in the ill-defined inverse problems is overlooked. In other words, the results of PPT or OHT from a single realization or experiment could be misleading. For these reasons, we advocate that Monte Carlo simulation must be employed to evaluate such inverse modeling's results. Moreover, the dense monitoring network and cost-effective data collection procedure are the keys to aquifer characterization. Such common sense is, however, often ignored.

Lastly, we do not discredit OHT or PPT. As noticed by many, the oscillating signals can easily extract from background signals such as the sensor noise, sensor drift, and other hydraulic influences (e.g., another nearby pumping, river stage changes, etc.). However, constant-rate pumping tests could overcome these issues if the temporal sampling is sufficiently dense and the signal-to-noise ratio is large enough (Mao et al., 2011, 2013b). Nevertheless, OHT is extremely useful in characterizing contaminated aquifers where contaminated groundwater and discharge are prohibited. For this reason, OHT is highly desirable for pump and treat remediation sites where groundwater is pumped out for treatments and reinjected back to aquifers.

10. Key points

1. Manipulating pumping rate frequency alone does not warrant estimate improvement.
2. Monte Carlo simulation shows that the estimates from different frequencies and multifrequency tests are indistinguishable on average.
3. A dense wellfield or Monte Carlo simulation is necessary to satisfy the ergodicity embedded in inverse modeling evaluation.

CRediT authorship contribution statement

Yu-Li Wang: Conceptualization, Methodology, Software, Investigation, Formal analysis, Writing - original draft, Visualization. **Tian-Chyi Jim Yeh:** Conceptualization, Methodology, Supervision, Project administration, Funding acquisition. **Dong Xu:** Validation. **Kuangjia Li:** Validation. **Jet-Chau Wen:** Funding acquisition. **Shao-Yang Huang:** Funding acquisition. **Wenke Wang:** Funding acquisition. **Yonghong Hao:** Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research is in part by the U.S. Civilian Research and Development Foundation (CRDF) under the award number DAA2-15-61224-1: Hydraulic tomography in shallow alluvial sediments: Nile River Valley, Egypt. The second author also acknowledges the support of the U.S. NSF grant EAR1931756. The program and the data used in this study are available from the corresponding author on reasonable request. The authors thank the editors and reviewers for their helpful and insightful comments which have significantly improved this work.

References

- Ahn, S., Horne, R.N., 2010. Estimating permeability distributions from pressure pulse testing. SPE Paper 134391 Presented at the SPE Annual Technical Conference and Exhibition held in Florence, Italy, 19-22 September 2010 doi:10.2118/134391-MS.
- Berg, S.J., Illman, W.A., 2011. Three-dimensional transient hydraulic tomography in a highly heterogeneous glaciofluvial aquifer-aquitard system. Water Resour. Res. 47, W10507. <https://doi.org/10.1029/2011WR010616>.
- Berg, S.J., Illman, W.A., 2015. Comparison of hydraulic tomography with traditional methods at a highly heterogeneous site. Groundwater 53 (1), 71–89. <https://doi.org/10.1111/gwat.2015.53.issue-110.1111/gwat.12159>.
- Black, J.H., Kipp Jr., K.L., 1981. Determination of hydrogeological parameters using sinusoidal pressure tests: A theoretical appraisal. Water Resour. Res. 17 (3), 686–692. <https://doi.org/10.1029/WR017i003p00686>.
- Butler, J.J., McElwee, C.D., 1990. Variable-rate pumping tests for radially symmetric nonuniform aquifers. Water Resour. Res. 26 (2), 291–306. <https://doi.org/10.1029/WR026i002p00291>.
- Cardiff, M., Barrash, W., 2011. 3-D transient hydraulic tomography in unconfined aquifers with fast drainage response. Water Resour. Res. 47, W12518. <https://doi.org/10.1029/2010WR010367>.
