ELSEVIER

Journal of Hydrology 243 (2001) 120–132

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Wellskins and slug tests: where's the bias?

C.W. Rovey II[a,*], W.L. Niemann[b]

[a]*Department of Geography, Geology, and Planning, Southwest Missouri State University, 901 S. National, Springfield, MO 65804, USA*
[b]*University of Dubuque, 2000 University Avenue, Dubuque, IA 52001, USA*

Received 2 November 1999; revised 19 October 2000; accepted 15 November 2000

Abstract

Pumping tests in an outwash sand at the Camp Dodge Site give hydraulic conductivities (**K**) approximately seven times greater than conventional slug tests in the same wells. To determine if this difference is caused by skin bias, we slug tested three sets of wells, each in a progressively greater stage of development. Results were analyzed with both the conventional Bouwer–Rice method and the deconvolution method, which quantifies the skin and eliminates its effects.

In 12 undeveloped wells the average skin is +4.0, causing underestimation of conventional slug-test **K** (Bouwer–Rice method) by approximately a factor of 2 relative to the deconvolution method. In seven nominally developed wells the skin averages just +0.34, and the Bouwer–Rice method gives **K** within 10% of that calculated with the deconvolution method. The Bouwer–Rice **K** in this group is also within 5% of that measured by natural-gradient tracer tests at the same site. In 12 intensely developed wells the average skin is < -0.82, consistent with an average skin of -1.7 measured during single-well pumping tests.

At this site the maximum possible skin bias is much smaller than the difference between slug and pumping-test **K**s. Moreover, the difference in **K** persists even in intensely developed wells with negative skins. Therefore, positive wellskins do not cause the difference in **K** between pumping and slug tests at this site.

Keywords: Hydraulic conductivity; Ground water movement; Slug tests, Tracer tests, Wellskins

1. Introduction

The cause of the common difference between hydraulic conductivity (**K**) measured by slug tests (lower) and pumping tests (higher) is a matter of contention. Some conclude that the "scale effect" (or increase) reflects a real increase in effective **K** (Neuman, 1994; Rovey and Cherkauer, 1995; Tidwell and Wilson, 1997; Rovey, 1998; Schulze-Makuch and Cherkauer, 1998; Schulze-Makuch et al., 1999). Others feel that the differences are artifacts of persistent test bias (e.g. Butler and Healey, 1998a,b). This bias is largely attributed to incomplete well development and the concomitant presence of low-conductivity wellskins, a zone of altered **K** around the wellscreen or borehole. Any such skin would bias small-scale slug tests toward artificially low **K**s but would scarcely affect the larger-scale pumping-test results. In this paper we address the artifact hypothesis by (1) measuring the skin and **K** for three sets of wells, each in a different stage of development, (2) comparing slug-test **K**s computed with the Bouwer–Rice method with those from the "deconvolution method" (Peres et al., 1989) which removes skin effects, and (3) comparing slug-test **K**s with values obtained from natural-gradient tracer tests. The results from this site

* Corresponding author. Fax: +1-417-836-6006.
E-mail addresses: charlesrovey@smsu.edu (C.W. Rovey II), wniemann@dbq.edu (W.L. Niemann).

PII: S0022-1694(00)00408-X

Nomenclature

C_D	"Dimensionless storage" coefficient used in generating type curves incorporating borehole storage effects, equal to $[2S]^{-1}$
C_N	Nominal value of the wellbore storage constant. For an open well equals area within the wellscreen
C_w	The actual wellbore storage constant. If the developed zone extends beyond the screen, C_w is greater than the nominal value C_N
FH	Falling-head slug test
H_D	Dimensionless head during a slug test. Equal to measured drawdown at a given time divided by the initial displacement H_0
H_0	Initial displacement during a slug test
K	Aquifer hydraulic conductivity
RH	Rising-head slug test
R_{eff}	The effective wellbore radius. In the case of a negative skin, the effective radius is greater than the nominal radius (r_{nom})
R_i	Radius of influence
R_{nom}	Nominal radius. For these wells, equal to the radius of the wellscreen
S	Aquifer storativity
σ	Skin. A positive value indicates lower **K** around the borehole caused by drilling disturbance. A negative value indicates higher **K** around the borehole, generally due to aggressive development

demonstrate the limitations of invoking skin bias as a universal explanation for the differences in **K** between slug and pumping tests.

2. Background

2.1. Scale and dimensionality

One explanation for the consistent difference between slug and pumping-test results is that small- and large-scale tests tend to average heterogeneity differently. Utilizing digital models, Rovey (1998) showed that mean **K** measured in a stressed well under transient radial flow increases with the radius of influence (R_i), reflecting an upward shift toward the arithmetic mean of the **K** field. Over short flow distances (small R_i) most flow is across heterogeneities; thus, the measured **K** is analogous to the lower vertical **K** in layered media. With increasing R_i, however, flow is progressively concentrated along high-conductivity heterogeneities. Therefore, **K** measured by larger (pumping) tests is analogous to the greater horizontal **K** of layered systems.

The same phenomenon also affects steady-state radial flow. Desbarats (1992) and Neuman and Orr (1993) simulated steady radial-convergent flow in synthetic heterogeneous media. The measured **K** around nodal positions, as determined by spatial gradient or distance-drawdown techniques, decreases systematically upon approaching the wellbore, reducing measured **K** to values just above the harmonic mean of the nodal values for 2D simulations. The strongly convergent flow increases drawdown because flow cannot bypass low-**K** zones and is forced across low-conductivity heterogeneity. At large distances, however, flowlines are sub-parrallel, following high-conductivity paths, and the measured **K** equals the larger geometric mean of the nodal values (or approaches the arithmetic mean in a 3D system).

