Predictions in Ungauged Basins: PUB Kick-off (Proceedings of the PUB Kick-off meeting held in Brasilia, 20–22 November 2002). IAHS Publ. 309, 2007.

Theoretical hydrology with respect to scaling and heterogeneity over sparsely-gauged or ungauged basins

M. L. KAVVAS[1], Z. Q. CHEN[1], J. Y. YOON[2] & J. YOSHITANI[3]

1 *Hydrological Research Laboratory, Civil & Envr. Engineering, UC Davis, California 95616, USA*

2 *Dept. of Environmental Engineering, Korea University, Jochiwon, Chungnam, Korea*

3 *Indep. Adm. Inst. Public Works Research Institute, Ministry of Land, Infrastructure and Transport, Japan*

Abstract Due to heterogeneity of lands over ungauged/sparsely gauged watersheds at which the hydrological flows occur, there is a need to upscale the hydrological conservation equations to the scale of the grid areas of the numerical mesh laid out over the watershed which is being modelled. A literature review of the various methods utilized in hydrology for the upscaling of hydrological conservation equations is provided. Then some recent formulae for the upscaling of hydrological processes are discussed with an application.

Key words conservation equation; cumulant expansion; decomposition; ensemble averaging; heterogeneity; hydrology; Lie operator; projector-operator; regular perturbation; scale; volume averaging

INTRODUCTION

For the quantification of hydrological water balances over sparsely gauged or ungauged regions/watersheds a central issue is how to model the hydrological processes at the scale of the grid areas of ungauged small-mesoscale watersheds (grid size ~1 km), and of sparsely gauged/ungauged regions (grid size ~10 km). Since there may be very sparse or no precipitation/runoff data over an ungauged/sparsely gauged watershed, it may be necessary to take a computational network with sufficiently large grid areas over such a watershed in order to be able to utilize the sparse data (if there is any) or to utilize remotely sensed observations as areally-averaged quantities over such grid areas. Then in order to have a scale-consistent description of the hydrological processes with respect to both numerical modelling and remotely sensed observations, it becomes necessary to develop upscaled hydrological conservation equations for the hydrological processes of interest over such grid areas.

Here, spatial scale is defined as the observational/computational grid size, while the time scale is defined as the observational/computational time interval. In the current state of hydrological science the hydrological conservation equations are generally known at "point-scale". The point-scale may be defined as the scale of a differential control volume. Currently, the conservation equations for mass, momentum and/or energy at a computational node are obtained at the scale of a differential control volume which surrounds that node. Each nodal point of a computational grid network represents a surrounding grid area which may range from ~10 m to ~100 km,

Copyright © 2007 IAHS Press

depending upon the domain being modelled. In order to utilize these hydrological conservation equations for modelling hydrological processes at the particular scale of a grid area, one makes the assumption that the conservation equation (usually a partial differential equation (PDE)) at the node represents the whole hydrological process evolving over the area that surrounds that node. This amounts to assuming homogeneity of soils, vegetation, geology, topography, and atmospheric inputs over the area (volume) that surrounds any nodal point of the computational grid network. However, soils, vegetation, geology, topography, and atmospheric inputs over an area (volume) that surrounds any nodal point of the computational grid network, are heterogeneous. Therefore, a hydrological conservation equation that is derived at the point-scale of a node, becomes uncertain (a stochastic PDE) over the grid area which it purports to represent, due to uncertainty of its parameters and boundary conditions over this area. As such, a point-scale conservation equation can not represent the general behaviour of the hydrological process which is taking place over the grid-scale area that surrounds that node. Consequently, a fundamental problem is how to upscale the existing point-scale hydrological conservation equations for mass, momentum (and/or energy) to the increasingly larger spatial scales, in order to obtain the conservation equations consistent with the scale of the grid areas over which they will describe the hydrological processes at ungauged/sparsely gauged watersheds.

There are various approaches to upscaling of hydrological conservation equations. In the hydrological literature the most commonly employed approach has been the averaging of hydrological conservation equations. The averaging approaches may further be classified as: (a) volume/areal averaging, and (b) ensemble averaging.

