ELSEVIER

Journal of Hydrology 265 (2002) 15–33

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Validation of aquifer parameter determination by extrapolation fitting and treating thickness as an unknown

Tien-Chang Lee[a,*], Thomas Perina[b], Cin-Young Lee[c]

[a]*Department of Earth Sciences, University of California, Riverside, CA 92521-0423, USA*
[b]*International Technology Corporation, 1425 South Victoria Court, Suite A, San Bernardino, CA 92408, USA*
[c]*Engineering Design Research Laboratory, California Institute of Technology, Pasadena, CA 91125, USA*

Received 17 August 2001; revised 10 March 2002; accepted 12 April 2002

Abstract

A genetic algorithm is used here to guess-estimate a close-to-true set of trial values as input to a three-staged quasi-linear inverse modeling scheme for the determination of aquifer parameters. To validate the parameter determination, in addition to the conventional measures of misfit root mean squares (rms) and distribution, the aquifer thickness is treated as an unknown parameter and the model parameters are further evaluated by comparing the expected drawdown with the observed drawdown at wells which are not used for parameter determination (extrapolation fitting). The method is tested with synthetic and observed drawdown data from five partially screened monitoring wells in a water-table aquifer. Test results for synthetic data doped with random errors indicate that modeling based on two or more well data can yield satisfactory parameter values and extrapolation misfits in an ideal aquifer. For field data, the results indicate that a model misfit on par with the standard error of the data is achievable for each individual well or a combination of two wells but the extrapolation misfit distributions are generally biased and their rms are far greater—possibly due to aquifer heterogeneity. Consistent parameter values can be obtained from the geometric means for multiple runs of the genetic-inverse modeling of one-, two-, three-, and four-well data. Our test aquifer can be represented by a set of parameters with 10 to 15% consistency, including transmissivity, storativity, vertical-to-horizontal conductivity ratio, and storativity-to-specific yield ratio, as affirmed by model aquifer thicknesses that deviate less than 10% from the actual thickness.

Keywords: Inverse modeling; Genetic algorithm; Unconfined aquifer; Aquifer parameters; Pumping test

1. Introduction

Curve fitting by inverse modeling of aquifer-test data is a commonly adopted technique for the determination of aquifer parameters. Such determination is made in order to characterize an aquifer for modeling groundwater flow and solute transport and for utilization of water resources. Regarding the test results, two questions are frequently asked. How reliable the determination is and how far are the determined parameters applicable? The answers obviously depend on the quality of data and the complexity of the test aquifer. This paper presents a method to test the reliability of results stemmed solely from good curve matching and assesses the determination by predicting the

* Corresponding author. Tel.: +1-909-787-4506; fax: +1-909-787-4324.
E-mail address: tien.lee@ucr.edu (T. C. Lee).

0022-1694/02/$ - see front matter © 2002 Published by Elsevier Science B.V.
PII: S0022-1694(02)00076-8

drawdown at wells where the observations have not been used for inverse modeling.

Good curve matching, as measured typically by root-mean-squares (rms) misfits that are near the standard errors of the data, can often be achieved with different combinations of parameter values for a given choice of aquifer type. Consequently, it would befit to report various ranges of parameter values for coping with non-unique determinations. Such uncertainty raises quality issue as well as doubt about the methodology of inverse modeling itself. For chemical quality assurance and control, a hydrogeologist may spike a few specimens with known concentrations of the desired constituents and deliver the spiked specimens along with his or her batch of water samples for laboratory analysis. So far, there is no equivalent spiking practice for inverse modeling, other than using synthetic data to verify the computer codes. Here, we attempt to treat aquifer thickness as an unknown parameter for inverse-modeling of the drawdown data. Among the aquifer parameters, the thickness is likely the best known because it can be estimated independently by drilling, wire-line logging, or surface geophysical mapping. Treating thickness as an unknown serves as a means to reassure the confidence of other parameter estimates, in addition to gauging the goodness of curve fitting, sensitivity analysis, parameter resolution, or post-processing variances of parameters.

Although aquifer thickness has been treated conventionally as a given value for the purpose of modeling, it is not always known with certainty. For example, drilling may not extend fully to the base of an aquifer and the thickness can only be guess-estimated. An erroneous estimate of thickness can lead to incorrect determination of aquifer properties as noted by Hudak (1993). Aquifer thickness can also become a fuzzy concept where sand, silt, and clay zones are interlaced and their lateral or vertical variations in thickness or particle size are gradational. In such cases, rather than using a thickness between ill-defined boundaries, effective thickness as a concept to be assessed by hydraulic response to pumping can be a better alternative. Therefore, by treating thickness as an unknown parameter, we would like to see if the drawdown responses could be used to estimate aquifer thickness where it is not known a priori or assign an effective thickness where it is more appropriate.

We do not intend to simultaneously reach both the goals of using thickness for reduction of parameter ambiguity and of defining an effective thickness although both are considered with the same methodology. Simultaneous inverse modeling of multiple well testing has been practiced to obtain a weighted set of aquifer parameters (Bohling and McElwee, 1992; Chen et al., 1999). Here, we practice likewise with multi-well testing but further exercise it from the context of extrapolation. Good extrapolation fitness serves to validate the results of inverse-modeling an aquifer that is free of significant lateral variation. Subject to the condition of good model fitting, poor extrapolation fitting indicates significant aquifer heterogeneity or it may signal other causes.

Our inverse-modeling algorithm is modified from the Gauss–Newton method described by Tarantola (1987) and Lee (1999). Such quasi-non-linear inversion methods work when the initial trial parameter values are close to the actual values. Otherwise, the iterative modeling may not converge or it may end with the incorrect answers at a local rms minimum in the parameter space. Several sets of trial values should therefore be run to ensure that the iterations have converged to the global minimum which is believed to associate with the desired solutions. For commonly encountered aquifers of several parameters, some parameters may range in values by several orders of magnitude and the task of manipulating trial-and-error input can be very daunting. Hence, we adopt a genetic algorithm, as a search engine, to find ‘close-to-true’ trial values for initiating an inverse modeling. A comprehensive review on the genetic or evolutionary algorithms can be seen in Lee et al. (2001) and references therein.

To test the idea of treating thickness as an unknown for model quality assurance and validating model results by extrapolation, drawdown data from five partially screened monitoring wells in the Mojave Desert of California are used. Reflecting the condition of data collection in a typical study for environmental remediation, our drawdown data were chanced upon rather than being the product of a designed test for our method. The aquifer is unconfined. We assume that it behaves like a Neuman water-table aquifer (Neuman, 1972, 1974, 1975) although an ideal Neuman aquifer

may not always be applicable in the field (Nwankwor et al., 1992). Depending on and varying with the nature of an unconfined aquifer, the issue whether the decline of water table is instantaneous (Neuman's) or non-instantaneous (Moench, 1995) will not be addressed here. A comparison study at one field site seems to favor the former (Chen and Ayers, 1998).

Aquifer response to pumping can be affected by well construction. According to Lee (1999, p. 179), from the viewpoint of discriminatable field observations, the effect of finite well diameter for full-penetration pumped wells is negligible at distances beyond 10 times well radius. In our case, the distance from the nearest observation well to the pumped well is 170 times the well radius. Hence, we can safely neglect the wellbore storage effect for our partial-penetration wells in an unconfined aquifer. A plethora of literature on the subject of storage effect can be traced from recent works of Novakowski (1989), Kabala and Cassiani (1997), Cassiani and Kabala (1998), and Cassiani et al. (1999) on confined aquifers. These cited authors also deal with the effect of finite or infinitesimal skin thickness. The skin effect varies greatly from well to well, depending on drilling method, type of drilling mud, and well completion method (e.g. well cleaning, gravel packing, and screening). Relative to intrinsic aquifer properties, the skin can enhance or suppress the hydraulic conductivity in the vicinity of the wellbore. The effect on drawdown can be acute at early time of pumping or observation. Early time data are hence crucial for deciphering the skin effect. Unfortunately, the noise-to-signal ratio is relatively high at early time when the drawdown is low or at distant wells where drawdown is also low. As seen in our records presented later, the skin effect is intractable and will be of secondary concern to us for the determination of aquifer parameters. The effect will be addressed again in the discussion of modeling the real data.

Pumping tests on a Neuman's aquifer can usually yield estimates of transmissivity, storativity, specific yield, and conductivity anisotropy to various extents of uncertainty. These four parameters plus the thickness are here chosen as the unknowns to test our methodology, which also has been adapted for other ideal aquifers including Theis' confined aquifer, Hantush's leaky aquifer, and Lee's (1999, p. 134) boundary flux model. Only the results for a Neuman aquifer are reported as follows.

2. Forward modeling

Assuming the test aquifer to behave like a Neuman's water-table aquifer and neglecting the skin effects and well-bore storage, the theoretically expected drawdown $\overline{\Delta h}$ in the Laplace-transform domain (the overbar designates a transformed variable) at radial distance r and elevation z due to pumping through a well screened between elevation Z_{top} and Z_{bot} at a rate of Q is (Moench, 1996; Lee, 1999, p. 170)

$$\overline{\Delta h}[r,z,s] = \frac{Q}{\pi T s}\sum_{n=0}^{\infty} {\lambda_n}^{\text{part}} \cos[\zeta_n z] K_0[\xi_n r] \tag{1}$$

where

$${\lambda_n}^{\text{part}} = \frac{b}{Z_{\text{top}} - Z_{\text{bot}}} \frac{\sin[\zeta_n Z_{\text{top}}] - \sin[\zeta_n Z_{\text{bot}}]}{\sin[\zeta_n b]} \lambda_n$$

$$\lambda_n = \frac{c}{(b + c + bc^2\zeta_n{}^2)\cos[\zeta_n b]} = \frac{\sin[\zeta_n b]}{\zeta_n b + 0.5\sin[2\zeta_n b]} = \frac{(-1)^n \sin[\theta_n]}{\zeta_n b + 0.5\sin[2\theta_n]} \tag{2}$$

where b is the aquifer thickness, T the transmissivity, $c = K_z/S_y s$, S_y the specific yield, K_z the vertical hydraulic conductivity, and s is the Laplace transform parameter. With the coordinate origin at the base of the unconfined aquifer, the elevation z is upward positive. $K_0[\xi_n r]$ is the zeroth-order modified Bessel function, and its argument is defined through

$$\xi_n = \sqrt{\zeta_n{}^2 k_D + q^2} \tag{3}$$

where ζ_n satisfies the root equation

$$\tan[\zeta_n b] = \frac{1}{c\zeta_n}, \qquad \zeta_n b = n\pi + \theta_n, \qquad 0 \le \theta_n \le \pi/2 \tag{4}$$

with vertical-to-horizontal conductivity ratio $k_D = K_z/K_r$, $q^2 = s/\kappa$, hydraulic diffusivity $\kappa = T/S$, storativity $S = S_s b$, and S_s the specific storage.

