Comment on "*The role of scaling laws in upscaling*" *by B.D. Wood*

Philippe Baveye

SIMBIOS Centre, Abertay University, Kydd Building, 40 Bell Street, Dundee DD1 1HG, UK.

In his recent article, Wood (2009) proposes a thorough and very interesting analysis of the process of coarse-graining in complex subsurface hydrologic systems. No doubt this article will eventually help elucidate a number of complicated aspects of upscaling. Nevertheless, one of Wood's (2009) key results is somewhat surprising. It implies that "the ability to develop a coarse-grained equation for describing transport in heterogeneous porous media has much less to do with the particulars of the mathematical methods used for averaging than it does the assumptions (scaling laws) adopted when developing the upscaled model" (p. 735). Specifically, Wood (2009) asserts that "averaging *per se* does not lead to a reduction in the information content of any particular problem". I would advocate that exactly the opposite is true, namely that the averaging process eliminates way too much information, and that a significant challenge in years to come will be to find ways to volume average while preserving some essential information about microscale spatial complexity.

One of the key features of the volume averaging formula used by Wood (2009, eq. (8)), which for the microscale solute concentration c(***y***,t) is

$$\langle c \rangle\big|_{x,t} = \frac{1}{\mathrm{V}} \int_{y \in V_x} c(y,t)\, dy \qquad (1)$$

where $\langle c \rangle|_{x,t}$ is the average (macroscopic) value of solute concentration at ($\boldsymbol{x}$,t), and V is the volume of the domain V_x over which the averaging is performed, is that it is designed to cause considerable information loss. Indeed, once the integral of eq. (1) is carried out, all the detailed information associated with the microscale concentration $c(\boldsymbol{y},t)$ effectively disappears. Only if $\langle c \rangle|_{x,t}$ is known rigorously at every single point $\boldsymbol{x}$ of the medium, and if V_x is strictly invariant in space and time, can eq. (1) be deconvoluted to reclaim the original information. However, this is never possible in practice.

Under these conditions, since volume averaging in and of itself clearly leads to massive information loss, where did Wood's conclusion come from? It seems to stem from several assumptions made in the analysis. The first is the introduction of a particular form of "deviation concentration", defined as the difference between the microscale and average concentrations at the point where the macroscale concentration is evaluated (Wood's eq. (11)):

$$c(x,t) = \langle c \rangle|_{x,t} + \delta c(x,t) \qquad (2)$$

Other types of deviation concentrations or local fluctuations could have been used instead, such as the "fluctuating components" of Efendiev et al. (2000, eq. (4a)), leading to different, coarse-grained equations that are not strictly continuous, but rather discretized.

The second assumption introduced by Wood (2009) is to consider that the microscale convection-dispersion equation

$$\frac{\partial c}{\partial t} = \nabla \cdot (\underline{\underline{\boldsymbol{D}}} \cdot \nabla c) - \nabla \cdot (\boldsymbol{v} c) \qquad (3)$$

where $\underline{\underline{\boldsymbol{D}}}$ is the microscale dispersion tensor and $\boldsymbol{v}$ is the microscale pore water velocity, is an acceptable starting point for the derivation of macroscale transport equations, via upscaling. Several other perspectives could have been adopted, such as starting from traditional conservation equations (of mass, momentum, or energy) at the microscale, volume averaging them, using averaging theorems (e.g., Baveye and Sposito, 1984, 1985) to simplify the volume-averaged equations, and introducing constitutive relations at the macroscopic scale.

A third assumption in Wood's analysis is that constitutive relations at the macroscale are considered in connection not with a volume-averaged form of eq. (3), but with the so-called "deviation equation"

$$\frac{\partial \tilde{c}}{\partial t} - \nabla \cdot (\underline{\underline{\boldsymbol{D}}} \cdot \nabla \tilde{c}) + \nabla \cdot (\boldsymbol{v}\tilde{c}) = \nabla \cdot (\tilde{\boldsymbol{f}}_1 + \tilde{\boldsymbol{f}}_2 + \boldsymbol{I}) \tag{4}$$

where

$$\tilde{\boldsymbol{f}}_1 = [\underline{\underline{\boldsymbol{D}}} \cdot \nabla \langle c \rangle - \langle \underline{\underline{\boldsymbol{D}}} \cdot \nabla \langle c \rangle \rangle] \tag{5}$$

$$\tilde{\boldsymbol{f}}_2 = -[\boldsymbol{v} \langle c \rangle - \langle \boldsymbol{v} \langle c \rangle \rangle] \tag{6}$$

$$\boldsymbol{I} = -[\langle \underline{\underline{\boldsymbol{D}}} \cdot \nabla \tilde{c} \rangle - \langle \boldsymbol{v} \tilde{c} \rangle] \tag{7}$$

The deviation expression in eq. (4) is obtained by replacing c by eq. (2) in eq. (3), volume averaging the resulting equation and subtracting eq. (3). As ingenious as this manipulation is, again alternative approaches could have been considered, leading to different intermediate transport equations, not necessarily involving variables that cannot be directly measured, as eq. (4) does. These different equations would require different heuristic constitutive relationships to end up with the same macroscale equation, eventually. Nevertheless, like Wood's (2009) perspective, the various possible approaches along those lines would all be equally defensible. As always in continuum mechanics, and particularly in the continuum mechanics of transport processes in polyphasic systems, the ultimate test of a particular heuristic construct is whether it is able to derive empirically-obtained transport expressions.