- Cardiff, M., Bakhos, T., Kitaniidis, P.K., Barrash, W., 2013. Aquifer heterogeneity characterization with oscillatory pumping: Sensitivity analysis and imaging potential. Water Resour. Res. 49 (9), 5395–5410. <https://doi.org/10.1002/wrcr.20356>.
- Cardiff, M., Barrash, W., 2015. Analytical and semi-analytical tools for the design of oscillatory pumping tests. Groundwater 53 (6), 896–907. <https://doi.org/10.1111/gwat.2015.53.issue-610.1111/gwat.12308>.
- Cheng, K.B., Rabinovich, A., Dagan, G., 2019. Stochastic modeling of oscillatory pumping in heterogeneous aquifers with application to Boise aquifer test. J. Hydrol. 569, 278–290. <https://doi.org/10.1016/j.jhydrol.2018.11.045>.
- Cooper Jr., H.H., Jacob, C.E., 1946. A generalized graphical method for evaluating formation constants and summarizing well-field history. Eos Trans. AGU 27 (4), 526–534. <https://doi.org/10.1029/TR027i004p00526>.
- Dagan, G., 1985. Stochastic modeling of groundwater flow by unconditional and conditional probabilities: the inverse problem. Water Resour. Res. 21 (1), 65–72. <https://doi.org/10.1029/WR021i001p00065>.
- Daranond, K., Yeh, T.-C., Hao, Y., Wen, J.-C., Wang, W., 2020. Identification of groundwater basin shape and boundary using hydraulic tomography. J. Hydrol. 588, 125099. <https://doi.org/10.1016/j.jhydrol.2020.125099>.
- Fischer, P., Jardani, A., Jourde, H., Cardiff, M., Wang, X., Chedeville, S., Lecoq, N., 2018. Harmonic pumping tomography applied to image the hydraulic properties and interpret the connectivity of a karstic and fractured aquifer (Lez aquifer, France). Adv. Water Resour. 119, 227–244. <https://doi.org/10.1016/j.advwatres.2018.07.002>.
- Gao, X., Yeh, T.-C. J., Yan, E.-C., Wang, Y.-L., Hao, Y., 2021. Conditional mean, effective, and realizations of hydraulic conductivity fields. J. Hydrol. 592 <https://doi.org/10.1016/j.jhydrol.2020.125606>.
- Gelhar, L.W., 1986. Stochastic subsurface hydrology from theory to applications. Water Resour. Res. 22 (9S), 135S–145S. <https://doi.org/10.1029/WR022i09Sp0135S>.
- Gelhar, L.W., 1993. *Stochastic Subsurface Hydrology*. Prentice Hall, p. 390. ISBN: 0138467676.
- Guadagnini, A., Neuman, S.P., 1999. Nonlocal and localized analyses of conditional mean steady state flow in bounded, randomly nonuniform domains: 1. Theory and computational approach. Water Resour. Res. 35 (10), 2999–3018. <https://doi.org/10.1029/1999WR900160>.
- Guiltinan, E., Becker, M.W., 2015. Measuring well hydraulic connectivity in fractured bedrock using periodic slug tests. J. Hydrol. 521, 100–107. <https://doi.org/10.1016/j.jhydrol.2014.11.066>.
- Gutjahr, A.L., 1989. *Fast Fourier Transforms For Random Field Generation, Project Report for Los Alamos grant, Contract 4-R58-2690R. Dep. of Math., N. M. Tech., Socorro*.
- Huang, S.-Y., Wen, J.-C., Yeh, T.-C.-J., Lu, W., Juan, H.-L., Tseng, C.-M., Lee, J.-H., Chang, K.-C., 2011. Robustness of joint interpretation of sequential pumping tests: Numerical and field experiments. Water Resour. Res. 47, W10530. <https://doi.org/10.1029/2011WR010698>.
- Illman, W.A., Liu, X., Takeuchi, S., Yeh, T.-C.-J., Ando, K., Saegusa, H., 2009. Hydraulic tomography in fractured granite: Mizunami Underground Research site, Japan. Water Resour. Res. 45, W01406. <https://doi.org/10.1029/2007WR006715>.