The time-domain solution is obtained numerically

from (Stehfest, 1970),

$$\Delta h[r,z,t] = \frac{\ln 2}{t}\sum_{j=1}^{M} v_j \overline{\Delta h}[r,z,s_j] \tag{5}$$

where $s_j = (j \ln 2)/t$, t is the time, M a computer-dependent optimal number (an even number, typically around 10 for personal computer) for maximum accuracy, and v_j is a weighting coefficient,

$$v_j = (-1)^{j+M/2} \sum_{i=m}^{\min[i,M/2]} \times \frac{i^{M/2}(2i)!}{(M/2-i)!i!(i-1)!(j-i)!(2i-j)!} \tag{6}$$

with m being the integer part of $(j+1)/2$.

Introducing three independent, dimensionless variables: $\sigma = S/S_y$, $\beta = (r/b)^2 k_D$, and $u = r^2 S/4Tt$, the drawdown in the time domain is

$$\Delta h[r,z,t] = \frac{Q}{4\pi T} W[u,\sigma,\beta]$$
$$W[u,\sigma,\beta] = \sum_{j=1}^{M} \frac{4v_j}{j} \sum_{n=0}^{\infty} \lambda_{n,j} \cos[\zeta_{n,j} z] K_0[\xi_{n,j} r] \tag{7}$$

where $\zeta_{n,j} b$ is the root of

$$\tan[\zeta_{n,j} b] = \frac{ju}{\sigma\beta}\frac{\ln 16}{\zeta_{n,j} b} \tag{8}$$

The argument of the modified Bessel function $K_0[\xi_{n,j} r]$ is

$$\xi_{n,j} r = \sqrt{\beta(n\pi + \theta_{n,j})^2 + ju \ln 16} \tag{9}$$

and $\lambda_{n,j}$ follows the definition of $\zeta_{n,j} b$.

The root finding is the most time-consuming procedure in the computation of drawdown. The θ_n is adopted for computational efficiency by bracketing root $\zeta_{n+1} b$,

$$\zeta_n b + \pi - (\theta_{n-1} - \theta_n) < \zeta_{n+1} b < \zeta_n b + \pi, \quad n \geq 1 \tag{10}$$

because θ_n decreases with increasing n and $(\theta_{n-1} - \theta_n)$ diminishes as n increases (Lee, 1999, p. 92). Further reduction in root-finding time can be achieved by noting the fact that $\theta_{n,j}$ also decreases with increasing j. In this study, the root $\zeta_{n,j} b$ is obtained with a bi-section method (Press et al., 1986) to a precision better than 10^{-8}.

For a given discharge rate Q, drawdown $\Delta h[r,z,t]$ is completely defined by four independent variables T, σ, β, and u, provided that the implicit pumping-screen interval between Z_{top}/b and Z_{bot}/b is given. In other words, any values of aquifer parameters that can be combined to give the same set of values for the four independent variables will yield identical $\Delta h[r,z,t]$. Hence, one can choose any four independent parameters to describe the drawdown in a Neuman water-table aquifer if these four can be re-combined to define T, σ, β and u completely. The four independent variables chosen for this study are transmissivity T, storativity S, vertical-to-horizontal conductivity ratio K_z/K_r, and the ratio of storativity to specific yield σ. The addition of aquifer thickness as an unknown makes our model a five-parameter water-table aquifer.

The average drawdown $\Delta h[r,t]$ for an observation well screened between elevations O_{bot} and O_{top} (as symbolized by deleting z as a variable) is

$$\Delta h[r,t] = \frac{1}{O_{top} - O_{bot}} \int_{O_{bot}}^{O_{top}} \Delta h[r,z,t] dz$$
$$= \frac{Q}{4\pi T} \frac{\sin[\zeta_{n,j} O_{top}] - \sin[\zeta_{n,j} O_{bot}]}{(O_{top}/b - O_{bot}/b)\zeta_{n,j} b} \tag{11}$$
$$\times \sum_{j=1}^{M} \frac{4v_j}{j} \sum_{n=0}^{\infty} \lambda_{n,j} K_0[\xi_{n,j} r]$$

which provides the expected drawdown for subsequent modeling.

3. Inverse modeling

The desired model parameters are obtained by a quasi-non-linear Gauss–Newton method. Minimization of the objective or misfit function

$$S = \tfrac{1}{2}\{(\mathbf{d}^{obs} - \mathbf{g})^t \mathbf{C}_d{}^{-1} (\mathbf{d}^{obs} - \mathbf{g}) + (\mathbf{p}^{guess} - \mathbf{p})^t \mathbf{C}_p{}^{-1} (\mathbf{p}^{guess} - \mathbf{p})\} \tag{12}$$

with respect to the parameter $\mathbf{p}$ yields an iterative algorithm for updating parameter $\mathbf{p}_k$ at step k to $\mathbf{p}_{k+1}$ at

step $k+1$ (Tarantola, 1987; Lee, 1999, p. 355)

$$\mathbf{p}_{k+1} = \mathbf{p}_k + \epsilon \mathbf{C}_{p'k}\left\{\mathbf{G}_k{}^{\mathrm{t}}\mathbf{C}_{\mathrm{d}}{}^{-1}\left(\mathbf{d}^{\mathrm{obs}} - \mathbf{g}_k\right) + \mathbf{C}_{\mathbf{p}}{}^{-1}\left(\mathbf{p}^{\mathrm{guess}} - \mathbf{p}_k\right)\right\} \tag{13}$$

where column vector $\mathbf{p}^{\mathrm{guess}} = (T, S, b, k_{\mathrm{D}}, \sigma)^{\mathrm{t}}$ is the guessed trial parameters, and superscript t designates transpose of a matrix or vector (a matrix is symbolized with a bold, capital letter while a vector with a bold, lower-case alphabet). The multiplier ϵ ($0 < \epsilon \leq 1$) is introduced here to adjust the size of iterative upgrade $\Delta\mathbf{p}_k$ from $\mathbf{p}_k$ to $\mathbf{p}_{k+1}$. Other variables are defined next.

Symbol $\mathbf{d}^{\mathrm{obs}}$ is the observed drawdowns expressed as an $n \times 1$ column vector (n being the number of data points) and $\mathbf{g}_k$ is the corresponding drawdowns Δh calculated according to Eq. (11) at iteration step k for parameter $\mathbf{p}_k$. The squares of the standard errors constitute the diagonal entries of the $n \times n$ data covariance matrix $\mathbf{C}_{\mathrm{d}}$. Assuming the data are independent of one another, the off-diagonal entries of $\mathbf{C}_{\mathrm{d}}$ are nullified. The five aquifer-defining parameters are presumably not correlated and consequently the $m \times m$ parameter covariance matrix $\mathbf{C}_{\mathbf{p}}$ is diagonal also ($m = 5$). The diagonal entries of $\mathbf{C}_p$ are the squares of the standard deviations of the aquifer parameters but the standard deviations are unknown and they are represented here by a fraction of their respective parameter values (e.g. 2%). The fractional uncertainty of $\mathbf{C}_p$ does not significantly affect the determination of $\mathbf{p}_{k+1}$ but it does affect the post-processing parameter covariance matrix $\mathbf{C}_{p'}$ and the resolution of parameter determination $\mathbf{R}_{pk}$,

$$\begin{aligned} \mathbf{H}_k &= \mathbf{C}_{p'k}{}^{-1} = \mathbf{C}_p{}^{-1} + \mathbf{G}_k{}^{\mathrm{t}}\mathbf{C}_{\mathrm{d}}{}^{-1}\mathbf{G}_k \\ \mathbf{R}_{pk} &= \mathbf{I} - \mathbf{C}_{p'k}\mathbf{C}_p{}^{-1} \end{aligned} \tag{14}$$

where the $\mathbf{H}_k$ is the Hessian. The $n \times m$ matrix $\mathbf{G}_k$ at step k represents the sensitivity of $\mathbf{g}_k$ (or theoretical drawdown Δh) to an infinitesimal change in parameter $\mathbf{p}_k$. Its entries G_{ij} at step k are

$$G_{ij} = \left.\frac{\partial g_i}{\partial p_j}\right|_{\mathbf{p}_k} \tag{15}$$

which is differentiated numerically in this study, typically at a ∂p_j of around 0.01 p_j.

Rewrite Eq. (13) as

$$\Delta\mathbf{p}_k = -\epsilon\mathbf{H}_k{}^{-1}\boldsymbol{\gamma}_k \tag{16}$$

where the $m \times 1$ column vector $\boldsymbol{\gamma}_k$ is the gradient of the misfit function (i.e. $\boldsymbol{\gamma}_k = \partial S_k/\partial\mathbf{p}$). It can be shown that $\boldsymbol{\gamma}_k$ is equal to the negative of the braced term in Eq. (13).