To the extent that it involves the local deviation $\tilde{c}$ at every point in the medium, eq. (4) formally encompasses the same information as eq. (3). However, clearly, this is in spite of the massive information loss caused by volume averaging, and results directly from the additional assumptions made by Wood (2009), which in effect re-introduce local microscale information in the picture. So, the statement that "averaging *per se* does not lead to a reduction in the information content of any particular problem" does not represent what really happens.

This discussion about the loss of information associated with the integration in eq. (1) has more than theoretical interest. In a number of different contexts, including the ecology of soil microorganisms (Balser et al., 2006) or the metagenomic analysis of soils (Baveye, 2009), it has significant practical implications. Fundamentally, the link with practice is due to the fact that eq. (1) mimics the measurement process in soils (Baveye and Sposito, 1984, 1985). A neutron moisture meter, a gamma-ray densitometer, or a time-domain reflectometer probe, located at a specific point $\boldsymbol{x}$ in a soil at a time t, are all affected by the position of myriads of atoms present around $\boldsymbol{x}$, but they manage to condense the information associated with the spatial coordinates of these atoms into a single number, either the volumetric water content or the bulk density of the soil at $(\boldsymbol{x},t)$. The phenomenal loss of information that results characterizes every measurement made in soils at this stage, even in the case of the minute (micron-sized) voxels delineated by computed tomography (e.g., Baveye et al., 2002).

In recent years, it has become clear that the volume averaging (or convolution product with instrumental response functions) carried out by measuring instruments, loses information that, in some situations, is crucially needed to describe soil processes, particularly (but not exclusively) those involving microorganisms. A vivid example of such a situation is the microscopic spatial distribution of Cu in vineyard soils in Burgundy (Jacobson et al., 2007). In these soils, bulk Cu concentrations reach levels of hundreds of parts per million that should strongly inhibit microbial growth and metabolism. Yet, microorganisms thrive. The key to this apparently contradictory observation, revealed by synchrotron-based X-ray fluorescence spectroscopy, is that Cu distribution is extremely heterogeneous at the micron scale, with "hotspots" in the vicinity of which microorganisms are likely to have a hard time colonizing the soil, and with relatively large portions of the soil that are devoid of Cu. This heterogeneity cannot be captured by classical bulk (macroscopic) Cu concentrations. At a minimum, beyond a simple volume average, some measure of the spatial dispersion of Cu and of microorganisms is needed, as well as a parameter quantifying the spatial disconnect between them.

Garnier et al. (2009) introduced such a parameter in their model of straw biodegradation in soils, and showed that with this parameter, which they termed a "contact factor", a satisfactory fit of their model to experimental data could be obtained. At this stage, the model of Garnier et al. (2009) is empirical and still has to be established on a firm theoretical basis. It might be possible to tweak Wood's (2009) upscaling scheme slightly, i.e., introduce different constitutive relations, to produce a macroscale convection-dispersion equation containing a disconnect parameter *à la* Garnier et al. (2009). Such an approach, despite its heuristic character, would be very enlightening, in that it would allow us to determine what information, lost during the classical volume averaging process, is required to describe things correctly. Reintroducing all the microscale information, in this context, would definitely be an overkill, but it is not clear what minimum set of data is absolutely needed and in what form.

REFERENCES

Balser TC, McMahon KD, Bart D, Bronson D, Coyle DR, Craig N, Flores-Mangual ML, Forshay K, Jones SE, Kent AE, Shade AL. Bridging the gap between micro - and macro-scale perspectives on the role of microbial communities in global change ecology. Plant and Soil 2006; 289: 59-70.

Baveye P, Sposito G. The operational significance of the continuum hypothesis in the theory of water movement through soils and aquifers. Water Resources Research 1984; 20: 521-530.

Baveye P, Sposito G. Macroscopic balance equations in soils and aquifers: the case of space- and time-dependent instrumental response. Water Resources Research 1985; 21: 1116-1120.

Baveye P, Rogasik H, Wendroth O, Onasch I, Crawford JW. Effect of sampling volume on the measurement of soil physical properties: Simulation with X-ray tomography data. Measurement Science and Technology 2002; 13: 775-784.

Baveye, P.C. To sequence or not to sequence the whole-soil metagenome? Nature Reviews Microbiology 2009; 7(10): 756.

Efendiev Y, Durlofsky LJ, Lee SH. Modeling of subgrid effects in coarse-scale simulations. Water Resources Research 2000; 36(8): 2031-2041.

Garnier P, Cambier C, Bousso M, Masse D, Chenu C, Recous S. Modeling the influence of soil-plant residue contact on carbon mineralization: Comparison of a compartmental approach and a 3D spatial approach. Soil Biology and Biochemistry 2009; 40: 2754-2761.

Jacobson AR, Dousset S, Andreux F, Baveye PC. Electron microprobe and synchrotron X-ray fluorescence mapping of the heterogeneous distribution of copper in high-copper vineyard soils. Environmental Science and Technology 2007; 41(18): 6350-6356.

Wood, E. The role of scaling laws in upscaling. Advances in Water Resources 2009; 32:723-736.