These model results demonstrate that convergent flow in heterogeneous media produces a *pseudoskin* manifested by increased total drawdown near the wellbore, above that predicted from **K** measured in either a more distant observation well or from the later time-rate of drawdown in the pumping well. In the latter two cases the measured drawdowns reflect flow conditions distant from the wellbore (e.g. Earlougher, 1977). This pseudoskin may have the same net effect as a true skin, but is conceptually

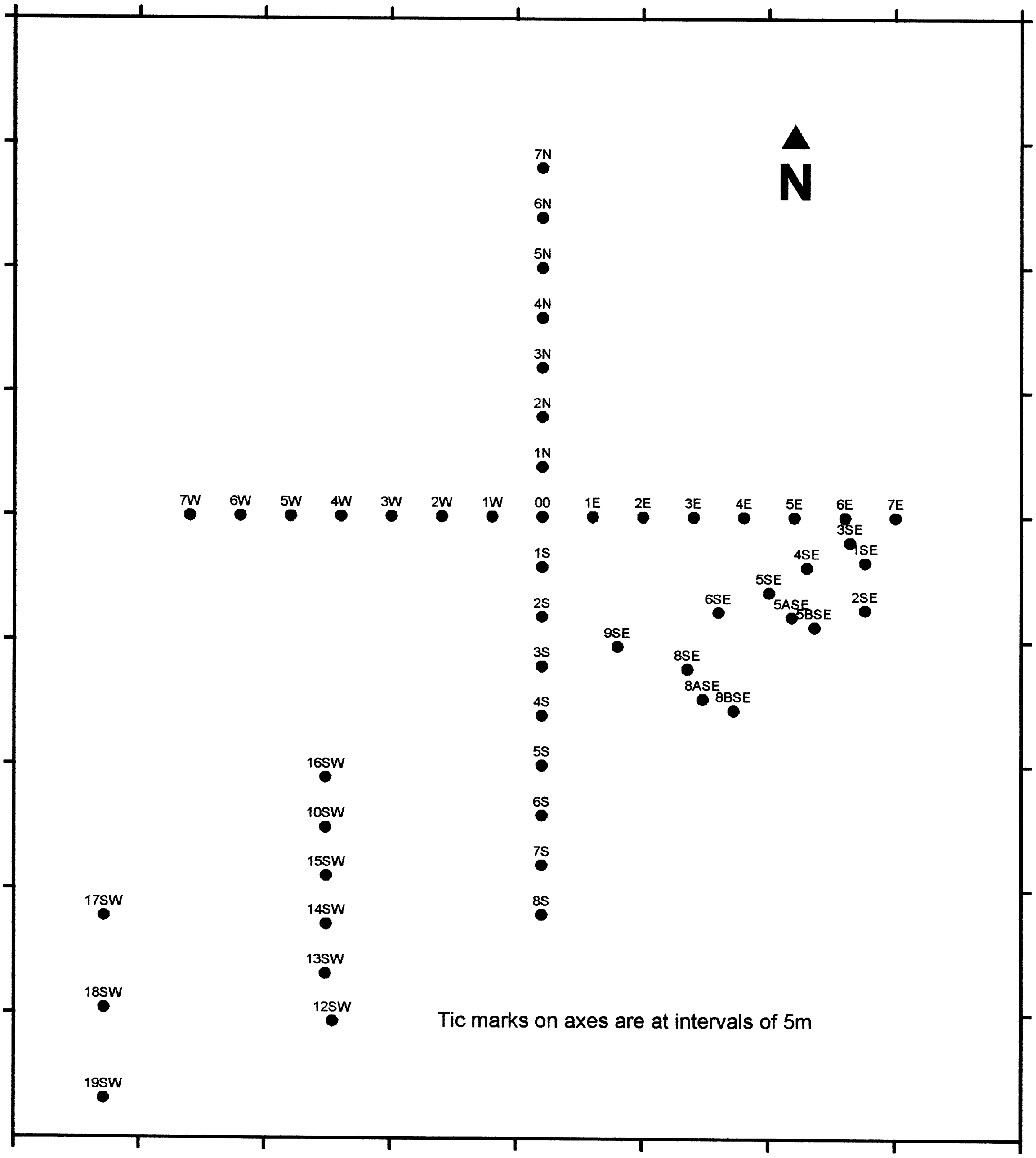

Fig. 1. Location of test wells. Wells with N, S, E, or W suffixes were installed initially for slug and pumping tests. Some of these wells were subsequently reinstalled as “SE” and “SW” wells to help monitor the natural-gradient tracer tests and to provide detailed measurements of the potentiometric surface.

Table 1
Summary of radial-convergent tests. Average **K**s are geometric means, numbers in parentheses are standard deviations in log cycles. Slug-test results (Bouwer–Rice method) are for 28 falling and 27 rising-head tests in 28 individual wells using a developed radius of 13.5 cm (see text)

Method	# Tests	Avg. **K** (cm/s)	Range, **K** (cm/s)	C_w (cm^2)	Skin	Developed radius (cm)
Slug	55	0.018 (0.14)	0.010–0.035			
Pumping	6	0.12 (0.36)	0.072–0.20			
Skin test	3			78–137	−1.5 to −1.8	12–15

distinct, particularly regarding its origin and implications for hydraulic testing.