If the multiplication of $\epsilon\mathbf{H}_k{}^{-1}$ is restricted to the diagonal entries of $\mathbf{H}_k{}^{-1}$ only, in other words if $\epsilon = 1/(1+\lambda\delta_{ij})$ (where $\delta_{ij} = 1$ if $i = j$ otherwise it vanishes), then Eq. (16) is equivalent to the Marquardt–Levenberg algorithm

$$(1+\lambda\delta_{ij})\mathbf{H}_k\Delta\mathbf{p}_k = -\boldsymbol{\gamma}_k \tag{17}$$

which represents a method of steepest descent

$$\Delta\mathbf{p}_k = -\lambda_k{}^{-1}\mathbf{H}_k{}^{-1}\boldsymbol{\gamma}_k, \qquad \text{if } \lambda_k \gg 1 \text{ or } \epsilon \ll 1 \tag{18}$$

or a Gauss–Newton method

$$\Delta\mathbf{p}_k = -\mathbf{H}_k{}^{-1}\boldsymbol{\gamma}_k, \qquad \text{if } \lambda_k \ll 1 \text{ or } \epsilon = 1 \tag{19}$$

In this study, ϵ is usually set at 0.75. Thus, we regard our method as a modified Gauss–Newton method.

Even though the misfit function S is minimized to obtain parameter $\mathbf{p}_{k+1}$ in Eq. (13), we use root-mean-squares at each iteration step,

$$\mathrm{rms}_k = \sqrt{\frac{\left(\mathbf{d}^{\mathrm{obs}} - \mathbf{g}_k\right)^t\left(\mathbf{d}^{obs} - \mathbf{g}_k\right)}{n-m}} \tag{20}$$

as a criterion to gauge misfit. The $\mathbf{p}$ associated with the minimum of all rms_k for a given $\mathbf{p}^{\mathrm{guess}}$ is the final for that particular $\mathbf{p}^{\mathrm{guess}}$. By varying $\mathbf{p}^{\mathrm{guess}}$, many plausible solutions, of which the rms are near the standard error of data, can be obtained if model constraints besides drawdown data are not available. The task of finding a correct $\mathbf{p}$ is to have a $\mathbf{p}^{\mathrm{guess}}$ that is close to the true but unknown $\mathbf{p}^{\mathrm{true}}$.

4. Finding model parameters

Our iterative computation (13) for $\mathbf{p}$ starts with a $\mathbf{p}^{\mathrm{guess}}$ picked from large ranges of potential parameter values. The process of picking is based on genetic algorithm (GA).

4.1. Genetic algorithm

The GAs are generally robust search engines that can typically find global optima in highly non-linear and multi-modal landscape. The algorithm borrows the concept in Darwinian evolution of natural selection in which the fittest individuals produce more offspring while the poorer perish. Being problem-dependent, genetic algorithms have many forms and names. A recent review can be found in Lee et al. (2001) and references therein. Here, we describe only what has been implemented for our inverse problem. It is essentially an iterative search procedure for $\mathbf{p}^{\text{guess}}$.

(1) A range of values is defined for every parameter. (2) A number N_p of individuals serving as the parent (initial) population is assigned and each individual bears five randomly selected parameter values, which are logarithmically transformed for the selection with the genetic algorithm. (3) Each of the N_p individuals yields a rms for n misfits ($\Delta h^{\text{obs}} - \Delta h^{\text{cal}}$) at the n observation points. The best (the minimum) among N_p rms is identified, rms^{min}. (4) A standard deviation σ_{rms} for N_p rms is calculated. (5) Any individual whose rms is within a fraction of σ_{rms} from the rms^{min} is deemed as a survivor. At the end of selection, there are N_s survivors ($N_s < N_p$) to produce the next generation of N_c children (offspring).

The N_c offspring are generated from the N_s survivors by four different ways, essentially parameter retention, exchange, perturbation, and recombination. Each way is selected at random and approximately has a 0.25 probability of being exercised if the random number generator is unbiased and N_c is sufficiently large (>30). Individuals from the N_s survivors are randomly selected to reproduce and a survivor may participate more than once to co-produce offspring as follows. (a) An offspring can inherit all parameter values of a survivor without any modification. (b) It can possess the geometric means of two survivors' respective parameter values. These means are inversely weighted by the squares of the two survivors' misfit rms. (c) Two survivors can cross-breed to produce an offspring through exchange of two randomly selected parameters. (d) An offspring can be produced by mutation of one survivor by increasing or decreasing some of its parameter values by a factor determined randomly between 0 and $\sqrt{10}$ (this factor is somewhat arbitrarily set). The mutation provides an opportunity for the search of $\mathbf{p}^{\text{guess}}$ to go outside the ranges encompassed by the survivors since the survivors may not be necessarily among the best fitted.

When the number of offspring reaches the size of the original parent population N_p (i.e. $N_c = N_p$), they become the new parents and go through the selection and reproduction processes to yield another generation of offspring. The population size in each generation stays the same in our scheme. These processes are recycled through N_g generations. There are N_p sets of parameters in each of N_g generations and each generation has a best-fitted parameter set that yields a rms^{min} as described above. Among those N_g best sets, the set associated with the smallest rms^{min} is the $\mathbf{p}_1{}^{\text{guess}}$. The preceding numerical process constitutes the first batch of our genetic algorithm (hence the subscript 1 in $\mathbf{p}_1{}^{\text{guess}}$). One can either stop operating the genetic algorithm here and proceed to use the inverse modeling, or continue similar batch processes for a total of N_b times.

The second batch process is essentially the same as the first except that the parameter ranges are now centered around $\mathbf{p}_i{}^{\text{guess}}$ by a factor of $12/n_b$, where n_b is the n_bth batch of the N_b batch processes ($i = 1, 2, \ldots, N_b$, and $N_b < 12$). There are N_b sets of $\mathbf{p}_i{}^{\text{guess}}$ and among them, the one with the least rms is named $\mathbf{p}_{\text{min}}{}^{\text{guess}}$. Those $\mathbf{p}_i{}^{\text{guess}}$ whose misfit rms are within 0.75 standard error of the data from the $\mathbf{p}_{\text{min}}{}^{\text{guess}}$ are combined in accordance with the weighting

$$\mathbf{p}^{\text{guess}} = \sum_{j=1}^{J} \mathbf{p}_j{}^{\text{guess}} \frac{1}{\text{rms}_j{}^{\text{min}} A}, \qquad A = \sum_{k=1}^{J} \frac{1}{\text{rms}_k{}^{\text{min}}} \tag{21}$$

to find a final weighted $\mathbf{p}^{\text{guess}}$ where J is the number of $\mathbf{p}_i{}^{\text{guess}}$ that are within the defined limit. This weighting is intended to move $\mathbf{p}^{\text{guess}}$ slightly from the $\mathbf{p}^{\text{guess}}$ generated by the selection and reproduction processes. By trial and error, N_b is set at 3 in this study.

This $\mathbf{p}^{\text{guess}}$ generally yields a rms that is near the standard error of data but the distribution of misfit as measured by $\mathbf{d}^{\text{obs}} - \mathbf{g}$ is frequently biased or skewed. Hence, such a $\mathbf{p}^{\text{guess}}$ is not necessarily the desirable

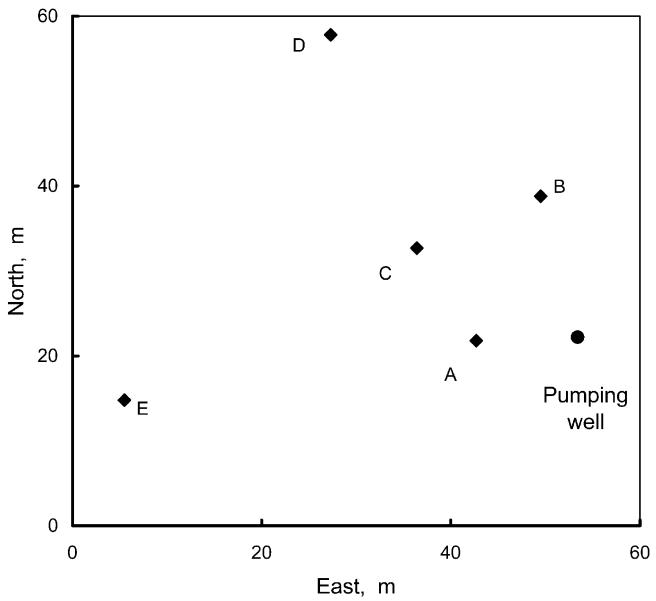

Fig. 1. Relative locations of monitoring wells, ordered alphabetically at increasing distances from the pumping well.

answer and it is appropriately labeled as a guessed set despite the possibility that its low rms may render a determination of acceptable **p**. An inverse modeling based on an algorithm like Eq. (13) is thus required to minimize misfit and improve misfit distribution for better estimates of parameters.

4.2. Inverse modeling

Inputting $\mathbf{p}^{\text{guess}}$ into Eq. (13) yields parameters **p** and associated standard deviations, given by the square roots of the diagonal entries of $\mathbf{C}_{p'}$ (the post-processing parameter covariance matrix). Compared to $\mathbf{p}^{\text{guess}}$, this **p** will usually improve the goodness of fitting but is not necessarily an acceptable set of parameters because, among a few plausible causes, the **p** could be located at a local minimum of misfit instead of a global minimum in the space of multivariate parameters. The standard deviations of the model parameters and the model resolution can be manipulated by adjusting the standard deviations of data and parameters through $\mathbf{C}_{\text{d}}$ and $\mathbf{C}_p$, which are practically unknown but assumed for modeling, and hence do not provide unequivocal assurance on the parameter determination. Thus, a three-staged inverse modeling is launched as a precautionary measure. The above description constitutes the first stage to produce $\mathbf{p}^{(1)}$.

In the second stage, $\mathbf{p}^{(1)}$ is used as a new $\mathbf{p}^{\text{guess}}$ to generate $\mathbf{p}^{(2)}$ while the parameter with the greatest resolution (the greatest diagonal entry in $\mathbf{R}_p$) is held stationary. This stage tests whether **p** stays stationary when $\mathbf{p}^{(1)}$ is slightly perturbed.