In the volume/areal averaging approach the point-scale hydrological conservation equation is integrated over a volumetric or areal domain, and then the resulting integrals are divided by the size of the domain. This approach was used in hydrology in order to reduce the hydrological conservation equations from their original PDE forms at point scale to ODE forms at larger spatial scales. Duffy (1996) reduced the unsaturated–saturated subsurface flow conservation equation from its original PDE form to a set of ordinary differential equations (ODEs) by means of volume averaging. Tayfur & Kavvas (1994, 1998) reduced rill and inter-rill overland flow equations from a 2-D PDE at point scale to an ODE at hillslope scale by volume averaging. However, this approach leads to closure problems where the hydrological fluxes at the boundaries of the flow domain require information on the state of flow within the flow domain (Tayfur & Kavvas, 1998).

In the ensemble averaging approach one recognizes that the point-scale hydrological conservation equations become uncertain (stochastic PDEs) due to the uncertain values of their point-scale parameters and boundary conditions at the grid-area scale. Accordingly, the aim is to obtain an ensemble average form of the original point-scale conservation equation (which is a stochastic PDE at grid area scale) that will represent its upscaled form at the scale of the modelling grid area.

One approach to ensemble averaging is the numerical probabilistic averaging of the conservation equations (Avissar & Pielke, 1989; Entekhabi & Eagleson,1989). In this approach one assigns probability distributions for the parameters of the point-scale conservation equation in order to describe the parameters' statistical variability within a grid area (subgrid variability). Then using these probability distributions, the point-scale conservation equations over the grid area are numerically averaged in order to obtain the grid area-scale behaviour of the corresponding hydrological process.

Meanwhile, there are several analytical approaches to the ensemble averaging of the hydrological conservation equations. One approach is the averaging based on analytical solutions to realizations (Serrano, 1992; Chen *et al*., 1994a,b; etc.). In this approach one obtains a pathwise analytical solution to the conservation equation, and then takes its ensemble average. It is possible to obtain exact analytical closures even in nonlinear problems. As such, this approach was successfully applied to the ensemble averaging of nonlinear Boussinesq equations (Serrano, 1992) and of nonlinear unsaturated soil water flow (Chen *et al*., 1994a,b). The main drawback of this approach is that the analytical solutions are cumbersome and difficult to understand/use by third parties.

Averaging based on regular perturbations is the most often used analytical ensemble averaging approach in hydrology (Gelhar & Axness, 1983; Dagan, 1984; Mantoglou & Gelhar, 1987; Graham & McLaughlin, 1989; Tayfur & Kavvas, 1994; Horne & Kavvas, 1997, etc.). In this approach one expresses each stochastic parameter and each state variable in the conservation equations by a sum of their corresponding mean and a small perturbation term. Then this perturbation expression is substituted in place of the original parameter/state variable within the conservation equation. The expectation of the resulting conservation equation is then taken in order to obtain an ensemble average equation for the considered hydrological process. The advantage of this approach is that it is straightforward to apply, even in nonlinear cases. However, it results in a closure problem where the equation for the mean requires information about the behaviour of higher moments. When one attempts to write an equation for the required higher moment, then that equation for the specific higher moment requires information about the behaviour of even higher moments. Hence, one can close the system of equations only by means of some ad hoc assumption. A small perturbation assumption is often invalid in highly heterogeneous domains where the hydraulic parameters may take large variance values.

Another analytical ensemble averaging approach is based upon Adomian's decomposition theory (Adomian, 1986; Serrano, 1992). In this approach the state variable in the original conservation equation is decomposed into a series of component functions. Then, starting with the deterministic analytical solution to the original conservation equation, the other terms in the decomposition are determined recursively, where each successive component in the series decomposition representation is determined in terms of the preceding component. The decomposition method can accommodate any size of fluctuation. It can be applied both to linear and nonlinear problems, and avoids closure problems by adding successively smaller magnitudes to the solution. Its main drawback is that it requires an analytical solution to the conservation equation in order to develop the corresponding ensemble average equation. However, such analytical solutions are unattainable in many nonlinear hydrological processes.