The scale effect can also be related to dimensionality of the test or flow system. Small-scale tests typically have closer boundary constraints, which induce flow systems of lower dimensionality. As the distance to boundaries increases in a heterogeneous medium, so does the dimensionality, as flow is increasingly free to follow any higher-**K** pathways which may exist. For example, the effective **K** under 1D flow (e.g. long permeameter samples) is a harmonic mean of the **K** field. Under 2D flow, however, the effective **K** increases to the geometric mean, while under fully 3D conditions, the effective **K** increases to a value between the geometric and arithmetic means (see for example the stochastic equations of Gelhar, 1993; Neuman, 1994 and simulation results of Deutsch, 1989; Desbarats, 1987). Tracer tests measure the velocity of a solute's center of mass (a point) along a linear flowpath, but with an additional degree of freedom in the vertical plane. Hence, tracer tests are effectively 2D, and should measure the geometric mean of the **K** field. Radial tests with a large R_i (pumping) are 3D. As the R_i shrinks to that of a slug test, however, the measurement becomes increasingly 2D. Therefore, if this reasoning is correct, **K**s measured by slug and tracer tests should be similar to each other, and both should be smaller than those measured by pumping tests, independently of any skin bias.

2.2. Wellskins

Slug-test results may be biased by multiple factors, and usually any bias is presumably toward lower **K** (Zlotnik, 1994; Hyder and Butler, 1995; Butler et al., 1996). The greatest potential bias is a true wellskin, a zone of altered **K** immediately surrounding the wellbore caused by drilling disturbance. Small-scale radial tests (slugs) typically have R_i between 1 and 2 m (Rovey, 1998), and hence, measure the **K** immediately around the borehole. Thus, slug tests analyzed with conventional methods are biased toward the skin **K**. Large-scale pumping tests, however, measure **K** over a much larger area, and are, therefore, insensitive to wellskins.

The skin's numeric value is a dimensionless constant which reflects an additional (or reduced in the case of a negative skin) drawdown under a constant pumpage rate (van Everdingen, 1953; Ramey, 1982). Positive wellskins indicate formation damage (lower **K**) around the screen or wellbore. Negative values imply aggressive development and higher **K** around the borehole relative to the formation background, because fines have been flushed out of the formation, locally increasing porosity and **K** around the wellbore (e.g. Fig. 15.3 in Driscoll, 1986).

3. Site and procedures

The data presented below were collected at the Camp Dodge Site near Des Moines, Iowa USA (see Niemann and Rovey, 2000). The tested aquifer is thin (~2 m) confined glacial outwash sand and gravel. Initially 29 5-cm-diameter wells were installed at 2-m intervals along two perpendicular lines (Fig. 1). To minimize the zone of disturbance, drilling was completed with a small (7.6 cm) diameter solid stem auger. After reaming the hole, screens were pushed through the natural slurry to near the base of the outwash, and any portion of the borehole remaining open was allowed to collapse around the wellscreen. The screens are typically 1.5 m in length with aspect ratios >50. Thus vertical anisotropy (another possible bias) would have only small effects on slug-test results (Fig. 3 in Hyder and Butler, 1995).

Twelve of the initial 29 wells were slug tested

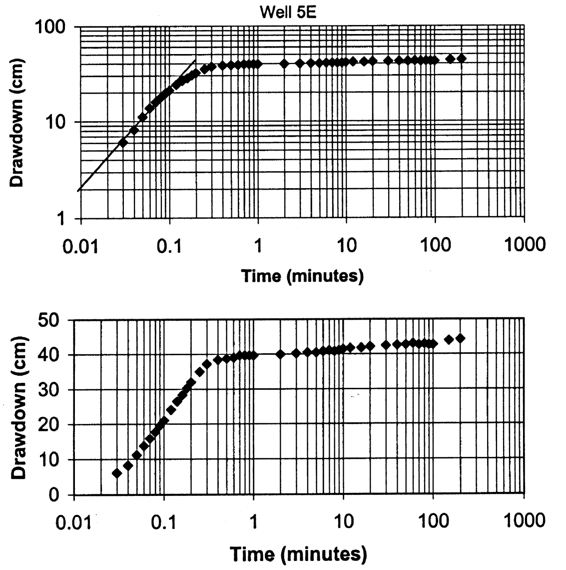

Fig. 2. Single-well pumping test, well 5E. (a) Log–log plot. Reference line shows a log–log unit slope. The unit slope at early time corresponds to the period of pure wellbore storage. Drawdown at time = 0.04 min is 8.2 cm. Discharge = 1.6×10^4 cm^3/min, so Δv = 640 cm^3. C_w = 640 cm^3/8.2 cm = 78 cm^2. (b) Semilog plot. The drawdown curve reaches a constant semilog slope within one log cycle after the end of pure wellbore storage. Δs per cycle = 2.2 cm; pumpage rate = 1.6×10^4 cm^3/min, b = 170 cm. **K** = 2.3(1.6×10^4 cm^3/min)/4π 170 cm(2.2 cm) = 7.8 cm/min(0.13 cm/s).

before any development whatsoever. Following these initial measurements all 29 wells were intensely developed immediately prior to a second round of slug and then pumping tests in 28 and six of these wells, respectively. Development was completed by twice surging along the wellscreen for several minutes with a conventional surge block. After both rounds of surging, the well was pumped until clear while moving the intake line along the entire screen. The slug and pumping tests in these intensely developed wells are the basis for the overall slug and pumping-test results summarized in Table 1.

Pumping tests (Niemann and Rovey, 2000) were completed with a small-diameter submersible pump which reaches a steady discharge in <1 s. After the pumping tests, additional wells were installed along the direction of hydraulic gradient to serve as monitoring wells for natural-gradient tracer tests. These last wells were developed only moderately, by bailing approximately 20 l, at which time the water had begun to clear, but was still somewhat turbid. Thus, these wells were developed to an intermediate degree between those in the first two sets. Seven of these moderately developed wells were slug tested prior to solute injection to measure **K** directly along the path of solute migration. The conservative solute velocities, combined with measured hydraulic gradient and porosity (Niemann, 1999; Niemann and Rovey, 2000), also provide another set of **K**s, independent of wellskins.