- Indelman, P., 2003. Transient pumping well flow in weakly heterogeneous formations. Water Resour. Res. 39, 1287. <https://doi.org/10.1029/2003WR002036>.

- Kitanidis, P.K., Vomvoris, E.G., 1983. A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. *Water Resour. Res.* 19 (3), 677–690. <https://doi.org/10.1029/WR019i003p00677>.
- Kitanidis, P.K., 1995. Quasi-Linear Geostatistical Theory for Inversing. *Water Resour. Res.* 31 (10), 2411–2419. <https://doi.org/10.1029/95WR01945>.
- Lavenue, M., deMarsily, G., 2001. Three-dimensional interference test interpretation in a fractured aquifer using the Pilot Point Inverse Method. *Water Resour. Res.* 37 (11), 2659–2675. <https://doi.org/10.1029/2000WR000289>.
- Li, F., Yeh, T.-C. J., Wang, Y.-L., Song, X., Lei, X., Wen, J.-C., Wang, W., Hao, Y., 2020. Potential of hydraulic tomography in identifying boundary conditions of groundwater basins. *Water Resour. Res.* 56 <https://doi.org/10.1029/2020WR028331>.
- Mao, D., Wan, L., Yeh, T.-C.-J., Lee, C.-H., Hsu, K.-C., Wen, J.-C., Lu, W., 2011. A revisit of drawdown behavior during pumping in unconfined aquifers. *Water Resour. Res.* 47, W05502. <https://doi.org/10.1029/2010WR009326>.
- Mao, D., Yeh, T.-C.-J., Wan, L., Hsu, K.-C., Lee, C.-H., Wen, J.-C., 2013a. Necessary conditions for inverse modeling of flow through variably saturated porous media. *Adv. Water Resour.* 52, 50–61. <https://doi.org/10.1016/j.advwatres.2012.08.001>.
- Mao, D., Yeh, T.-C., Wan, L., Wen, J.-C., Lu, W., Lee, C.-H., Hsu, K.-C., 2013b. Joint interpretation of sequential pumping tests in unconfined aquifers. *Water Resour. Res.* 49 (4), 1782–1796. <https://doi.org/10.1002/wrcr.20129>.
- Neuman, S.P., Orr, S., 1993. Prediction of steady state flow in nonuniform geologic media by conditional moments: Exact nonlocal formalism, effective conductivities, and weak approximation. *Water Resour. Res.* 29 (2), 341–364. <https://doi.org/10.1029/92WR02062>.
- Priestley, M.B. (1979), Spectral Analysis and Time Series, ISBN: 9780125649223.
- Rabinovich, A., Dagan, G., Miloh, T., 2013. Effective conductivity of heterogeneous aquifers in unsteady periodic flow. *Adv. Water Resour.* 62, 317–326. <https://doi.org/10.1016/j.advwatres.2013.09.002>.
- Notay, Y., 2010. An aggregation-based algebraic multigrid method. *Electron. Trans. Numer. Anal.* 37, 123–146.
- Notay, Y., 2012. Aggregation-based algebraic multigrid for convection-diffusion equations. *SIAM J. Sci. Comput.* 34 (4), A2288–A2316. <https://doi.org/10.1137/110835347>.
- Notay, Y., Napov, A., 2012. An algebraic multigrid method with guaranteed ! convergence rate. *SIAM J. Sci. Comput.* 34 (2), A1079–A1109. <https://doi.org/10.1137/100818509>.
- Rabinovich, A., Barrash, W., Cardiff, M., Hochstetler, D.L., Bakhos, T., Dagan, G., Kitanidis, P.K., 2015. Frequency dependent hydraulic properties estimated from harmonic pumping tests in an unconfined aquifer. *J. Hydrol.* 531, 2–16. <https://doi.org/10.1016/j.jhydrol.2015.08.021>.