A promising analytical ensemble average upscaling approach is based on projector-operator theory (Zwanzig, 1960; Cushman, 1991). In this approach one considers an operator which projects quantities onto their averages ($Pu = \langle u \rangle$). Then applying this operator together with an operator that represents the difference between the actual variable and its mean ($Du = u - \langle u \rangle$), one derives an exact integro-differential equation for the ensemble average which is non-local. This approach does avoid the closure problem. However, it is applicable only to linear problems. Also, the obtained integro-differential equation is implicit in the state variable. Therefore, it requires further approximations for its explicit solution.

A popular analytical ensemble average upscaling approach is based on cumulant expansion theory (Kubo, 1962; van Kampen, 1976; Kabala & Sposito, 1991; Kavvas & Karakas, 1996; Karakas & Kavvas, 2000; Kavvas, 2001). In this approach one expresses the original conservation equation in terms of an operator equation with an average component and a fluctuating dynamic component. One then solves the resulting initial value problem in order to obtain the ensemble average equation, expressed in terms of a series of cumulants (correlation functions) of increasing order. Truncation at any order cumulant yields an exact closure at that order. However, the resulting ensemble average equation is in terms of operators which need to be expressed explicitly for practical applications. There are two different approaches for obtaining explicit expressions. One approach is the cumulant expansion combined with spectral theory (Kabala & Sposito, 1991). This approach takes the Fourier transform of the cumulant expansion expression in order to develop an equation for the ensemble average in the Fourier space. However, it is still necessary to invert the expression in the Fourier space to the real time-space for practical applications. The second approach is the cumulant expansion combined with Lie group theory (Kavvas & Karakas, 1996; Wood & Kavvas, 1999; Karakas & Kavvas, 2000; Kavvas, 2001, 2002). This approach recognizes that the operators in the cumulant expansion representation of the ensemble average conservation equation are Lie operators. Then it employs the Lie operator properties (Olver, 1993) in order to obtain an explicit expression for ensemble average conservation equation in real time-space.

By utilizing the combined cumulant expansion-Lie operator theory, a general formula for the upscaling of any linear hydrological conservation equation from point-scale to the next larger spatial scale was developed (Kavvas, 2001). Any linear hydrological conservation equation can be written in the operator form:

$$\frac{\partial h(\mathbf{x},t)}{\partial t} = A(\mathbf{x},t)h(\mathbf{x},t) \tag{1}$$

where h is the state variable and A is the operator coefficient function, and $\mathbf{x}$ is any vectorial location. Then using the cumulant expansion-Lie operator theory it may be shown that the upscaled ensemble average form of equation (1) is (Kavvas, 2001):

$$\frac{\partial < h(\mathbf{x}_t,t) >}{\partial t} = < A(\mathbf{x}_t,t) >< h(\mathbf{x}_t,t) > + \int_0^t ds Cov_o[A(\mathbf{x}_t,t); A(\mathbf{x}_{t-s},t-s)] < h(\mathbf{x}_t,t) > \tag{2}$$

to the order of the covariance time of operator A (exact second order). Also, in equation (2):

$$\mathbf{x}_{t-s} = \overleftarrow{\exp}\,[\int_{t-s}^{t} d\tau A_L(\mathbf{x}_\tau,\tau)]\ \mathbf{x}_t \tag{3}$$

where A_L is that portion of <A> which is made up of the linear combination of the first spatial derivatives and $\overleftarrow{\exp}$ is the time-ordered exponential which is a Lie operator. Fundamentally, this operator displaces the location x_t at time t to the location x_{t-s} at time t–s. Using the upscaling formula (2), the upscaled conservation equation for non-reactive transport by unsteady flow within a heterogeneous aquifer at a spatial scale one step larger than the Darcy-scale may be obtained as (Kavvas & Karakas, 1996):