4. Conventional results

4.1. Slug and pumping tests

All results summarized below were obtained within an area smaller than 35×35 m^2 (Fig. 1). The highest slug-test **K** is approximately one-half the lowest pumping-test **K**, and the average slug value is approximately one-seventh of the pumping-test mean (Table 1). This difference in magnitude between the slug and pumping tests illustrates the issue at hand. Is the difference due to a true scale dependency or to skin bias?

Wellskins were first estimated from measurements completed during three single-well pumping tests. For a pumped well the earliest portion of the drawdown curve reflects properties of the borehole, rather than the formation. Thus, under certain circumstances, the wellbore storage constant (C_w) may provide an estimate of skin:

$$C_w = \Delta V/\Delta H = Q\Delta t/\Delta H \quad (1)$$

where C_w is the wellbore storage constant (l^2), ΔV = change in fluid volume in the borehole (l^3) = volume of pumpage, ΔH = change in water level within borehole (l), Q = pumping rate (l^3/t), and Δt = elapsed time of pumpage (t). Both ΔV and ΔH must be measured during the earliest portion of the test when all pumpage is derived from wellbore storage, as marked by a log–log unit slope in the drawdown curve (Fig. 2).

For open wells with positive or zero skin, C_w is simply the cross-sectional area within the screen. If the skin is negative, however, the functional borehole radius is greater than that of the screen, and C_w > the

Table 2
Summary of natural-gradient tracer-tests. **K**s are given along segments of the flow path between successive detection points. **K** is calculated from tracer velocities with Darcy's Law using independent measurements of porosity and hydraulic gradient. The cumulative **K** for both tests (avg. = 0.0078 cm/s) is calculated independently of the segment **K**s using a time-averaged gradient over the entire test duration. Slight differences between values in Table 1 and those in Rovey and Niemann (1998) are due to using more accurate hydraulic gradients, and average aquifer thickness in converting transmissivity to **K**

Test	Distance (m)	**K** (cm/s)
Tracer 1	2	0.020
	15	0.0090
	12	0.011
	Cumulative (29)	0.010
Tracer 2	8	0.0070
	16	0.0060
	Cumulative (24)	0.0060

nominal wellbore storage (C_n). The larger coefficient implies that the developed zone is much more permeable than the background formation, so that the developed zone functions as part of the borehole. Other methods may be used to estimate skins (Earlougher, 1977) which involve comparison between measured drawdowns and theoretical values corresponding to measured transmissivity. However, these methods actually give pseudoskins, because they do not distinguish extra drawdown caused by true skins from that produced by radial-convergent flow and other factors. Therefore, we utilize C_w measurements, noting that any well with $C_w > C_n$ must have a negative skin.

The three tested wells have C_w ranging from 78 to 137 cm^2 (Table 1), considerably larger than the C_n of approximately 20 cm^2. Therefore, all three wells have negative skins. Given a measured value of porosity (Niemann, 1999; Niemann and Rovey, 2000) C_w can be converted to an effective borehole radius, which is then related to skin:

$$r_{eff} \approx r_{nom} e^{-\sigma} \tag{2}$$

where r_{eff} is the effective borehole radius, r_{nom} is the nominal radius of the screen, and σ is the value of skin (Earlougher, 1977). The calculated r_{eff} (or developed radius) ranges from 12 to 15 cm, corresponding to skins between −1.5 and −1.8 (Table 1). Thus, the developed zone extends well beyond any distance possibly disturbed by an auger with a 3.8-cm radius.

Initial results confirm that the intense development produced negative skins. C_w, and hence skins, are obtained from the first few seconds of the pumping test (Fig. 2), and moreover, flow was completely clear over the duration of the test with no accumulation of sediment at the well bottom. Therefore, the negative skins must have been produced during the preceding development, not during the pumping tests themselves.

4.2. Tracer tests

Results are given in Table 2 for two conservative (Cl^-) natural-gradient tracer tests. The flow distance through any possible wellskin (around injection and detection wells) is $\ll$ the total flow distance, except for the shortest (2 m) test. If solute velocities were slowed by positive wellskins, this bias would diminish with distance as time within the skin becomes insignificant relative to the total time of migration. However, the **K**s do not increase with travel distance. Considering the intense development, the resulting negative skins, and the close agreement between slug and tracer-test **K**s (both quasi-2D), skins are an unlikely cause of the difference between slug and pumping-test **K**s.

5. Deconvolution method

The deconvolution method of slug-test analysis (Peres et al., 1989) is highly recommended by those questioning a true scale dependence of **K** (Butler, 1998; Butler and Healey, 1998b). An initial convolution (Peres et al., 1989; Butler, 1998) converts the slug-test response into a dimensionless drawdown which would be produced in the same well by constant pumpage (Fig. 3). Thus, the procedure gives results equivalent to a short pumping test with a radius of influence equal to that of the original slug test.

The convolved response is that for a pumped (not observation) well. Therefore, the early-time response is dominated by wellbore storage (Fig. 3b), and a unique type-curve match is nearly impossible. To help obtain a unique match between field and type curves, drawdown and derivative-drawdown curves,