In the third stage, $\mathbf{p}^{(2)}$ serves in return as $\mathbf{p}^{\text{guess}}$ to yield $\mathbf{p}^{(3)}$. According to our testing for synthetic and field data, the misfits associated with $\mathbf{p}^{(3)}$ are usually better than that with $\mathbf{p}^{(1)}$ but the differences are insignificant as gauged by the standard error of data. Occasionally, cases of instability occur, as diagnosed by rapid increase or decline in one or more of the parameter values as the modeling progresses from $\mathbf{p}^{\text{guess}}$ through $\mathbf{p}^{(1)}$ and $\mathbf{p}^{(2)}$ to $\mathbf{p}^{(3)}$. So, the three-stage inversion modeling can detect a wrong determination of **p** with an acceptable rms but still cannot assure a correct answer.

To procure an acceptable **p**, we re-run the genetic algorithm and inverse modeling a few times to test the sensitivity of **p** to various $\mathbf{p}^{\text{guess}}$ as generated by different seeding random numbers. The variations in those **p** are indicative of parameter uncertainty. The variances in $\mathbf{C}_{p'}$ are also indicative of uncertainty but their sizes can be manipulated with the unknown

Table 1
Distances to the pumping well and screen intervals in depths below the pre-pumping water table

Well	Distance (m)	Top (m)	Bottom (m)
Pump		5.0	11.0
A	10.7	0.0	7.9
B	17.1	6.1	7.6
C	20.4	3.1	4.6
D	44.2	0.0	1.6
E	48.8	1.5	4.6

standard deviations of data and parameters themselves. In addition to misfit itself and misfit distribution, the model thickness as a parameter acts as a measure of uncertainty.

5. Test and discussion

Our methodology was tested with synthetic and observed drawdown data at five monitoring wells in a water-table aquifer.

5.1. Data

The test aquifer consists of Pleistocene unconsolidated fluvial deposits in the southwestern Mojave Desert in California (Cox and Hillhouse, 2000). It is underlain by a well-defined aquitard of lacustrine clay, dipping about 1% northwest. Boring records indicate that the aquifer consists of interfingering zones of sand, silt, and clay with high degrees of lateral variation. Correlation of lithologic units between boreholes is generally poor. High-resolution seismic profiles are not available. At the study site, the ground is fairly flat (slope $<1\%$) and the aquifer thickness is about 21 m (Sibbett, 1999).

Fig. 1 shows the relative positions of the pumping and observation wells. The five observation wells are sequenced from A to E in the order of increasing distances from the pumping well. All observation wells were rotary-drilled with hollow-stem auger and cased with 4-inch slotted PVC pipes (10.16 cm in outer diameter); and the pumping well was drilled using air-rotary and cased with 5-inch slotted PVC pipe (12.7 cm). Their screen intervals at depths below the pre-pumping water table are listed in Table 1. A screen-top at zero depth implies that the physical screen at the well of concern was installed astride the pre-pumping water table.

At a constant pumping rate of 8.14×10^{-4} m^3 s^{-1}, the drawdowns were measured with pressure transducers and recorded at sampling intervals from less than one second at early pumping time to one minute at late time for 60 h. In order to reduce computation time for modeling, a subset of data was extracted from the original records. Each decade of time in seconds in the data file for modeling contains about 4–6 data points at log-equidistant time intervals. Those are the observed data in this paper.

The data are assumed to have a standard error of 0.01 m, which represents inseparable errors due to measurement, atmospheric pressure changes, as well as deviation of a real, heterogeneous aquifer from an ideal aquifer of infinite lateral extent, uniform thickness and homogeneous properties. Variations of the assumed error by a factor of two or three does not adversely affect the parameter estimates but it does change the values of chi squares, chi-square probability, post-processing covariance matrix, and parameter resolutions (Lee, 1999, p. 346). Those standard-error-dependent values were computed but are not tabulated here. Any drawdown less than the standard error is excluded from modeling. This 0.01-m cutoff is set low enough to retain a sufficient number of data points for modeling. A computed drawdown less than 0.01 m is also excluded because

Table 2
Wide and narrow ranges of parameters used in genetic algorithm to yield $\mathbf{p}^{guess}$

		T (m^{2}s^{-1})	S	b (m)	K_z/K_r	S/S_y
Wide	Low	1.00×10^{-6}	1.00×10^{-7}	18	0.01	1.00×10^{-7}
	High	1.00×10^{-1}	1.00×10^{-2}	30	2.00	1
Narrow	Low	1.67×10^{-4}	3.33×10^{-4}	18	0.10	3.33×10^{-3}
	High	1.50×10^{-3}	3.00×10^{-3}	22	0.90	3.00×10^{-2}

Table 3
Modeling results for synthetic drawdown data

Model	T	S	b	K_z/K_r	S/S_y	rms A	rms B	rms C	rms E
Input	6.53	1.21	23.2	0.059	2.66				
SS1B2									
Gen	5.18	1.00	21.5	0.083	1.84		**0.009**		
Inv1	5.19	1.04	22.3	0.084	1.84	0.017	**0.007**	0.006	0.009
Inv3	5.23	1.08	23.2	0.087	1.84	0.021	**0.007**	0.006	0.010
Error	−20%	−11%	0%	+47%	−69%				
SS2B2									
Gen	6.44	1.09	22.7	0.060	2.30	**0.009**		&	
Inv1	6.22	1.15	22.3	0.061	2.30	**0.005**	0.006	&	0.005
Inv3	6.14	1.19	22.3	0.062	2.30	**0.005**	0.007	&	0.005
Error	−6%	−2%	−4%	5%	−14%				

T transmissivity in 10^{-4} m^2 s^{-1}, S storativity in 10^{-3}, b aquifer thickness in m, K_z/K_r vertical-to-horizontal conductivity ratio, S/S_y storativity-to-specific yield ratio in 10^{-3}. The input parameters for the synthetic data are the output of model L2B4 in Table 4. Model SS1B2 and SS2B2 represent one- and two-well models, respectively. Model Gen is the output of genetic algorithm while Inv1 and Inv3 are respectively the output of stage-1 and 3 inverse modeling. Bold-faced rms indicates the well has been used in modeling; and an & means the marked well has also been used in modeling. Errors are percent deviation of Inv3 from the respective parameters used to generate synthetic data. Italicized model names indicate that the fitting is also presented graphically.

the Stehfest's inverse Laplace transform may yield erratic results at early time (Tseng and Lee, 1999). Any set of model parameters that produces an rms around 0.01 m is deemed acceptable from the viewpoint of rms alone.

The ranges of parameters for searching $\mathbf{p}^{\text{guess}}$ with the genetic algorithm are listed in Table 2. There are two sets of ranges: wide and narrow. The wide ranges for T, S, and S/S_y cover several orders of magnitude. The vertical-to-horizontal conductivity ratio is allowed to range by a factor of 100. The range for the aquifer thickness is set between 18 and 30 m that bracket the reported thickness of 21 m. For the narrow ranges, the aquifer thickness is bracketed between 18 and 22 m and other parameters scan over a factor of 9 from the low to the high limits. Depending on the seeding of the random number generator, desirable results can be obtained from both the wide and narrow ranges, of which the latter are marked with an asterisk in Table 4. Once the $\mathbf{p}^{\text{guess}}$ is determined by the genetic algorithm, the range limits are relaxed from the three-staged inverse modeling.

Based on geological information, the test aquifer is presumed to be a Neuman's type of unconfined aquifer, which can be represented with different levels of uncertainty by hundreds or more sets of model parameter values. As used here, each set is a model. Numerous models can be generated from the drawdown data at one well or any combination of wells. Models based on one-well data are herein categorized as one-well models; and a two-well model means that the set of model parameter values is obtained from drawdown data at two observation wells. In the listing of model identifications (file names) in Tables 3 and 4 and labeling of figures, the first numeral after the leading alphabetic symbols signifies the number of observation wells used to construct the model. Other symbols are for filename usage only. An observation well used for modeling is marked with a bold-faced rms (also, marked with ampersand for multiple-well models) in Tables 3 and 4 or with solid-model curves in Figs. 2–10. Extrapolation rms is marked in light-faced font and extrapolation curves are dashed. Extrapolation or interpolation fitting serves as a means to validate the model parameters for a homogenous aquifer, which is the case for the synthetic data. For the cases of real data, however, misfit by extrapolation or interpolation may be compounded by the aquifer heterogeneity. It is noted that the absolute differences between computed and observed (or synthetic) values are on the average greater at locations with larger drawdowns. Hence, rms cannot be compared from well to well;

