Transp Porous Med (2007) 68:413–421
DOI 10.1007/s11242-006-9045-8

BRIEF COMMUNICATION

A discussion on the effect of heterogeneity on the onset of convection in a porous medium

D. A. Nield · Craig T. Simmons

Received: 21 February 2006 / Accepted: 10 August 2006 / Published online: 28 September 2006

Abstract The effect of heterogeneity on the onset of convection induced by a vertical density gradient in a saturated porous medium is discussed. The distinction between moderate and high heterogeneity of permeability is emphasized. Topics discussed include the limitations of the use of the log–normal distribution for permeability, the different effects of vertical and horizontal heterogeneity, and the use of an electrical resistor analogy applied to an illustrative quartered-square example. In a postscript, a new approximate criterion for instability applicable to the case of strong heterogeneity is proposed.

Keywords Heterogeneity · Stability · Natural convection

We would like to examine more closely the effect of heterogeneity on the onset of convection in a porous medium and to evaluate whether the classical analysis based on the concept of a Rayleigh number is valid for heterogeneous systems. To do this, we closely re-examine previous works by Simmons et al. (2001) and Prasad and Simmons (2003), amongst others, and then discuss important unresolved research challenges in what we believe to be a complex area of inquiry that is as yet far from fully described and understood.

Simmons et al. (2001) and Prasad and Simmons (2003) correctly point out that in many heterogeneous geologic systems, hydraulic properties such as the hydraulic conductivity of the system under consideration can vary by many orders of magnitude and sometimes rapidly over small spatial scales. Geologic systems, characterized by fractured rock environments or lenticular mixes of sand and clay, are common

D. A. Nield (✉)
Department of Engineering Science, University of Auckland, Private Bag 92019, Auckland, New Zealand
e-mail: d.nield@auckland.ac.nz

C. T. Simmons
School of Chemistry, Physics and Earth Sciences, Flinders University, GPO Box 2100, Adelaide, SA 5001, Australia

in many hydrogeologic systems. Such heterogeneity occurs over many spatial scales and variable density flow phenomena may be triggered, grow, and decay over a very large mix of different spatial and temporal scales. Dense plume problems in these geologic environments are, in general, expected to be inherently transient in nature and often involve sharp plume interfaces whose spatiotemporal development is very sensitive to initial conditions. They are also expected to be complicated. Importantly, Simmons et al. (2001) and Prasad and Simmons (2003) made the critical observation that the onset of instability in transient, sharp interface problems is controlled by very local conditions in the vicinity of the evolving boundary layer and not by the global layer properties or indeed some average property of that macroscopic layer. They also pointed out that any averaging process would remove the very structural controls and physics that are expected to be important in controlling the onset, growth, and/or decay of instability in a highly heterogeneous system. These authors reported that in the case of dense plume migration in highly heterogeneous environments the application of an average global Rayleigh number based upon average hydraulic conductivity of the medium appeared problematic. In these cases, an average Rayleigh number is unable to predict the onset of instability accurately because the system is characterized by unsteady flows and large amplitude perturbations. For statistically equivalent geologic systems, and hence identical average global *Ra*, dense plume behavior was observed by Simmons et al. (2001) and Prasad and Simmons (2003) to vary between highly unstable and highly stable, suggesting that the global Rayleigh number is not a consistent and accurate predictor of plume behavior.

A number of factors clearly limit the application of the Rayleigh number in highly heterogeneous geologic environments and support the above conclusion. These include the inability to accurately quantify both time-dependent non-dimensionalizing length scales and dispersion in plume problems. In real field-scale settings the critical transition regions between stable and unstable flow and transport behavior in groundwater systems are rarely known, and the idealized boundary condition assumptions underlying the classic Horton–Rogers–Lapwood problem are not met. Thus there are inherent difficulties in not only defining an appropriate average Rayleigh number in a heterogeneous system but also in then determining whether or not any *Ra* number, should it be accurately quantified, implies stable or unstable flows are theoretically expected. Also it is likely that in the case of transient development of fingers elements of Rayleigh–Taylor instability are involved and the effect is accentuated when there is heterogeneity. Thus one may expect that the use of a single average Rayleigh number has severe limitations in the highly heterogeneous situations investigated by Simmons and his colleagues.