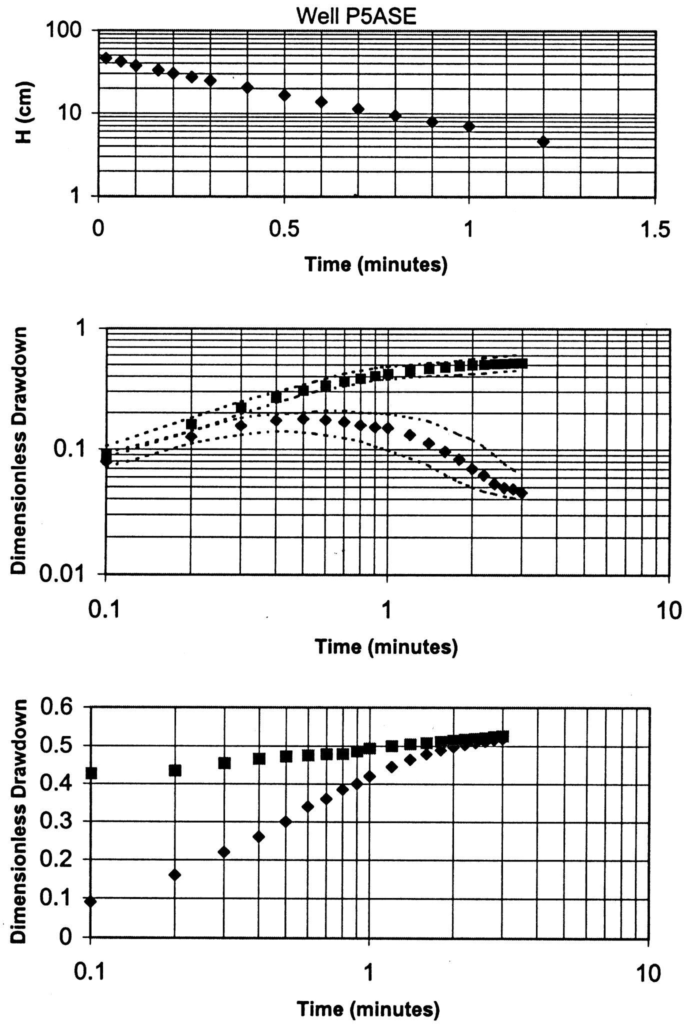

Fig. 3. Slug test results, well P5ASE. (a) Original response. The recovery curve (rising head) gives H_0 indistinguishable from the theoretical value of 47 cm. Bouwer–Rice **K** = 0.0024 cm/s. (b) Convolved response, log–log plot. The upper curve is the dimensionless drawdown; the lower curve is the corresponding derivative with respect to natural log of time. Dashed lines show type curves for $C_D e^{2\sigma} = 10^6$ and 10^4 (Bourdet et al., 1983). Match with field data gives $C_D e^{2\sigma} = 10^5$, and skin = +0.35. (c) Dimensionless drawdown, semilog plot. The upper curve is the deconvolved response, without wellbore storage effects. The lower is the convolved response which includes wellbore storage effects. **K** = 2.3(r^2)/4$b\Delta s$, which is the semilog equation for dimensionless drawdown (Butler, 1998). Δs per cycle = 0.076, b = 227 cm, r = 2.54 cm, **K** = 2.3(2.54 cm)^2/4(227 cm[0.076/min]) = 0.22 cm/min(0.0036 cm/s).

are matched simultaneously. Both drawdown and derivative type curves are plotted for values of the dimensionless grouping $C_D e^{2\sigma}$, where $C_D = [2S]^{-1}$ (for an open well), S = aquifer storativity, and σ = skin. Therefore, given storativity, the match between field and type curves also yields the skin value:

$$\sigma = 1/2 \ln\{2S(C_D e^{2\sigma}_M)\} \quad (3)$$

where $C_D e^{2\sigma}_M$ is the value of the type curve matched to the field data, and all other parameters are as previously defined. Note that large errors in storativity or the match value have only small effects on calculated skins.

Table 3
Slug-test results for a subset of wells. Results are for those wells which were slug tested both before any, and after, intense development. Bouwer–Rice **K**s for the postdevelopment tests are given for two radii. The developed radius of 13.5 cm is calculated from the single-well pumping tests, and the nominal radius is that of the wellscreen, 2.5 cm. “sw” denotes a single-well pumping test, “mw” multiwell pumping test. The pumping-test **K** for well 2S is calculated from drawdown measurements in that well; the pumped well was 4 m distant. A blank indicates that the specific test or analysis was not completed

Slot	Well	Predevelopment tests				Postdevelopment tests				Pumping tests
		Bouwer–Rice		Deconvolution		Bouwer–Rice developed *R*		Bouwer–Rice nominal *R*		
		Falling	Rising	**K**	Skin	Falling	Rising	Falling	Rising	
10	3N	0.011	0.0015	0.0018	0.35	0.011	0.016	0.018	0.027	
25	4N	0.005	[a]	0.0074	− 0.8	0.012	0.015	0.02	0.025	
10	5N	0.0079	0.0097	0.035	2.6	0.014	0.019	0.022	0.032	
25	6N	0.0022	[a]	0.0057	0.4	0.012	0.012	0.02	0.019	0.17(sw)
25	7N	0.0087	0.0071	0.033	18					
25	2S	0.019	[b]			0.017	0.016	0.028	0.028	0.19(mw)
10	3S	0.0052	0.0069	0.014	2.6	0.028	0.016	0.047	0.027	
25	4S	0.013	0.023	0.069	6.1	0.025	0.026	0.043	0.045	
10	5S	0.021	0.023	0.023	1.5	0.016	0.023	0.028	0.04	
25	6S	0.0066	0.011	0.04	5	0.026	0.024	0.046	0.041	0.09(mw)
10	7S	0.02	0.026	0.08	6.1	0.013	0.03	0.022	0.052	
10	8S	0.019	0.027	0.049	2.6	0.015	0.03	0.025	0.051	
Geometric mean		0.0095	0.011	0.021		0.016	0.02	0.027	0.033	0.14
Standard deviation			0.41	0.51			0.14			
Arithmetic mean					+ 4.0					

[a] Nonideal effects preclude calculating an RH **K**. Deconvolution **K** calculated from FH test.
[b] Nonideal effects (see text) preclude calculating RH **K** or deconvolution **K** from either RH or FH tests.