Table 4
Modeling results for field data

Model	T	S	b	K_z/K_r	S/S_y	rms A	rms B	rms C	rms E
L1A1*	3.58	1.02	19.9	0.173	1.61×10^{-3}	**0.007**	0.037	0.029	0.018
L1A2	5.15	1.31	23.1	0.103	2.46×10^{-3}	**0.006**	0.024	0.015	0.006
L1B1*	7.68	0.895	20.7	0.048	5.23×10^{-3}	0.041	**0.008**	0.013	0.006
L1B2	2.59	0.483	19.9	0.249	7.89×10^{-4}	0.055	**0.005**	0.020	0.018
L1B3	5.28	0.742	19.8	0.086	1.86×10^{-2}	0.025	**0.007**	0.013	0.008
L1C1*	7.41	1.06	23.8	0.054	4.21×10^{-3}	0.019	0.022	**0.007**	0.008
L1C2*e*	6.10	1.04	23.6	0.070	3.01×10^{-5}	0.017	0.030	**0.007**	0.006
L1C3e	6.43	1.06	23.5	0.017	1.18×10^{-5}	0.017	0.030	**0.007**	0.007
L1E1 *	10.3	0.541	20.5	0.039	2.60×10^{-3}	0.087	0.047	0.035	**0.002**
L1E1*e*	9.23	0.160	21.8	0.050	1.60×10^{-1}	0.224	0.217	0.106	**0.007**
L1E3*e*	13.7	1.02	25.9	0.028	8.28×10^{-5}	0.087	0.032	0.034	**0.005**
G1m	4.78	0.818	21.0	0.087	3.19×10^{-3}				
L2A1	4.71	0.946	21.1	0.104	1.69×10^{-3}	**0.016**	&	0.015	0.013
L2B1	6.23	1.23	23.3	0.065	1.24×10^{-3}	**0.009**	0.033	&	0.008
L2B4	6.53	1.21	23.2	0.059	2.66×10^{-3}	**0.008**	0.034	&	0.007
L2E4*	7.10	0.854	19.8	0.032	4.97×10^{-3}	0.032	**0.011**	&	0.060
L2E7	5.04	0.827	22.8	0.104	2.17×10^{-3}	0.022	**0.009**	&	0.010
G2m	5.85	0.999	21.9	0.067	2.27×10^{-3}				
L3A1*e*	5.98	1.18	25.5	0.090	3.79×10^{-5}	**0.015**	&	&	0.012
L3B1	5.04	0.930	20.4	0.087	4.77×10^{-4}	**0.014**	&	0.012	&
L3C1*e*	6.99	0.942	21.6	0.055	7.93×10^{-6}	0.035	**0.011**	&	&
L3D1	7.49	1.38	26.7	0.059	3.63×10^{-3}	**0.008**	0.032	&	&
G3m	6.14	1.13	23.3	0.072	1.32×10^{-3}				
G*G*m	5.94	0.962	22.1	0.074	2.12×10^{-3}				
L4A1*	6.40	1.15	25.7	0.087	4.47×10^{-3}	**0.011**	&	&	&
L4A2*	6.08	1.09	24.4	0.088	3.99×10^{-3}	**0.011**	&	&	&
L4A3	4.85	0.885	19.6	0.086	3.13×10^{-4}	**0.014**	&	&	&
L4A4	5.85	1.01	22.7	0.082	2.61×10^{-3}	**0.011**	&	&	&
G4m	5.76	1.03	23.1	0.086	1.95×10^{-3}				

Units and symbols are footnoted at Table 3 except for S/S_y. Model G*n*m designates geometric means for *n*-well models and G*G*m designates geometric means of G*n*m for $n = 1, 2, 3$. Results for italicized model names are plotted in Figs. 4–10. Models marked with symbol * are run within narrow ranges of parameters (Table 2) for the genetic algorithm to yield $\mathbf{p}^{guess}$. Parameter values for model names ended with an *e* are excluded from the computation of geometric means because of poor fitting at late time. Bold-faced rms indicates the well has been used in modeling; and an & means the marked well has also been used in modeling.

at the same well, however, the rms can be compared for different models.

5.2. Modeling with synthetic data

A set of drawdown data at five observation wells was calculated with the aquifer parameter values listed in the 'input' row of Table 3. These parameters were obtained from one example of modeling the field data, as described later. Random errors up to ± 0.01 m were imposed on the calculated drawdown to form the synthesized data. The modeling presented here is a recast because the method had been tested with other synthetic data before its being applied to field data. In comparison with modeling the real data, this recast for an idealized aquifer not only tests the methodology but also provides a glimpse of heterogeneity in a real aquifer. Note that well E is farther away from the pumping well than well D is but, like the observed data trend, it has experienced greater drawdown because of different screening depths and intervals.

5.2.1. Models based on one observation well

Initiated with a seed of random number, the output

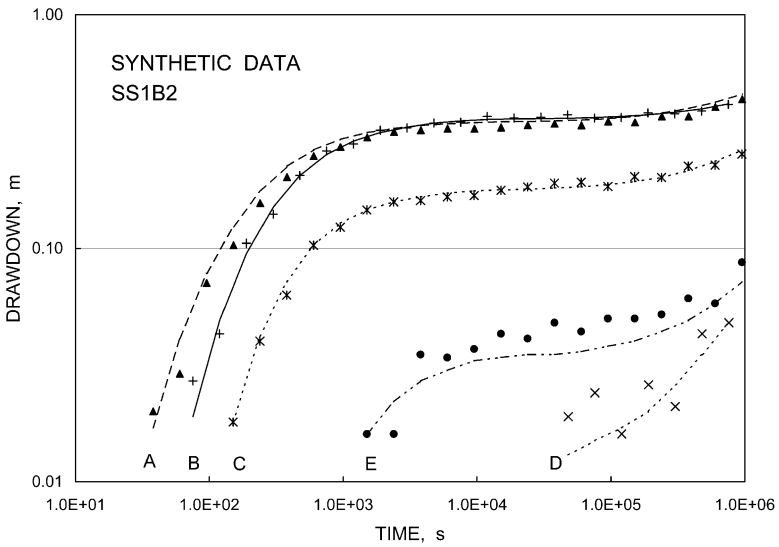

Fig. 2. Model curve (solid) based on synthetic drawdown at well B (crosses). Dashed curves represent extrapolations based on model parameter values (Table 3, SS1B2 Inv1). Note good extrapolation fitting at well C but poor elsewhere.

of the genetic algorithm has a rms of 0.009 m (submodel headed by 'Gen', Table 3). The output of submodel Gen is used as input $\mathbf{p}^{\text{guess}}$ to the first-stage inversion modeling, which yields submodel Inv1. The rms for submodel Inv1 is 0.007 m (column rms B, bold-faced font) while the rms for the extrapolations range from 0.006 to 0.017 m (rms C, E, and A, light-faced font). The name of submodel Inv1 is italicized to indicate that the curve fitting is graphically displayed. As shown in Fig. 2, the solid curve signifies that the model is based on drawdown at well B and the dashed curves represents extrapolation, based on a well-B model, to other wells. By keeping stationary one parameter of submodel Inv1 which is associated with the largest parameter resolution as defined in Eq. (14) and inputting the rest as $\mathbf{p}^{\text{guess}}$, a second-stage inversion output is produced. Then, the output of the second-stage inversion (unlisted) generates the third-stage inversion, submodel Inv3. The errors in the parameter estimations at Inv3 are tabulated as percentage deviations from the input values, which are used to construct the synthetic drawdown. Those errors range from -69% for S/S_y to $+47\%$ for K_z/K_r, despite the results that the model rms is on par with the imposed random error for the synthetic drawdown.

As shown in Fig. 2, model misfits at well B (solid curve) are symmetric in the sense that there are about equal numbers of positive and negative misfits and that the positives and negatives are more-or-less evenly distributed around the model curve. Listed in Table 3 (model SS1B2), the changes in parameter values from stage-1 to 3 inverse modeling are insignificant (a few percent). These attributes constitute good fitting but the errors in parameter estimates are unacceptably large.

Misfit distribution for the genetic output (not shown) is typically biased or skewed in comparison to the distributions for subsequent inversion. Submodel Inv1 is chosen for the plotting, instead of Inv3, because the extrapolation-misfit rms at wells A and E are slightly but insignificantly better although the model thickness for submodel Inv3 duplicates exactly the input thickness, 23.2 m (note that the aquifer thickness for the synthetic data is 2.2 m greater than the real thickness). Excepting this example, results of submodel Inv3 for other modeling with the synthetic or observed data are always taken as the final results.

Extrapolation fitting at nearby well C appears satisfactory (Figs. 1 and 2) but exhibits systematically biased distribution. Toward the pumping well, extrapolation overestimates the drawdown (well A) while outward from it, underestimation prevails

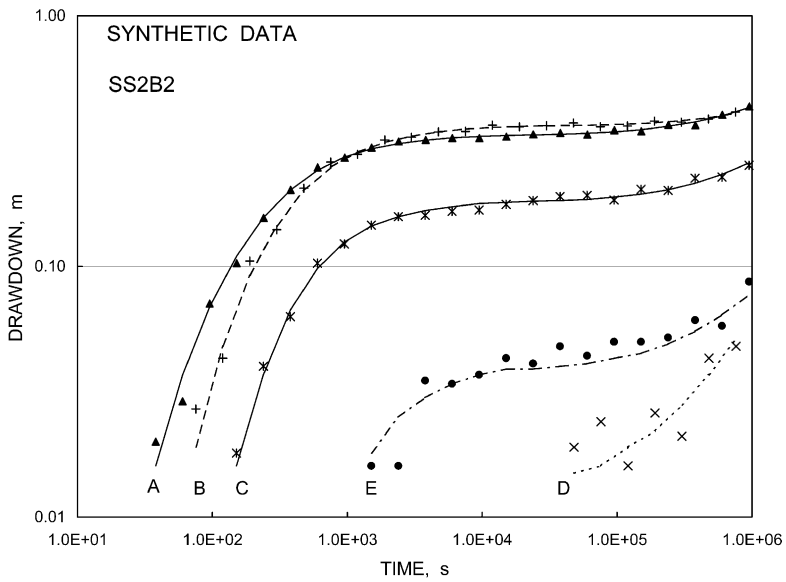

Fig. 3. Model curves (solid) based on synthetic drawdown at wells A (triangles) and C. Dashed curves represent extrapolations based on model parameter values (Table 3, SS2B2 Inv3). Note the improvement in model and extrapolation misfits from the two-well model over the one-well model (Fig. 2).

(wells D and E). At well D, the extrapolation curve tracks the trend of synthetic data but because of the graphical amplification at small numbers in a log–log plot, the extrapolation misfits seem disproportionately greater as compared to other wells.