It is now argued here that as a criterion for the onset of instability (as distinct from its subsequent transient development, a very important distinction) in the case of moderately heterogeneous media the average Rayleigh number may do a better job than Simmons and his colleagues have reported for cases of high-level heterogeneity. Here we intentionally make a clear and very important distinction between moderate and high-level heterogeneity. Furthermore, because they worked in terms of a log permeability field they employed a poor estimate (too low) of the average permeability. That means that they did not give the traditional theory a fair test.

In the remainder of the paper we talk about vertical heterogeneity and horizontal heterogeneity. To reduce possible confusion, we emphasize that by *vertical heterogeneity* we mean *variation in the vertical direction*. An extreme example is when we have a series of horizontal layers in each which the permeability is uniform.

The extent to which an equivalent Rayleigh number based on averaged permeability might work was investigated by Nield (1994), who considered the case of vertical heterogeneity of permeability and other parameters. He showed that provided the variation of each of the various parameters lies within one order of magnitude, a rough and ready estimate of an effective Rayleigh number can be made that is useful as a criterion for Rayleigh–Bénard convection. This effective Rayleigh number is based on the arithmetic mean quantities (such as the permeability) that appear in the numerator, and the harmonic mean of quantities (such as the viscosity) that appear in the denominator of the defining expression. It is clear that the arithmetic mean of the logarithm of the permeability is not equal to the logarithm of the mean of the permeability.

For example, Eq. (6) of Simmons et al. (2001) reads:

$$K_{x,y}(x,y) = K_{\mathrm{AV}} 10^{\psi\left(\sin\frac{2\pi x}{\lambda}\right)}. \tag{1}$$

This refers to the hydraulic conductivity at the point (x, y), and the authors refer to the "average conductivity, K_{AV}". This is misleading. A correct statement is that $\log K_{\mathrm{AV}}$ is the average of the logarithm of the conductivity. The true average (arithmetic mean) of the quantity in Eq. (1) is

$$\overline{K} = \frac{1}{\lambda}\int_0^{\infty} K_{x.y}(x,y)dy = \frac{K_{\mathrm{AV}}}{2\pi}\int_0^{2\pi} 10^{\psi\sin\theta}d\theta. \tag{2}$$

This integral can be evaluated by introducing a complex variable and using the residue theorem. The result is

$$\frac{\overline{K}}{K_{\mathrm{AV}}} = 1 + \left(\frac{\mu^2}{1!}\right)^2 + \left(\frac{\mu^2}{2!}\right)^2 + \left(\frac{\mu^2}{3!}\right)^3 + \cdots \tag{3}$$

where $\mu = (1/2)\psi \ln 10$.

It is clear that $\overline{K}$ is greater than K_{AV}, and it increases as the amplitude ψ increases. Since the Rayleigh number involves the first power of the permeability, it is clear that using K_{AV} instead of the more appropriate value $\overline{K}$ leads to an underestimate of a predicted critical Rayleigh number. The effect is accentuated in the situation considered by Prasad and Simmons (2003). They considered a normal distribution of the log permeability field, an approach that is consistent with standard stochastic practice in generating correlated random spatial functions for heterogeneity. However, in this case the arithmetic mean of the permeability distribution is theoretically infinite and this is problematic in the calculation of an effective Rayleigh number. However, it is possible to calculate an arithmetic mean of a finite heterogeneous domain once it has been sampled from a log–normal distribution, but this was not the method employed by Prasad and Simmons (2003). Clearly, care must be taken in defining appropriate *Ra* numbers in heterogeneous distributions that are generated using traditional log–normal permeability distributions to ensure that the true arithmetic mean of the permeability distribution is employed. What is most important here, however, is that the results of Simmons and his colleagues have not ruled out the possibility that an equivalent Rayleigh number, based on the arithmetic mean of the permeability, may be useful as a criterion for the onset of convection in moderately heterogeneous systems. Clearly, in the limiting case that heterogeneity reduces to ultimately reach homogeneous conditions, one would expect progressively better predictive behavior.