- Rasmussen, T.C., Haborak, K.G., Young, M.H., 2003. Estimating aquifer hydraulic properties using sinusoidal pumping at the Savannah River Site, South Carolina. *USA. Hydrogeol. J.* 11 (4), 466–482. <https://doi.org/10.1007/s10040-003-0255-7>.
- Renner, J., Messar, M., 2006. Periodic pumping tests. *Geophys. J. Int.* 167, 479–493. <https://doi.org/10.1111/j.1365-246X.2006.02984.x>.
- Robin, M.J.L., Gutjahr, A.L., Sudicky, E.A., Wilson, J.L., 1993. Cross-correlated random field generation with the direct Fourier transform method. *Water Resour. Res.* 29 (7), 2385–2397. <https://doi.org/10.1029/93WR00386>.
- Schuite, J., Longuevergne, L., Bour, O., Guihneuf, N., Becker, M.W., Cole, M., Burbey, T. J., Lavenant, N., Boudin, F., 2017. Combining periodic hydraulic tests and surface tilt measurements to explore in situ fracture hydromechanics. *J. Geophys. Res. Solid Earth* 122 (8), 6046–6066. <https://doi.org/10.1002/2017JB014045>.
- Song, I., Renner, J., 2008. Hydromechanical properties of Fontainebleau sandstone: Experimental determination and micromechanical modeling. *J. Geophys. Res.* 113, B09211. <https://doi.org/10.1029/2007JB005055>.
- Sun, A.Y., Lu, J., Hovorka, S., 2015. A harmonic pulse testing method for leakage detection in deep subsurface storage formations. *Water Resour. Res.* 51 (6), 4263–4281. <https://doi.org/10.1002/2014WR016567>.
- Sun, A.Y., Lu, J., Islam, A., 2017. Laboratory validation study of using time-lapse oscillatory pumping test for leakage detection in geological repositories. *J. Hydrol.* 548, 598–604. <https://doi.org/10.1016/j.jhydrol.2017.03.035>.
- Sun, N.-Z., Yeh, W.-W.-G., 1990. Coupled inverse problems in groundwater modeling: 1. Sensitivity analysis and parameter identification. *Water Resour. Res.* 26 (10), 2507–2525. <https://doi.org/10.1029/WR026i010p02507>.
- Sun, R., Yeh, T.-C., Mao, D., Jin, M., Lu, W., Hao, Y., 2013. A temporal sampling strategy for hydraulic tomography analysis. *Water Resour. Res.* 49 (7), 3881–3896. <https://doi.org/10.1002/wrcr.20337>.
- Sykes, J.F., Wilson, J.L., Andrews, R.W., 1985. Sensitivity Analysis for Steady-State Groundwater Flow Using Adjoint Operators. *Water Resour. Res.* 21 (3), 359–371. <https://doi.org/10.1029/WR021i003p00359>.
- Tartakovskiy, D.M., Neuman, S.P., 1998. Transient flow in bounded randomly heterogeneous domains: 1. Exact conditional moment equations and recursive approximations. *Water Resour. Res.* 34 (1), 1–12. <https://doi.org/10.1029/97WR02118>.
- Theis, C.V., 1935. The relation between the lowering of the Piezometric surface and the rate and duration of discharge of a well using ground-water storage. *Eos Trans. AGU* 16 (2), 519–524. <https://doi.org/10.1029/TR016i002p00519>.
- Thiem, G., 1906. *Hydrologische Methoden*. Gebhardt, Leipzig.
- Wang, Y.-L., Yeh, T.-C.-J., Wen, J.-C., Gao, X., Zhang, Z., Huang, S.-Y., 2019. Resolution and ergodicity issues of river stage tomography with different excitations. *Water Resour. Res.* 55, 4974–4993. <https://doi.org/10.1029/2018WR023204>.
- Wen, J.-C., Chen, J.-L., Yeh, T.-C., Wang, Y.-L., Huang, S.-Y., Tian, Z., Yu, C.-Y., 2019. Redundant and nonredundant information for model calibration or hydraulic tomography. *Groundwater* 58 (1), 79–92. <https://doi.org/10.1111/gwat.v58.110.1111/gwat.12879>.