5.2.2. Models based on two observation wells

The results of modeling the drawdowns at wells A and C together are also listed in Table 3 (Model SS2B2). Again, the parameter values resulted from the sequence of genetic-inverse modeling are fairly

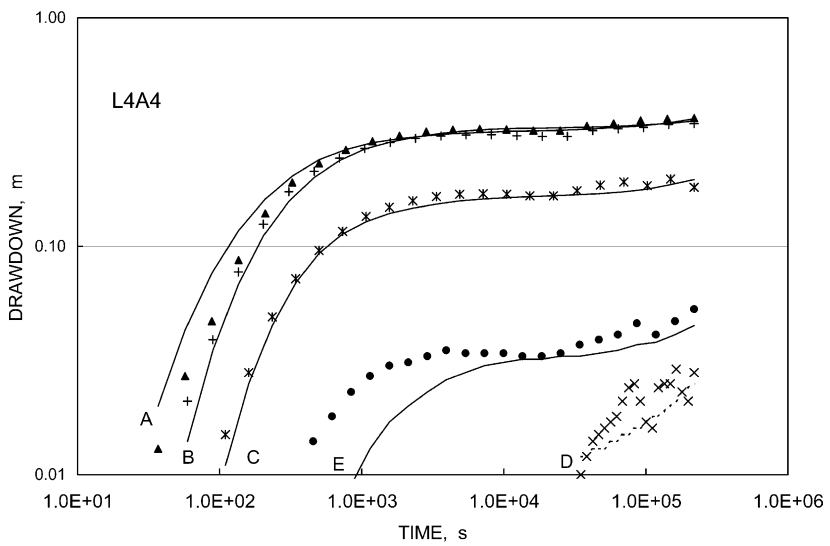

Fig. 4. Model curves (solid) based on observed drawdown at four wells. Dashed curve represents extrapolation to well D based on the model parameter values (Table 4, L4A4). All drawdowns less than 0.01 m are excluded from modeling and plotting. Note the biased misfit distribution at well E. At early time, overestimates of drawdown at well A (triangles) are compensated by underestimates at wells B and C.

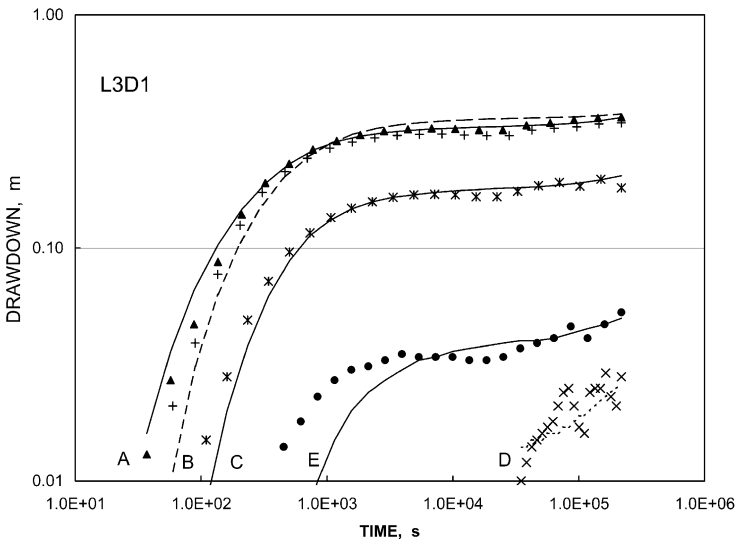

Fig. 5. Model curves (solid) based on observed drawdown at three wells. Dashed curves represent extrapolation to wells B and D based on model parameter values (Table 4, L3D1). Note the biased distributions of model misfit at each well but the overall distributions are unbiased; for example, at early time, overestimates of drawdown at well A (triangles) are compensated by underestimates at well C.

consistent. Misfits for submodel Inv3 are shown in Fig. 3. The improvement in fitting by the two-well model can readily be seen by comparing Figs. 2 and 3, especially for wells D and E. The extrapolation fitting at well B (actually, an interpolation between wells A and C) is visibly as good as the model fitting in Fig. 2.

Errors in parameter estimates are greatly reduced for the two-well model. The error in conductivity ratio K_z/K_r is reduced from 47% for the single-well model to 5% for the two-well model. For the transmissivity, it improves from −20 to −6%. The aquifer thickness is within 4% of the input value. However, the error in the estimate of the ratio of storativity to specified yield (S/S_y) at −14% for the two-well model is still uncomfortably large.

5.2.3. Discussion

Large errors in the one-well model for the synthetic data suggest that parameter determination based on good curve fitting alone should still be viewed with caution. Some parameters may not be resolvable because the computed drawdowns are insensitive to parameter variations within the confine of data-time window. For example, the ratio S/S_y is poorly determined as evidenced by the fact that the S/S_y remains unchanged from submodels Gen to Inv3 (Table 3). Poor resolution in S/S_y is expected because the data recording ended shortly after the leveled (delayed) drawdown started to rise again. Its estimate can be improved by extending the recording time. The improvement in the estimate of K_z/K_r for the two-well model also suggest that drawdown data collected from different screen intervals at different distances from the pumping well can help to better resolve parameter estimates.

One dozen sets of parameters have been obtained for the one- and two-well models. In one case of one-well modeling, starting at an initial model thickness (b^{guess}) of 25.2 m as outputted by the genetic algorithm, the inverse modeling improves iteratively the rms from 0.044 to 0.023 m but the final model thickness is inflated to 37.3 m. This set is unacceptable because of the relatively large rms and excessive changes in value through the three stages of inverse modeling. Apparently, in this case, the genetic algorithm did not provide a close-to-true set of $\mathbf{p}^{\text{guess}}$.

In another case of one-well modeling, the final model thickness is 30.3 m at a rms of 0.007 m as started from an initial value b^{guess} of 27.2 m. Such a small rms is usually indicative of good model fitting. However, poor extrapolation fittings at other wells raise an alarm of potential pitfalls in

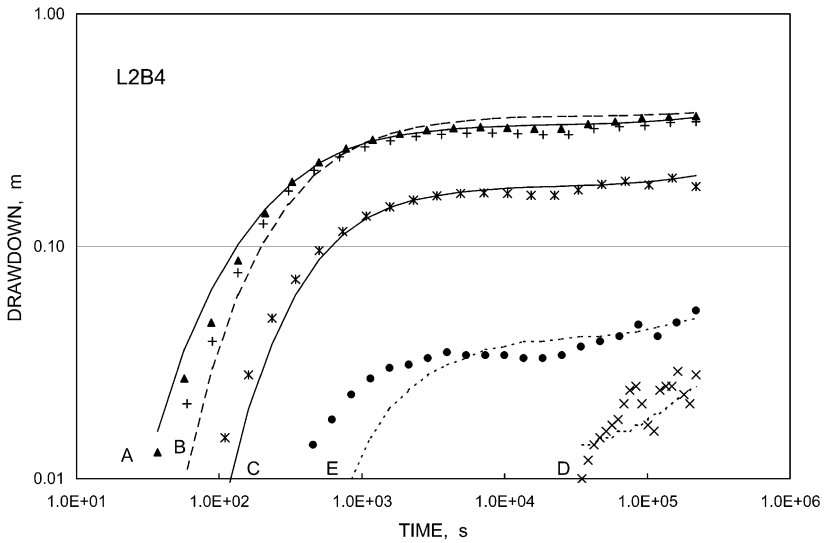

Fig. 6. Model curves (solid) based on observed drawdown at well A (triangles) and C. Dashed curves represent extrapolations based on model parameter values (Table 4, L2B4), which are also used to generate synthetic data for Figs. 2 and 3. Note that the misfit distribution for the real data is inferior to that for the synthetic data (Fig. 3), implying inhomogeneity in the real aquifer.

inverse modeling. Should that well be a lone monitoring well for field data and were there no constraint on the aquifer thickness, the final **p** could have been accepted as the aquifer parameters.

These two examples for one-well modeling are extreme but similar misleading results can and will happen. It demonstrates the need to use the like of genetic algorithm to generate and apply more sets of $\mathbf{p}^{guess}$ as a remedy for the shortcoming of relying on a single monitoring well. The case shown in Fig. 2 demonstrates that some aquifer parameters can be estimated to within 20% from one single-well pumping test. Most can be

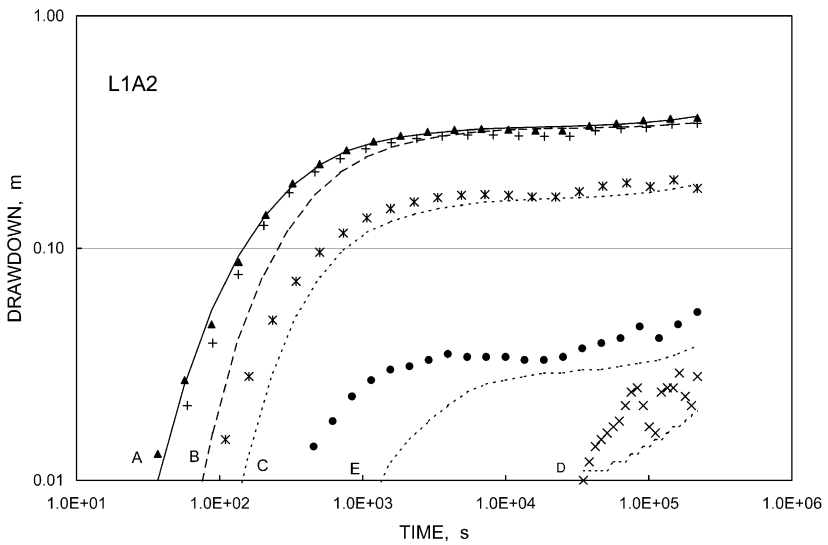

Fig. 7. Model curve (solid) based on drawdown at well A (triangles). Dashed curves represent extrapolation based on model parameter values (Table 4, L1A2). Despite good model fitting at well A, extrapolation fits poorly.