Indeed, one of the important findings of Prasad and Simmons (2003, p. 17) supports the present conclusion. They note that as the standard deviation of the permeability field increases, the standard deviation of output variables increases. The logical consequence of this observation is therefore that as the standard deviation of the permeability field increases, the probability of a successful prediction decreases. They also note that there is a critical standard deviation for any given mean, below which the mean output (instability indicator) variables are essentially unaltered from the homogeneous values and where probabilities of successful predictions would therefore improve greatly. A probability of exceedence analysis also demonstrated that analyses based upon homogeneous assumptions typically underestimated, often quite significantly, the value of key measurable output characteristics.

There is an urgent need to distinguish explicitly and quantitatively between high and moderate heterogeneity, and to investigate precisely where the transition takes place and what factors control the nature of that transition. A step in this direction has already been taken by Nield (1994). His results indicate that the equivalent Rayleigh number leads to a good quantitative estimate provided that the permeability variation is not more than threefold. Importantly, the sensitivity of onset conditions to the geometrical structure of heterogeneity warrants exploration at moderate heterogeneity. Whether a moderate or highly heterogeneous field is considered, structural details of the heterogeneous field are not captured by an average permeability and hence effective Rayleigh number. Whilst clearly important in highly heterogeneous cases, the extent to which the explicit structure of the heterogeneity controls the onset of instability conditions in moderately heterogeneous cases is not yet clear and warrants investigation. Clearly, two different moderately heterogeneous structures with identical arithmetic means, and hence effective *Ra*, could lead to entirely different onset behavior locally at the boundary layer depending upon the precise distribution of the permeability distribution very locally in the vicinity of the dense plume source. Such an effect could not possibly be captured by the average *Ra*. We speculate that then one might need to distinguish between structures by the introduction of at least two Rayleigh numbers, one based on the mean of the permeability and the other based on the standard deviation of the permeability.

In further discussion of the heterogeneity problem, one should be aware of the results in the various papers surveyed in Section 6.13 of any edition of Nield and Bejan (1992, 1999, 2006). These papers deal with some limiting cases, and the results can serve as guides for further work. Particularly pertinent is the work of McKibbin and O'Sullivan (1980, 1981) on layered media for the case of vertical heterogeneity (e.g., horizontal layering) and McKibbin (1986) and Nield (1987) for the case of horizontal heterogeneity.

In the case of strictly vertical heterogeneity an obvious way of dealing with the general permeability distribution is to use a piecewise-constant approximation. That is, one can divide the field into horizontal layers, within each of which one can employ an arithmetic mean permeability with the mean taken over that particular layer. (In this way a continuously varying permeability is discretized in terms of the vertical position coordinate.) One would expect that this procedure would give an acceptable prediction if the heterogeneity was moderate or small. It is true that for high heterogeneity this procedure has some obvious limitations. Consider, for example, the case in which the permeability is low except in a single thin layer. Then one might have instability (convection) within that layer whereas the overall equivalent Rayleigh number might indicate stability (conduction of heat, diffusion of solute). If one had in mind an industrial process in which any instability was deleterious then clearly the faulty prediction would be a serious matter. On the other hand, if one just wanted to estimate whether the global heat transfer would be significantly above the conduction level, then the faulty prediction might be acceptable. If the permeability varies in the horizontal direction as well as in the vertical direction, one could consider dividing each sublayer into a number of vertical columns. Progressively one would arrive at an increasingly localized and spatially distributed *Ra* number assessment throughout the heterogeneous domain. Furthermore, this analysis could be performed at any given time leading to the idea of a spatiotemporally distributed Rayleigh space that could be used in assessing stability problems.

Some general remarks can be made about the horizontally layered situation.