$$\frac{\partial < c(\mathbf{x}_t,t) >}{\partial t} = \{- < v_i(\mathbf{x}_t,t) > + \int_0^t ds Cov_o[v_j(\mathbf{x}_t,t); \frac{\partial v_i(\mathbf{x}_{t\text{-}s},t-s)}{\partial x_j}]\} \frac{\partial < c(\mathbf{x}_t,t) >}{\partial x_i}$$

$$+ \{D_{ji} + \int_0^t ds Cov_o[v_j(\mathbf{x}_t,t); v_i(\mathbf{x}_{t\text{-}s},t-s)]\} \frac{\partial^2 < c(\mathbf{x}_t,t) >}{\partial x_j \partial x_i} \qquad (4)$$

to exact second order. In equation (4):

$$\mathbf{x}_{t\text{-}s} = \overleftarrow{\exp} \, [\int_{t-s}^t d\tau < v_l(\mathbf{x}_\tau,\tau) > \frac{\partial}{\partial x_l}]\mathbf{x}_t \qquad (5)$$

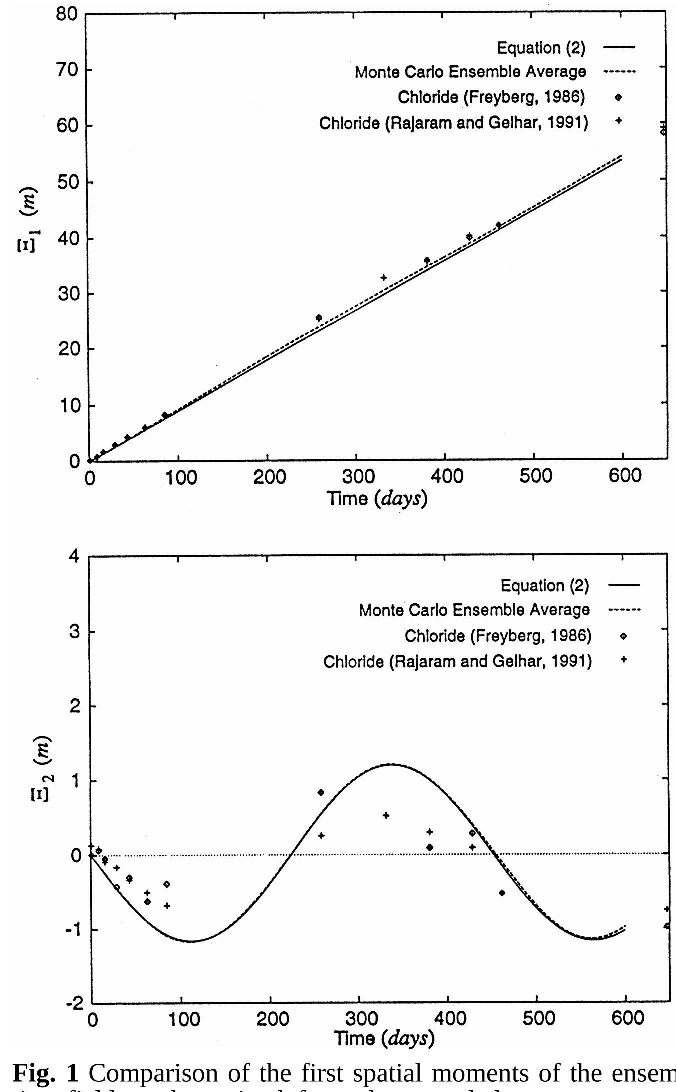

Fig. 1 Comparison of the first spatial moments of the ensemble-averaged concentration field as determined from the upscaled transport equation, from Monte Carlo simulations, and from field data of the Borden Site (from Wood & Kavvas, 1999). Field data are shown by diamond or plus points.