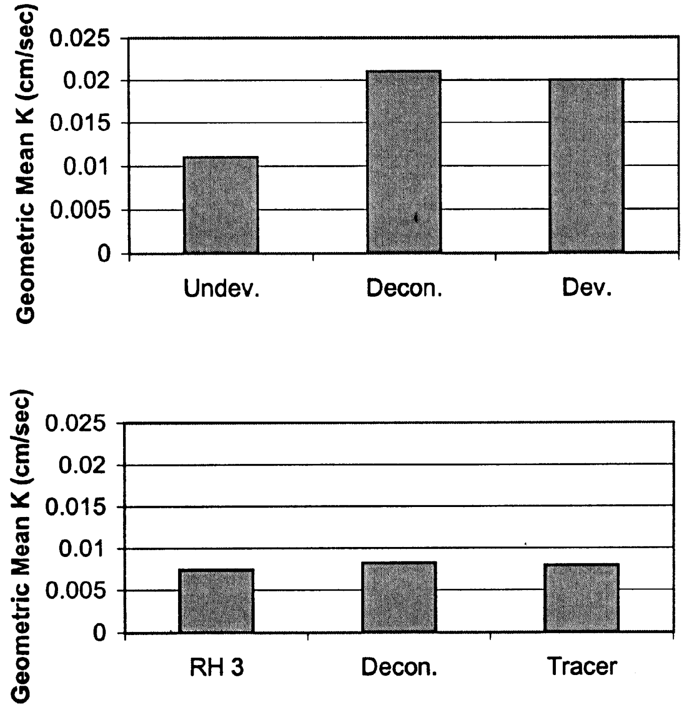

Fig. 4. Comparison of mean hydraulic conductivities. (a) Slug tests completed in identical wells prior to and after intense development. “Undev.” refers to rising-head tests conducted in undeveloped wells and analyzed with the Bouwer–Rice method (Table 2). “Decon.” refers to the deconvolution values obtained from the same undeveloped wells. “Dev” refers to rising-head tests after development analyzed with the Bouwer–Rice method. (b) Tracer tests and slug tests conducted along the path of plume migration. “RH 3” refers to rising-head tests conducted in nominally developed wells and analyzed with the Bouwer–Rice method; “Decon.” refers to deconvolution values obtained from the same tests. “Tracer” is the value obtained from measured velocities of two conservative tracer tests, combined with measured hydraulic gradient and effective porosity.

A unique match, however, still requires data over several log cycles, generally to times coinciding with a dimensionless head (H_D) < 0.01 during the original slug test. Because most equipment does not accurately measure a slug test’s H_D below this ratio, a further step is required. A second (deconvolution) procedure removes the wellbore storage effects so that an earlier portion of the response can be analyzed with the semilog time-drawdown method (Fig. 3c). This calculation is unaffected by wellskins, because it is based on the slope of the time-drawdown curve (not the total drawdown), which rapidly becomes independent of any true skin (e.g. Earlougher, 1977). Finally, the vertical match point for the convolved drawdown curve (Fig. 3b) is calculated using the transmissivity obtained from the semilog method. Given the match-point value, the field curves can then be matched to type curves, and hence skin. By utilizing these steps, our field data could be matched to values of $C_D e^{2\sigma}$ within a factor of approximately 2. To calculate skins, S was taken as 10^{-5} in all cases, based on eight pulse tests yielding values between 5.9×10^{-6} and 2.5×10^{-5}. Therefore, the expected accuracy of any individual calculated skin is approximately ± 1.5 (see Eq. (3)).

6. Undeveloped wells

Several slug tests (Table 3) within undeveloped wells had nonideal recovery (hardly surprising considering the total lack of development). Semilog recovery plots of these tests consist of multiple linear slopes interspersed between time intervals with little recovery. This behavior generally occurred during rising-head (RH) tests (flow into the well) and was apparently caused by intermittent clogging of the wellscreen. The most highly affected wells have 25-slot screens (Table 3), which is slightly too large for the grain-size distribution of this aquifer (Niemann, 1999) according to standard criteria (Driscoll, 1986). Likewise, three of the four largest skin values were present within the 25-slot wells. Apparently the larger 25-slot screens did not adequately stabilize the formation and initially allowed excess movement of aquifer material into the well during some of the RH tests, causing bridging against and perhaps within the screen.

The intermittent clogging noted above precluded calculating **K** for three RH tests (Table 3). A similar pattern was present to a lesser extent during the latest portions of two falling-head (FH) tests. In these two cases, however, **K** was obtained from the earlier well-behaved segment of the recovery curve.

FH tests (slug introduced into casing) displayed a second type of nonideal recovery. Over the earliest ~2.5 s measured heads generally oscillated above and below the theoretical initial head (H_0). In contrast, RH tests (slug removed from casing) displayed no discernible, or very limited, oscillation, and that over a much shorter duration (<1 s). Therefore, we generally use results of the RH tests for group comparisons (all statistical comparisons are at the 90% confidence level).

The deconvolution method utilizes a time derivative of the recovery curve and therefore is sensitive to nonideal behavior, even after such effects have apparently dissipated within the original recovery curve. The deconvolved drawdown values of those FH tests with the greatest early-time oscillation are numerically unstable, oscillating throughout the entire recovery period without reaching a consistent semilog slope. Therefore, deconvolution values were obtained from RH results, except for two tests in wells with 25-slot screens (Table 3) which had nonideal effects throughout the RH tests, but fortunately had relatively smooth recovery during the FH tests.