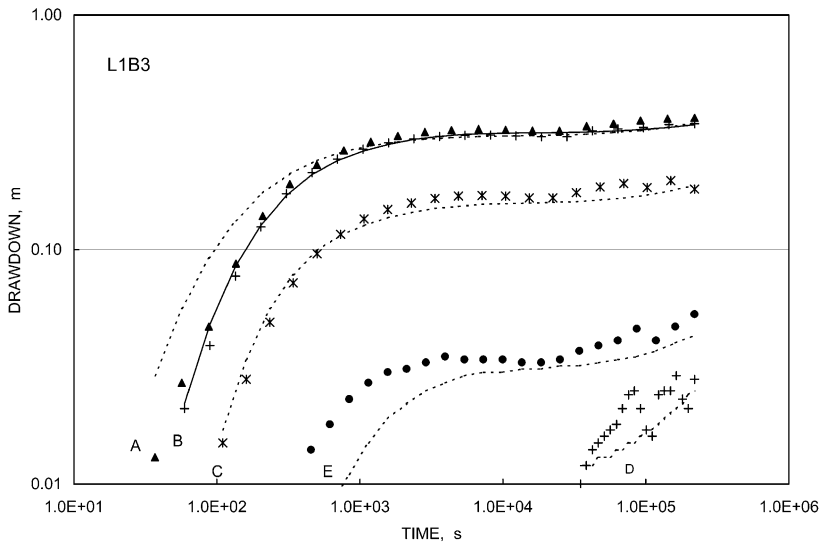

Fig. 8. Model curve (solid) based on drawdown at well B (crosses). Dashed curves represent extrapolation based on model parameter values (Table 4, L1B3). Note poor extrapolation fittings.

estimated to within 10% for a two-well test (Fig. 3). However, the 10% goal may not be reachable for one or two parameters if the record length is short or the well locations and screen intervals are improperly configured even in a homogenous aquifer. Optimal well configurations are being investigated by us to minimize the uncertainty of parameter determination.

5.3. Modeling with field data

Modeling results of field data are summarized in Table 4. The plotted models (ones with italicized names) do not necessarily represent the best in fitting; some are plotted for the convenience of discussion. An asterisked model name indicates the model employed narrow ranges of parameter values to find $\mathbf{p}^{\text{guess}}$ with the genetic algorithm (Table 2). The following presentation starts from models based on the drawdown at four observation wells and ends at models based on one observation well. Well D has been excluded for modeling because its drawdown readings are low and will contribute relatively little to the misfit rms. Nevertheless, well D is used for testing the extrapolated trend of the time-drawdown curves. The rows headed by G*n*m in Table 4 designate geometric means of parameter values, where *n* is the number of wells used in the modeling. Another row, G*G*m, designates the geometric means of G1m, G2m, and G3m. A model name affixed with a character *e* is excluded from the computation of geometric means because those models yield poor fitting at late time, caused mainly by too high or too low a determined value of S/S_y.

5.3.1. Models based on four observation wells

As shown in Fig. 4, well C at the mid-range distance to the pumping well yields the best fitting among the L4 series (Table 4). For time less than 5000 s, the model overestimates the drawdown at well A but underestimates it at well B; whereas the reverse trend of over- or under-estimations is true at later time. The model underestimates the drawdowns at the far distant well E. At the next distant well D, model extrapolations follow the trend of the noisy time-drawdown data but misfit distribution is biased.

Model L4A3 is associated with a slightly worse rms (0.014 m instead of 0.011 m for the other L4-series models) but yields an S/S_y ratio that is less than the others by one order of magnitude. In addition, other parameters for L4A3 are comparatively smaller except for the K_z/K_r ratio. In comparison with the fitting in Fig. 3 for the synthetic data, the relatively poor fitting for the field data (Fig. 4, L4A4) suggests that aquifer inhomogeneity is a major cause of misfit. In case of significant aquifer inhomogeneity, the Neuman's water table model is inadequate to describe the drawdown.

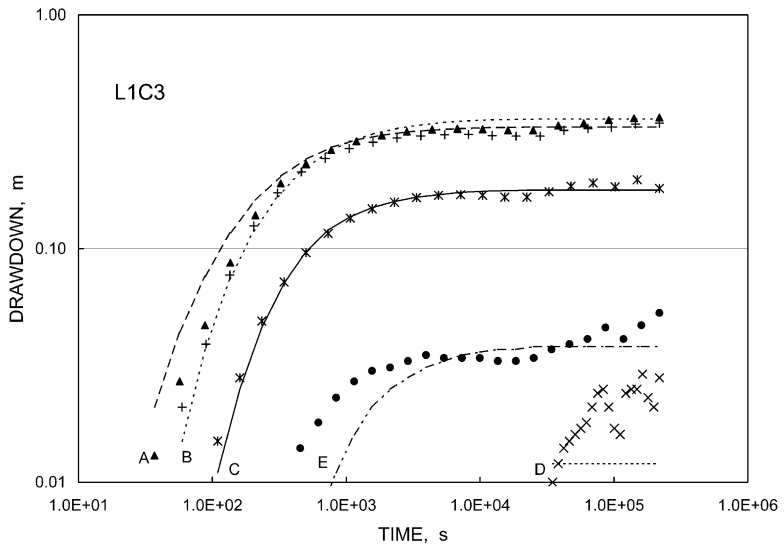

Fig. 9. Model curve (solid) based on drawdown at well C; extrapolation based on model parameter values (Table 4, L1C3). Poor misfits at late time are caused by a low S/S_y determination, as diagnosed with a flattened or delayed drawdown response. Triangles denote drawdown at well A.

5.3.2. Models based on two or three observation wells

The L3 series models in Table 4 indicate that, depending on which of the three wells are chosen, the modeled S/S_y ratios can differ by three orders of magnitude. Generally, a decrease in the S/S_y ratio will lengthen the delay time in drawdown response in a Neuman aquifer (Fig. 9). Model L3D1 (Fig. 5) provides the best model fitting among the four tabulated three-well models. Despite a three-order-of-magnitude, uncertainty in S/S_y, which is likely a consequence of inadequate record length as mentioned in the case of modeling synthetic data, other parameters vary by less

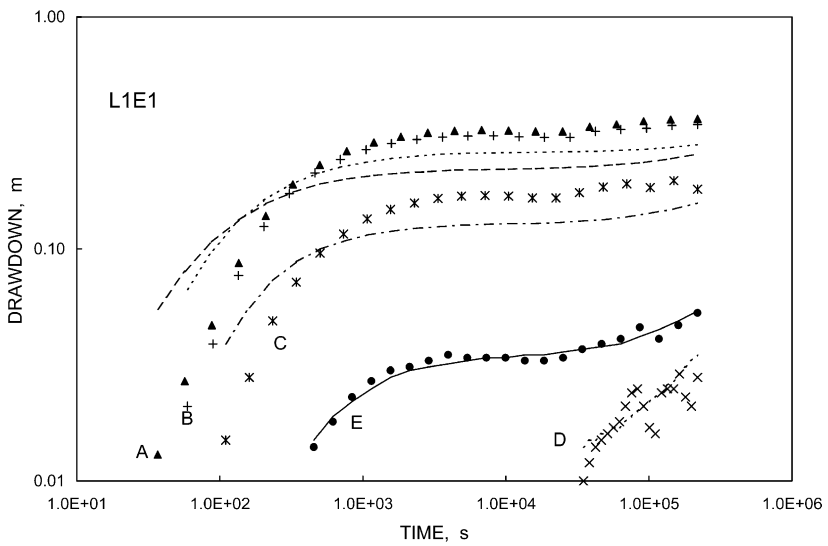

Fig. 10. Model curve (solid) based on drawdown at well E, where drawdown has been poorly predicted by model parameter values derived from other wells (Figs. 4–9). Extrapolation fitting based on model parameter values (Table 4, L1E3) is generally poor except at well D. Triangles denote drawdown at well A.

than 20% from their respective geometric means as listed in the row of G3m.

Among the five listed two-well models, parameters T, S, and b show less variations from model to model. However, the ratios of K_z/K_r or S/S_y can vary by a factor of two from their respective geometric means (row G2m). The model parameters for L2B4 (Fig. 6) were used to generate the synthetic data used for Figs. 2 and 3. The fitting for the synthetic drawdown (Fig. 3) is definitely superior to the fitting for the real data.

5.3.3. Models based on one observation well

Four of the 11 one-well models listed in Table 4 are depicted in Figs. 7–10. The representative model at each well exhibits good fitting but extrapolation fittings at other wells are generally poor.

The three models for well C (L1C series) exhibit two-order-of-magnitude variations in the S/S_y ratio despite the fact that all three have the same rms of 0.007 m and that the variations of transmissivity T, storativity S, and thickness b are less than 10% from their respective geometric means (not tabulated). These contrasts in model-to-model parameter variations lead to the conclusion that the S/S_y ratio and hence the specific yield S_y cannot be determined from the drawdown data alone at well C. Model L1C3 in Fig. 9 demonstrates the lengthy delay response in a Neuman aquifer at low S/S_y ratio. The flattened portion of the L1C3 model curve does not fit well with the late-time data ($\geq 10^4$ s); nor the model parameters yield a time-drawdown trend that resembles the data trend at well D, all because of an erroneously determined low S/S_y ratio. In this L1C3 example, the misfit distribution rather than rms is clearly discriminatory against the model results. A better misfit distribution can be achieved by narrowing the range of S/S_y ratio to initiate $\mathbf{p}^{guess}$ with the genetic algorithm, as done in model L1C1 (illustration omitted).

As seen from Figs. 4–9, all model and extrapolation fittings are poor at the far-distant well E. Nevertheless, a reasonably good fitting can be obtained at well E itself but this set of model parameters (model L1E1, Fig. 10) yields poor extrapolation fittings at other wells except for well D.