(i) Upper and lower bounds for the critical value of the vertical density gradient can be obtained using the minimum and maximum values of the permeability, respectively.
(ii) Assessments of stability of time-dependent situations can be made by freezing the time and examining the instantaneous situation. If the criterion (whatever one chooses to use) predicts stability for all frozen times then one has stability overall. (If instability is predicted at particular time further investigation is needed in order to determine whether the disturbance continues to grow or later decays.)
(iii) The effect of heterogeneity depends on how close to instability one is. If one is well away from instability then one can tolerate relatively high amounts of heterogeneity when employing the equivalent Rayleigh number as a criterion for stability. Of course, extremely high heterogeneity, and especially high-localized heterogeneity, will always cause problems for any general criterion of stability.

The case of horizontal heterogeneity is more difficult to deal with. Indeed, except for the studies of McKibbin (1986) and Nield (1987), it appears that little theoretical work on this aspect has been published. The reciprocal of the permeability is a measure to the resistance to motion, and we can invoke an analogy here with electrical resistance. In the case of convection induced by buoyancy the primary flow is vertical and the net heat flow is vertical, and for horizontal layering the contributions of the individual layers are in series for the flow of mass and heat. In particular, the resultant of two resistances of magnitudes 1 and ε is $1 + \varepsilon$ and if ε is small then the resultant differs little from unity. In contrast, for the case of vertical layering the resistances are in parallel. Then the resultant of two resistances of magnitudes 1 and ε is $\varepsilon/(1 + \varepsilon)$, a quantity that differs greatly from unity when ε is small.

We are not sure how far the resistor analogy should be pushed, but we thought it worthwhile to do one exploratory exercise based on it. Consider a square box with vertical and horizontal sides and quartered by vertical and horizontal midlines. Suppose that the resistances of the individual boxes have the following values:

$$\text{top left:}1 - \varepsilon - \eta, \quad \text{top right:}1 + \varepsilon - \eta,$$
$$\text{bottom left:}1 - \varepsilon + \eta, \quad \text{bottom right:}1 + \varepsilon + \eta,$$

where ε (representing horizontal heterogeneity) and η (representing vertical heterogeneity) are small compared with unity (the homogeneous case). Suppose that items in the same column (the vertical direction) are treated in series (for the net heat flow) and those in the same row (the horizontal direction) are treated in parallel. Then one finds that, to second order in the small quantities, the calculated overall resistance does not depend on whether rows are treated before columns or vice versa, the resulting resistance being $1 - \varepsilon^2$ in each case. Thus to this order of approximation, the vertical heterogeneity has no effect while the effect of horizontal heterogeneity is to reduce the resistance. The conclusion for our convection problem is that weak vertical heterogeneity is expected to have little effect while weak horizontal heterogeneity is expected to increase the likelihood of instability.

This resistance analogy provides simple but convincing evidence in support of a claim that the structural orientation of the heterogeneous field relative to the primary vertical buoyancy flow direction is likely to be very important in controlling both instability onset conditions, as well as subsequent transient behavior. The extent to which structure can therefore be removed through any averaging process involved in *Ra* computations requires further elucidation, even at moderate levels of heterogeneity. It is likely that the nature of the transition between moderate and highly heterogeneous fields will not only be controlled by an arithmetic mean and how close that mean is to the point of instability but will also be controlled by the geometrical details of the structure itself. Understanding what level of structural simplification is possible and under what conditions is critical in developing a more quantitative appreciation of the conditions under with the Rayleigh number is both expected to work and not work, and importantly what spatial scale it must be applied at. We contend that these matters are not simple and do warrant further investigation.

This discussion has highlighted a number of approaches, resolutions and challenges that exist in understanding the complex area of convective flow in heterogeneous porous media. Indeed, we believe that the discussion has provided compelling evidence in support of our opening remarks about the state of this obviously complex field, namely that many important issues remain largely unexplored and poorly understood and should form the basis for further research investigation.