Test results for the undeveloped wells are listed in Table 3 as “predevelopment tests.” The deconvolution method gives a geometric mean of 0.021 cm/s, not statistically different from the overall mean of 0.018 cm/s for the intensely developed wells calculated with the Bouwer–Rice method (Table 1). Note the even closer agreement with the Bouwer–Rice RH mean of 0.020 cm/s, (“postdevelopment” RH tests, Table 3, Fig. 4a). Within the undeveloped subset, however, the Bouwer–Rice RH mean (0.011 cm/s) is statistically smaller than the deconvolution mean (0.021 cm/s), consistent with the positive skins (avg. of +4.0) measured by these tests. The Bouwer–Rice values for the undeveloped wells are indeed biased, on average by a factor of approximately 2. This bias toward a lower **K** is small, however, compared to the difference in magnitude between slug and pumping tests (Tables 1 and 3).

7. Nominally developed wells

The nonideal behavior during slug tests in undeveloped wells, along with the difference in skins between the 10- and 25-slot screens, suggests that the positive skins were largely caused by migration of aquifer material toward the well, not formation damage caused directly by drilling. Such skins should be relatively easy to eliminate during development, and this conjecture is confirmed by tests in the nominally developed wells (Table 4). The calculated skins range from −0.8 to +2.6 with an arithmetic average of just +0.34.

The Bouwer–Rice RH average for this second

Table 4
Slug-test results for nominally developed wells

Slot	Well	Nominally developed wells			
		Bouwer–Rice		Deconvolution	
		Falling	Rising	**K**	Skin
10	13SE	0.0041	0.0036	0.0062	1.5
25	16SE	0.018	0.012	0.012	0.4
25	17SE	0.0024		0.0021	− 0.8
25	18SE	0.016	0.011	0.012	− 0.8
10	5ASE	0.0032	0.0024	0.0036	0.4
10	8ASE	0.015	0.012	0.012	− 0.8
10	8BSE		0.012	0.032	2.6
Geometric mean		0.0072	0.0074	0.0082	
Standard deviation			0.34	0.39	
Arithmetic mean					+ 0.34

group (0.0074 cm/s) is not statistically different from the deconvolution mean (0.0082 cm/s) calculated from the same tests, even though the Bouwer–Rice values in two instances (wells 13SE and 8BSE) are slightly biased by small positive skins. The difference between the Bouwer–Rice and deconvolution values of the first (undeveloped) set of wells (the maximum possible skin bias) is smaller than the variability in deconvolution **K**s between the two sets.

The wells in this last set were installed along the path of solute migration. Results from these wells, therefore, are best for comparing slug and tracer-test **K**s (Fig. 4b). Neither the mean deconvolution **K** (0.0082 cm/s, Table 4) nor the RH Bouwer–Rice value (0.0074 cm/s) is statistically different from the tracer-test mean (0.0078 cm/s, Table 2). This close agreement supports the earlier conjecture that slug and tracer tests measure a similar 2D **K**, in contrast with the larger 3D **K** measured by pumping tests (see also Döll and Schneider, 1995; Muldoon and Bradbury, 1999).

8. Intensely developed wells

The majority of tests within the set of intensely developed wells (postdevelopment tests, Table 3) could not be analyzed with the deconvolution method. Most of these tests reached $H_D < 0.01$ (below the accuracy limits of the transducer) before the respective deconvolved values reached a constant semilog slope. In four such tests, however, the recovery lasted longer, allowing application of the deconvolution technique. In these four instances, the calculated **K**s are very close (within ~10%) to respective values measured during the earlier undeveloped tests in the same wells. The average skin calculated from these four tests is -0.82, confirming the presence of negative skins around most of the intensely developed wells. Although this skin value is larger (a smaller negative number) than that inferred from the single-well pumping tests (Table 1), it is actually an upper bound to the true average, necessarily being biased toward wells with a lower degree of development.

The average deconvolution value of 0.021 cm/s (predevelopment tests, Table 3) is virtually identical with that calculated for the same wells after development with the Bouwer–Rice RH method (0.020 cm/s), and using a developed radius of 13.4 cm. The nominal (2.5 cm) screen radius, which would apply if skins were not negative, produces Bouwer–Rice values approximately 80% greater (still significantly less than the pumping tests). Therefore, the developed radius calculated from the three single-well pumping tests is representative, confirming an aggregate negative skin of approximately -1.7 for the intensely developed wells.

9. Summary and conclusions

The question posed in the title of this paper can be answered in several ways. In one sense bias is in two directions. A true positive skin may indeed cause **K** to be underestimated with conventional slug-test analysis, but skins are not necessarily positive. As shown in this study, common development techniques easily produce negative skins which, if unrecognized, cause conventional slug-test **K**s to be *overestimated*.

In a second sense, skin bias is not present within the slug-test results originally presented for the Camp Dodge site (Rovey and Niemann, 1998). The difference between slug and pumping-test **K** is not an artifact of positive wellskins, because the true wellskins were negative, and the Bouwer–Rice calculations were corrected for this condition. Negative skins for these wells were measured independently with both pumping tests and the deconvolution technique of slug-test analysis. Moreover, the close agreement between slug and tracer-test **K**s cannot be reconciled with a skin bias affecting the slug-test results.

In a third sense, an extreme skin bias is unlikely for any well completed without mud in a sand and gravel aquifer. In jointed media, skins are naturally negative (McConnel, 1993), and in the porous outwash at Camp Dodge, statistically significant bias toward lower **K** was limited to wells in a completely undeveloped state. Even in this group, bias was relatively small, approximately a factor of 2. Very modest development eliminated virtually all skin effects.

Results here should be fairly general. The aquifer, screen lengths, and installation at the Camp Dodge site are comparable to those of other studies reporting a similar difference in **K** (e.g. Bradbury and Muldoon, 1990). The "bias" toward lower **K** typically obtained with slug tests, compared to pumping tests, is caused not by wellbore damage (a true skin), but by a pseudo-skin, a natural consequence of radial-convergent flow in heterogeneous media. Radial-convergent flow produces a quasi-2D flow system which averages heterogeneity geometrically, and therefore, gives a lower effective **K** than large-scale, 3D tests which provide values closer to the arithmetic mean of the **K** field. In final summary, conventional slug-test procedures are robust. With modest development, slug tests provide an accurate 2D **K**; the lower values obtained in heterogeneous media are simply a different type of **K** than that obtained from pumping tests.