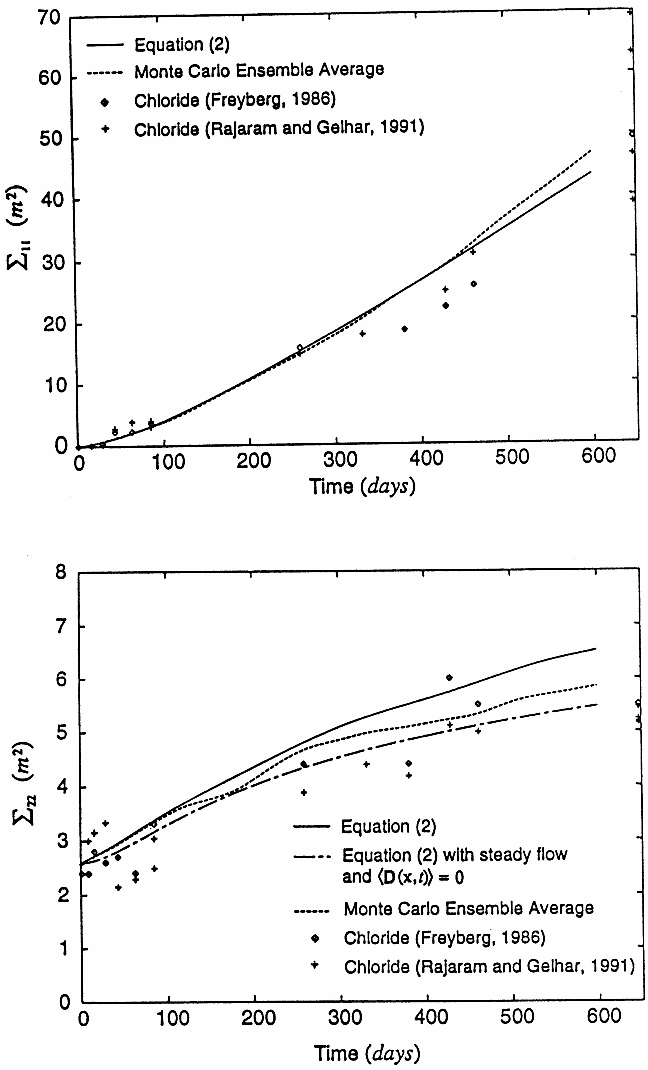

Fig. 2 Second spatial moments of the ensemble-averaged concentration field as determined from the upscaled transport equation and from the Monte Carlo simulation, as compared to those from field data. The second moments calculated directly from the Borden aquifer data appear as circle or plus points (from Wood & Kavvas, 1999).

This theory was applied to Borden Aquifer field experimental data on solute transport within a heterogeneous aquifer with satisfactory results (Wood & Kavvas, 1999). In Figs 1 and 2, some of these application results from Wood & Kavvas (1999) are given. Figure 1 compares the first spatial moments of the solute plume in the longitudinal (Ξ_1) and transverse (Ξ_2) directions from the field data, and those estimated from the above upscaled transport equation (4) as a special case of the upscaling formula equation (2). Similarly, Fig. 2 compares the second spatial moments of the solute plume in the longitudinal (Ξ_{11}) and transverse (Ξ_{22}) directions, as obtained from the field data and from the upscaling theory above.

Again, by utilizing the combined cumulant expansion-Lie operator theory, a general formula for the upscaling of any nonlinear hydrological conservation equation from point-scale to the next larger spatial scale was also developed recently, and is given in Kavvas (2003).

DISCUSSION

In the world of upscaled hydrological conservation equations one may note the following features: (1) While the original point-scale conservation equations are Eulerian, the upscaled conservation equations are mixed Eulerian–Lagrangian. Hence, their solutions will require new computational approaches that can accommodate mixed Lagrangian–Eulerian frameworks. (2) While the parameters of the existing point-scale conservation equations are at point-scale, the parameters of the upscaled conservation equations are at the scale of the grid areas being modelled (e.g. areal median saturated hydraulic conductivity, areal variance of log hydraulic conductivity, areal covariance of flow velocity, etc.). Hence, new parameter estimation methodologies will be required for the estimation of these grid area-scale parameters. (3) The spatial heterogeneities due to topography, soils, vegetation, land use/land cover, geology are incorporated explicitly into the upscaled conservation equations by means of the newly emerging parameters on the areal variance/covariance of the point-scale parameters. Especially, the areal dispersion of the point-scale hydrological dynamics (due to heterogeneity in land conditions and atmospheric boundary conditions) is explicitly modelled in the upscaled equations. (4) The hydrological models which are based upon point-scale conservation equations with effective parameters may yield significantly incorrect predictions over highly heterogeneous ungauged basins due to the effect of hydrological nonlinearity. In such basins it may be necessary to utilize upscaled hydrological conservation equations with their upscaled parameters. (5) It is essential to establish a hydrological model intercomparison project by which existing models (point-scale or upscaled) are tested for their performance when they are provided no atmospheric/hydrological data over a large basin for predicting hydrological processes at various spatial scales within that basin.