5.3.4. Discussion

Based on rms, the 24 models listed in Table 4 reveal that the model transmissivity T varies from the lowest to highest values by a factor of 2, storativity S by 3, aquifer thickness b by 1.3, conductivity ratio K_z/K_r by 10, and the storativity-to-specific yield ratio S/S_y by an astonishing factor of 20,000. Some criteria other than rms must therefore be invoked to reduce their ranges of uncertainty. It is noted that all models associated with high ($>10^{-2}$) or low ($<10^{-4}$) S/S_y ratios do not fit well with the late-time data (Fig. 9). Excluding those models (model names ended with an *e*), consistent results emerge as shown by the geometric means for each parameter of the one-, two-, three-, and four-well models (rows G*n*m, for $n = 1–4$).

The geometric means of the respective parameters in G1m, G2m, and G3m are listed at row GGm of Table 4. The parameters in this GGm set are remarkably close to the respective geometric means for the four-well models (G4m), being less than 7% from their respective arithmetic means. Because each well is not used at equal frequency to obtain the geometric means, except the L4 series of which each is equally represented by four observation wells, the various geometric mean values as obtained here do not necessarily represent the properties of a heterogeneous aquifer. Nevertheless, these geometric means can reasonably represent the aquifer and their variations are the uncertainties inherited from a non-ideal aquifer, data error, plus parameter resolution in our methodology.

The wellbore skin effect, if significant, will vary from well to well systematically. We do not see any set of drawdown data or combination of data set that yields consistently higher or lower parameter values. The uncertainty cited above is not of systematic error. Failure to see the effect, if any, stems from the fact that the early-time or distant data have high noise-to-signal ratios as typified by the drawdown at wells D and E, and consequently the effect is overshadowed by the noises. Also, the early time data that may reflect the skin effect represents only a fraction of the total data used for the modeling and any possible effect could have been masked. If the skin effect is to be included in the modeling, two or more parameters (e.g. skin thickness, conductivity) will be needed for every well, including the pumped well. Presumably, adding more model parameters can improve the fitting. However, without consideration of wellbore

skin or storage effects, the model rms is already on par with the data error. To improve misfit rms further by including the skin effect cannot be justified unless information independent of drawdown data is available to constrain the modeling, the data quality is improved, and the test aquifer is indeed ideal.

6. Conclusions

Our method for the determination of aquifer parameters involve using a genetic algorithm to find a set of close-to-true parameters $\mathbf{p}^{guess}$, then inputting the $\mathbf{p}^{guess}$ to a three-staged quasi-linear inverse algorithm for refinement toward the final answers. A model is deemed acceptable if its misfit rms is near the standard error of the drawdown data and the misfit distribution is symmetric around the model time-drawdown curve. The selection of a representative set of parameters from many sets of acceptable models are based on two criteria: (1) the model aquifer thickness which is treated as an unknown parameter in the genetic-inverse modeling must agree with the true thickness to within a few percent if the truth is known, and (2) extrapolations based on the model parameters must fit drawdown data elsewhere if the aquifer is homogeneous and uniform in thickness within the zone of pumping influence.

The making of $\mathbf{p}^{guess}$ is a random process subject to the survival constraint of minimal rms through selective combinations of parameter retention, exchange, perturbation, and recombination. Therefore, a $\mathbf{p}^{guess}$ may not be necessarily close to the true $\mathbf{p}^{true}$ or the global minimum in rms in the parameter space. Such situations can be detected by an rms that is significantly greater than the standard error of data or a misfit distribution that is asymmetric around the model curve. Our three-staged inversion scheme (the middle stage keeps stationary the parameter with the highest model resolution) is another safe guard for excluding undesirable models in the sense that if a $\mathbf{p}^{guess}$ is close to the $\mathbf{p}^{true}$, the final $\mathbf{p}$ in the iterative inverse modeling should be close to that $\mathbf{p}^{guess}$. The scheme is effective for a parameter that can sensitively perturb the drawdown (e.g. thickness in our test examples).

However, a computed $\mathbf{g}$ (the theoretical drawdown Δh) may not be sensitive to slight change in one or more parameters for a given set of observed data. As shown in Table 3, the S/S_y ratio remains stationary as the modeling proceeds from submodels Gen to Inv3. In such cases, the scheme of three-staged inversion is not effective to resolve S/S_y. Misfit rms and distribution are therefore indispensable criteria. Like many other non-linear inversion schemes, it is prudent to run the genetic-inverse algorithm a few times with different random-number seedings in order to assure a final $\mathbf{p}$ with the global minimum in rms.

Our method has been demonstrated with synthetic and observed drawdown data for five partially screened monitoring wells in a Neuman-type water-table aquifer. In our testing, models with acceptable goodness of fitting can always be obtained for one single monitoring well. Some of those yield poor extrapolation fitting and their model aquifer thicknesses do not match with the known thickness, or in the case of synthetic data, model parameters deviate significantly from the true values. Since the locations and screen intervals of our monitoring wells are not designed for an optimal determination of aquifer parameters, many runs of the genetic-inverse modeling are needed to ensure a final $\mathbf{p}$ that yields acceptable model- and extrapolation-fitting.

Comparing Fig. 2 for a one-well model, with Fig. 3 for a two-well model, the inclusion of one additional well certainly improves the model and extrapolation fitting with the synthetic data. For the real, observed data, however, the fitting improvement is not as great (from Figs. 7 to 6). For the two-well model (L2B4) in Fig. 6, the model fitting at well A or C alone is biased, especially at the early time. These biases complement each other and hence, in the senses of least-squares curve fitting, the model misfits are symmetric around zero if the misfits are lumped together and plotted as a function time. So, model L2B4 is considered to have made good fitting but extrapolation fitting is biased at wells B and E. Assuming the methodology has been validated by synthetic data, such biases are indicative of aquifer inhomogeneity. Similar misfit results appear in Figs. 4 and 5, respectively, for the four- and three-well models, re-affirming lateral variations in aquifer properties.

Depending on the severity of aquifer inhomogeneity, one single set of parameter values may not be

representative of the aquifer. In an attempt to obtain a representative set for our test aquifer, we have used the geometric means to summarize the parameter values for various combinations of monitoring wells (Row G*n*m for *n* from 1 to 4, Table 4). Row G*G*m (Table 4), which are geometric means of G1m, G2m, and G3m, agree with row G4m to within 7% from their arithmetic means. Any set of the geometric means can be representative of the aquifer to within 10–15%. Each of the geometric mean for the model aquifer thickness is definitely within 10% of the given aquifer thickness, 21 m, as estimated from borehole logs.

Acknowledgements

We are grateful to two anonymous reviewers whose comments improved the manuscript.

References

Bohling, G.C., McElwee, C.D., 1992. SUPRPUMP: an interactive program for well test analysis and design. Ground Water 30, 262–268.

Cassiani, G., Kabala, Z.J., 1998. Hydraulics of a partially penetrating well: solution to a mixed-type boundary value problem via dual integral equations. Journal of Hydrology 211, 100–111.

Cassiani, G., Kabala, Z.J., Medina, M.A., 1999. Flowing partially penetrating well: solution to mixed-type boundary value problem. Advances in Water Resources 23, 59–68.

Chen, X., Ayers, J.F., 1998. Aquifer properties determined from two analytical solutions. Ground Water 36, 783–791.

Chen, X., Goeke, J., Summerside, S., 1999. Hydraulic properties and uncertainty analysis for an unconfined alluvial aquifer. Ground Water, 845–854.

Cox, B.F., Hillhouse, J.W. 2000. Pliocene and Pleistocene evolution of the Mojave River and associated tectonic movement of the Transverse Ranges and Mojave Desert, based on borehole stratigraphy studies near Victoville, California. US Geological Survey, Open-File Report 00-147, 66 pp.

Hudak, P.F., 1993. Effects of improper characterization of aquifer thickness on estimates of hydraulic conductivity from pumping tests. Ground Water Monitoring Review, 113–117.

Kabala, Z.J., Cassiani, G., 1997. Well hydraulics with Weber–Goldstein transforms. Transport in Porous Media 29, 225–246.

Lee, T.-C., 1999. Applied Mathematics in Hydrogeology, Lewis Publishers, New York.

Lee, C.-Y., Ma, L., Antonsson, E.K., 2001. Evolutionary and adaptive synthesis methods. In: Antonsson, E.K., Cogan, J. (Eds.), Formal Engineering Design Synthesisit, Cambridge University Press, Cambridge.

Moench, A.F., 1995. Combining the Neuman and Boulton models in a homogeneous, anisotropic water table aquifer. Ground Water 33, 378–384.

Moench, A.F., 1996. Flow to a well in a water-table aquifer: an improved Laplace transform solution. Ground Water 34, 593–596.

Neuman, S.P., 1972. Theory of flow in unconfined aquifers considering delayed response of the water table. Water Resources Research 8, 1031–1044.

Neuman, S.P., 1974. Effects of partial penetration on flow in unconfined aquifers considering delayed gravity response. Water Resources Research 10, 303–312.

Neuman, S.P., 1975. Analysis of pumping test data from anisotropic unconfined aquifers considering delayed gravity response. Water Resources Research 11, 329–342.

Novakowski, K.S., 1989. A composite analytical model of analysis of pumping tests affected by well bore storage and finite thickness skin. Water Resources Research 25, 1937–1946.

Nwankwor, G.I., Gillham, R.W., van der Kamp, G., Akindunni, F.F., 1992. Unsaturated and saturated flow in response to pumping of unconfined aquifer: field evidence of delayed drainage. Ground Water 30, 690–700.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vettering, W.T., 1986. Numerical Recipes: the Art of Scientific Computing, Cambridge University Press, New York.

Sibbett, B.S., 1999. Pleistocene channels of the Mojave River near Victoville, California. San Bernardino County Museum Association Quarterly 46, 65–68.

Stehfest, H., 1970. Numerical inversion of Laplace transform. Communications of the Association of Computing Machinery 13, 47–49.

Tarantola, A., 1987. Inverse Problem Theory, Elsevier, New York.

Tseng, P.-H., Lee, T.-C., 1998. Numerical evaluation of exponential integral: Theis well function approximation. Journal of Hydrology 205, 38–51.