Acknowledgements

This project was made possible through the help and cooperation of the Iowa Geological Survey Bureau; Art Bettis, University of Iowa; Mary Jones, Iowa National Guard; and John Niemann. Work was partially funded by Grant #1015-22-0896 Southwest Missouri State University Center for Graduate Studies. The authors thank Vincent Tidwell for critically reviewing an early manuscript draft.

References

Bourdet, D., Whittle, T.M., Douglas, A.A., Pirard, Y.M., 1983. A new set of type curves simplifies well test analysis. World Oil May, 95–106.

Bradbury, K.R., Muldoon, M.A., 1990. Hydraulic conductivity determinations in unlifthified glacial and fluvial materials. Ground Water and Vadose Zone Monitoring, Nielsen, D.M., Johnson, A.I. (Eds.). ASTM STP 1053, 138–151.

Butler Jr., J.J., 1998. The Design, Performance, and Analysis of Slug Tests. Lewis, London (252pp.).

Butler Jr., J.J., Healey, J.M., 1998a. Relationship between pumping-test parameters: scale effect or artifact?. Ground Water 36 (2), 305–313.

Butler Jr., J.J., Healey, J.M., 1998b. Discussion of papers: author's reply. Ground Water 36 (6), 867–868.

Butler Jr., J.J., McElwee, C.D., Liu, W., 1996. Improving the quality of parameter estimates obtained from slug tests. Ground Water 34 (3), 480–490.

Desbarats, A.J., 1987. Numerical estimation of effective permeability in sand/shale formations. Water Resources Research 23 (2), 273–286.

Desbarats, A.J., 1992. Spatial averaging of transmissivity in heterogeneous fields with flow toward a well. Water Resources Research 28 (3), 757–767.

Deutsch, C., 1989. Calculating effective absolute permeability in sandstone/shale sequences. SPE Formation Evaluation September, 343–348.

Döll, P., Schneider, W., 1995. Lab and field measurements of the hydraulic conductivity of clayey silts. Ground Water 33 (6), 884–891.

Driscoll, F.G. (Ed.), 1986. Ground Water and Wells 2nd ed.. Johnson Division, St. Paul, MN (1089pp.).

Earlougher Jr., R.C., 1977. Advances in well test analysis. Society of Petroleum Engineers Monograph 5.

van Everdingen, A.F., 1953. The skin effect and its influence on the productive capacity of a well. Petroleum Transactions, AIME 198, 171–176.

Gelhar, L.W., 1993. Stochastic Subsurface Hydrology. Prentice-Hall, New York (390pp.).

Hyder, Z., Butler Jr., J.J., 1995. Slug tests in unconfined formations: an assessment of the Bouwer and Rice technique. Ground Water 33 (1), 16–22.

McConnel, C., 1993. Double porosity well testing in the fractured carbonate rocks of the Ozarks. Ground Water 33 (1), 75–83.

Muldoon, M.A., Bradbury, K.R., 1999. Relationships of measurement scale to observed magnitude and variability of hydraulic conductivity in fractured dolomite. GSA Abstracts with Programs 31 (7), A213.

Neuman, S.P., 1994. Generalized scaling of permeabilities: validation and effect of support scale. Geophysical Research Letters 21(5), 349–452.

Neuman, S.P., Orr, S., 1993. Prediction of steadystate flow in non-uniform geologic media by conditional moments: exact non-local formalism, effective conductivities and weak approximation. Water Resources Research 29 (2), 341–364.

Niemann, W.L., 1999. Relationship between scale effects in hydraulic conductivity and scale effects in dispersivity: a case study. PhD dissertation, University of Missouri-Rolla.

Niemann, W.L., Rovey II, C.W., 2000. Comparison of hydraulic conductivity values obtained from aquifer pumping tests and conservative tracer tests. Ground Water Monitoring and Remediation 20 (3), 122–128.

Peres, A.M.M., Onur, M., Reynolds, A.C., 1989. A new analysis procedure for determining aquifer properties from slug test data. Water Resources Research 25 (7), 1591–1602.

Ramey Jr., H.J., 1982. Well-loss function and the skin effect: a review. Geological Society of America Special Paper 189, 265–271.

Rovey II, C.W., 1998. Digital simulation of the scale effect in hydraulic conductivity. Hydrogeology Journal 6, 216–225.

Rovey II, C.W., Cherkauer, D.S., 1995. Scale dependency of hydraulic conductivity measurements. Ground Water 33 (5), 769–780.

Rovey II, C.W., Niemann, W.L., 1998. Discussion of "Relationship between pumping-test and slug-test parameters: Scale effect or artifact?" by J.J. Butler Jr and J.M. Healey. Ground Water 36 (6), 866–867.

Schulze-Makuch, D., Cherkauer, D.S., 1998. Variations in hydraulic conductivity with scale of measurement during aquifer tests in heterogeneous, porous carbonate rocks. Hydrogeology Journal 6, 204–215.

Schulze-Makuch, D., Carlson, D.A., Cherkauer, D.S., Malik, P., 1999. Scale dependency of hydraulic conductivity in heterogeneous media. Ground Water 37 (6), 904–919.

Tidwell, V.C., Wilson, J.L., 1997. Laboratory method for investigating permeability upscaling. Water Resources Research 33 (7), 1607–1616.

Zlotnik, V., 1994. Interpretation of slug and packer tests in anisotropic aquifers. Ground Water 32 (5), 761–766.