Postscripts

Weak heterogeneity

Since we submitted the first version of this paper we have been engaged on detailed analyses of weak heterogeneity, defined as the case where the variation in a property is a small fraction of its mean value. Our approach is essentially a perturbation one, starting with the solutions for the temperature and velocity fields for the homogenous case, and using those as trial functions with a Galerkin method involving functions of two variables (the horizontal and vertical coordinates). The case of a square enclosure or a tall rectangular enclosure with "conducting" (constant temperature) top and bottom was treated by Nield and Kuznetsov (2006a). The case of a shallow rectangular enclosure with "insulating" (constant flux) top and bottom was studied by Nield and Kuznetsov (2006b). We are currently investigating further extensions, namely to the cases of double diffusion, local thermal non-equilibrium, a bidisperse porous medium and temperature-gradient heterogeneity (nonuniformity of the basic thermal temperature gradient, resulting from the freezing of a weakly transient temperature distribution or otherwise). We have found that weak heterogeneity does not affect the critical value of the Rayleigh number (defined in terms of mean properties) at first order in the small parameters expressing the heterogeneity. The second order corrections have been obtained. The results show no dramatic difference between vertical heterogeneity (e.g., horizontal layering) effects and horizontal heterogeneity effects at this order of approximation, once one takes into account the geometrical effect of the aspect ratio of the enclosure.

Strong heterogeneity

We now propose a rough-and-ready criterion for the onset of convection in the situation where there is strong heterogeneity. This criterion is not restricted to the two-dimensional situation. It is based on the famous Figs. 2 and 3 of Beck (1972) that have been reproduced many times, for example as Figs. 6.22 and 6.23 in each edition of Nield and Bejan (1992, 1999, 2006). These show the variation of the critical Rayleigh number, and the preferred cellular mode, as functions of the aspect ratios $A_x = H/L_x$ and $A_y = H/L_y$ for a three-dimensional box with height H and horizontal dimensions L_x and L_y. (Actually Beck worked with the reciprocals of those aspect ratios.) The figures apply to a box with impermeable conducting top and bottom and impermeable insulating sidewalls, and occupied by a porous medium for which the Darcy model is applicable. Beck showed that the critical Rayleigh number for a homogeneous medium is given by

$$Ra = \pi^2 \min \left(b + \frac{1}{b} \right)^2 \tag{4}$$

where

$$b = \left[(pA_x)^2 + (qA_y)^2 \right]^{1/2} \tag{5}$$

and the minimum is taken over the set of nonnegative integers p and q. Beck's figures show that in the region $A_x < 1, A_y < 1$, the value of Ra does not exceed 40.7. Also, in the region $A_x > A_y > 1$, the critical mode is $p = 1, q = 0$, so that

$$Ra = \pi^2 \left(A_x + A_x^{-1} \right)^2. \tag{6}$$

Furthermore, when $A_x > 1$ and $A_y < 1$ the value of Ra does not exceed the value given by the expression in Eq. (6).

Similarly, in the region $A_y > A_x > 1$, the critical mode is $p = 0, q = 1$, so that

$$Ra = \pi^2 \left(A_y + A_y^{-1} \right)^2, \tag{7}$$

and when $A_y > 1$ and $A_x < 1$ the value of Ra does not exceed the value given by the expression in Eq. (7).

We now construct a generalized Rayleigh number for a heterogeneous box in the following way. Our basic idea is that if at any stage we find instability in any part of the enclosure at any time, then the whole system can be considered to be unstable. We start with a domain consisting of the box and consider subdomains. Each subdomain is taken to be a rectangular box, of arbitrary size and with arbitrary aspect ratios, bounded by planes $x = x_1, x = x_2$, $y = y_1, y = y_2$, $z = z_1, z = z_2$. We call these things sub-boxes. Then for each sub-box we calculate the aspect ratios and a local Rayleigh number Ra_l based on the height of the sub-box and with other properties given the mean value over the sub-box. In particular, the sub-box mean of the basic temperature gradient at a particular time is employed here. We then define a geometrically adjusted Rayleigh number Ra_{g} defined by

$$Ra_{\mathrm{g}} = \begin{cases} Ra_{\mathrm{l}} & \text{if } A_x < 1 \text{ and } A_y < 1, \\ \frac{4Ra_{\mathrm{l}}}{\left(A_x + A_x^{-1}\right)^2} & \text{if } A_x > 1 \text{ and } A_x > A_y, \\ \frac{4Ra_{\mathrm{l}}}{\left(A_y + A_y^{-1}\right)^2} & \text{if } A_y > 1 \text{ and } A_y > A_x. \end{cases} \tag{8}$$

We then define an overall Rayleigh number Ra_{o} defined by

$$Ra_{\mathrm{o}} = \max Ra_{\mathrm{g}} \tag{9}$$

where the maximum is taken over all the subdomains and all times.