- Wu, C.-M., Yeh, T.-C.-J., Zhu, J., Lee, T.H., Hsu, N.-S., Chen, C.-H., Sancho, A.F., 2005. Traditional analysis of aquifer tests: Comparing apples to oranges? *Water Resour. Res.* 41, W09402. <https://doi.org/10.1029/2004WR003717>.
- Xiang, J., Yeh, T.-C., Lee, C.-H., Hsu, K.-C., Wen, J.-C., 2009. A simultaneous successive linear estimator and a guide for hydraulic tomography analysis. *Water Resour. Res.* 45 (2) <https://doi.org/10.1029/2008WR007180>.
- Yeh, T.-C., Srivastava, R., Guzman, A., Harter, T., 1993. A Numerical Model For Water Flow And Chemical Transport In Varily Saturated Porous Media. *Groundwater* 31 (4), 634–644. <https://doi.org/10.1111/gwat.1993.31.issue-410.1111/j.1745-6584.1993.tb00597.x>.
- Yeh, T.-C., Gutjahr, A.L., Jin, M., 1995. An iterative cokriging-like technique for groundwater flow modeling. *Groundwater* 33 (1), 33–41. <https://doi.org/10.1111/gwat.1995.33.issue-110.1111/j.1745-6584.1995.tb0260.x>.
- Yeh, T.-C.-J., Jin, M., Hanna, S., 1996. An iterative stochastic inverse method: conditional effective transmissivity and hydraulic head fields. *Water Resour. Res.* 32 (1), 85–92. <https://doi.org/10.1029/95WR02869>.
- Yeh, T.-C.-J., Liu, S., 2000. Hydraulic tomography: Development of a new aquifer test method. *Water Resour. Res.* 36 (8), 2095–2105. <https://doi.org/10.1029/2000WR000114>.
- Yeh, T.-C., Mao, D., Zha, Y., Hsu, K.-C., Lee, C.-H., Wen, J.-C., Lu, W., Yang, J., 2014. Why hydraulic tomography works? *Groundwater* 52 (2), 168–172. <https://doi.org/10.1111/gwat.12129>.
- Yeh, T.-C.J., Carroll, K.C., 2015a. *Flow through Heterogeneous Geologic Media*. Cambridge Press. ISBN: 9781107076136.
- Yeh, Tian-chyi J., Mao, De-qiang, Zha, Yuan-yuan, Wen, Jet-chau, Wan, Li, Hsu, Kuo-chin, Lee, Cheng-haw, 2015b. Uniqueness, scale, and resolution issues in groundwater model parameter identification. *Water Sci. Eng.* 8 (3), 175–194. <https://doi.org/10.1016/j.wse.2015.08.002>.
- Zha, Tian-chyi J., Mao, De-qiang, Zha, Yuan-yuan, Wen, Jet-chau, Wan, Li, Hsu, Kuo-chin, Lee, Cheng-haw, 2016. An application of hydraulic tomography to a large-scale fractured granite site, Mizunami, Japan. *Groundwater* 54 (6), 793–804. <https://doi.org/10.1111/gwat.12421>.
- Zhao, Z., Illman, W.A., 2018. Three-dimensional imaging of aquifer and aquitard heterogeneity via transient hydraulic tomography at a highly heterogeneous Field Site. *J. Hydrol.* 559, 1–11. <https://doi.org/10.1016/j.jhydrol.2018.09.058>.
- Zhou, YaoQuan, Lim, David, Cupola, Fausto, Cardiff, Michael, 2016. Aquifer imaging with pressure waves - Evaluation of low-impact characterization through sandbox experiments. *Water Resour. Res.* 52 (3), 2141–2156. <https://doi.org/10.1002/2015WR017751>.
- Zhu, J., Yeh, T.-C.J., 2005. Characterization of aquifer heterogeneity using transient hydraulic tomography. *Water Resour. Res.* 41, W07028. <https://doi.org/10.1029/2004WR003790>.