We then expect that if $Ra_{\mathrm{o}} > 41$ then instability will occur. The criterion for instability will be met in at least one sub-box, and hence the whole system will be unstable. If $Ra_{\mathrm{o}} \ll 41$ then it is unlikely that instability will occur. If Ra_{o} is only slightly less than 41 then a closer examination of the particular situation is needed to determine whether or not instability will occur.

Our expectation that $Ra_{\mathrm{O}} > 41$ is a sufficient condition for instability is based on the fact that the impermeable conducting boundaries are the most restrictive boundaries pertaining to the Darcy model. The boundary conditions on the top and bottom of the sub-boxes are undetermined, but we can be sure our criterion is conservative.

(Specifying the value of a variable on the boundary is more restrictive than, for example, specifying the value of the derivative of that variable. In an eigenvalue problem with a given differential equation, the more restrictive the boundary conditions then the greater the eigenvalue. In Table 6.1 of Nield and Bejan (1992, 1999, 2006), which gives the values of the critical Rayleigh number for various boundary conditions, the largest entry for Ra_{c} corresponds to the impermeable conducting boundaries.)

In practice the process of maximization over the sub-boxes can be approximated by restricting the faces of the sub-boxes to a numerical grid, so the maximum will then be found subject to such constraints. The process of finding the maximum can be prematurely terminated if at any stage it is found that the value of Ra_{g} for a sub-box has a value greater than 41. We recognize that a challenge when applying our criterion to a practical hydrogeological system is that the permeability distribution will need to be approximated before the stability analysis can be conducted.

Acknowledgement The authors have benefited from discussions with Dr Robin Wooding.

References

Beck, J.L.: Convection in a box of porous material saturated with fluid. Phys. Fluids **15**, 1377–1383 (1977)

McKibbin, R., O'Sullivan, M.J.: Onset of convection in a layered porous medium heated from below. J. Fluid Mech. **96**, 375–393 (1980)

McKibbin, R., O'Sullivan, M.J.: Heat transfer in a layered porous medium heated from below. J. Fluid Mech. **111**, 141–173 (1981)

McKibbin, R.: Heat transfer in a vertically layered porous medium heated from below. Transport in Porous Media **1**, 361–370 (1986)

Nield, D.A.: Convective heat transfer in porous media with columnar structure. Transport in Porous Media **2**, 177–185 (1987)

Nield, D.A.: Estimation of an effective Rayleigh number for convection in a vertically inhomogeneous porous medium or clear fluid. Int. J. Heat Fluid Flow **15**, 337–340 (1994)
Nield, D.A., Bejan, A.: Convection in Porous Media. Springer, New York (1992, 1999, 2006)
Nield, D.A., Kuznetsov, A.V.: The effects of combined horizontal and vertical heterogeneity on the onset of convection in a porous medium. Int. J. Heat Mass Transfer, submitted (2006a)
Nield, D.A., Kuznetsov, A.V.: The onset of convection in a shallow box occupied by a heterogeneous porous medium with constant flux boundaries. Transport in Porous Media, to appear (2006b)
Prasad, A., Simmons, C.T.: Unstable density-driven flow in heterogeneous porous media: A stochastic study of the Elder [1967b] "short heater" problem. Water Resour. Res. **39** (1), 1007 (2003)
Simmons, C.T., Fenstemaker, T.R., Sharp, J.M.: Variable-density flow and solute transport in heterogeneous porous media: Approaches, resolutions and future challenges. J. Contam. Hydrol. **52**, 245–275 (2001)