REFERENCES

Adomian, G. (1986) *Nonlinear Stochastic Operator Equations.* Academic Press, New York, USA.

Avissar, R. & Pielke, R. A. (1989) A parameterization of heterogeneous land surface for atmospheric numerical models and its impact on regional meteorology. *Mon. Weath. Rev*. **117**, 2113–2136.

Chen, Z. Q., Govindaraju, R. S. & Kavvas, M. L. (1994a) Spatial averaging of unsaturated flow equations under infiltration conditions over areally heterogeneous fields: 1. Development of models. *Water Resour. Res.* **30**, 523–533.

Chen, Z. Q., Govindaraju, R. S. & Kavvas, M. L. (1994b) Spatial averaging of unsaturated flow equations under infiltration conditions over areally heterogeneous fields: 2. Numerical simulations. *Water Resour. Res*. **30**, 535–548.

Cushman, J. H. (1991) On diffusion in fractal porous media. *Water Resour. Res*. **27**, 643–644.

Dagan, G. (1984) Solute transport in heterogeneous porous formations. *J. Fluid Mech., London* **145**, 151–177.

Duffy, C. J. (1996) A two-state integral-balance model for soil moisture and groundwater dynamics in complex terrain. *Water Resour. Res*. **32**, 2421–2434.

Entekhabi, D. & Eagleson, P. (1989) Land surface hydrology parameterization for atmospheric general circulation models including subgrid-scale spatial variability. *J. Climate* **2**(8), 816–831.

Gelhar, L. W. & Axness, C. L. (1983) Three-dimensional stochastic analysis of macrodispersion in aquifers. *Water Resour. Res*. **19**, 161–180.

Graham, W. & McLaughlin, D. (1989) Stochastic analysis of nonstationary subsurface solute transport, 1. Unconditional moments. *Water Resour. Res*. **25**, 215–232.

Horne, F. E. & Kavvas, M. L. (1997) Physics of the spatially averaged snowmelt process. *J. Hydrol*. **191**, 179–207.

Kabala, Z. J. & Sposito, G. (1991) A stochastic model of reactive solute transport with time-varying velocity in a heterogeneous aquifer. *Water Resour. Res*. **27**, 341–350.

Karakas, A. & Kavvas, M. L. (2000) Conservation equations for groundwater velocity in general conditions. *J. Hydrol. Engng, ASCE* **5**(2), 206–216.

Kavvas, M. L. (2001) General conservation equation for solute transport in heterogeneous porous media. *J. Hydrol. Engng, ASCE* **6**(4), 341–350.

Kavvas, M. L. (2003) Nonlinear hydrological processes: conservation equations for determining their means and probability distributions. *J. Hydrol. Engng, ASCE* **8**(2), 44–53.

Kavvas, M. L. & Karakas, A. (1996) On the stochastic theory of solute transport by unsteady and steady groundwater flow in heterogeneous aquifers. *J. Hydrol., Amsterdam* **179**, 321–351.

Kubo, R. (1962) Generalized cumulant expansion method. *J. Phys. Soc. Japan* **117**, 1100–1120.

Mantoglou, A. & Gelhar, L. W. (1987) Capillary tension head variance, mean soil moisture content, and effective specific soil moisture capacity of transient unsaturated flow in stratified soils. *Water Resour. Res*. **23**, 47–56.

Olver, P. J. (1993) *Applications of Lie Groups to Differential Equations*. Springer-Verlag, New York, USA.

Serrano, S. (1992) The form of the dispersion equation under recharge and variable velocity, and its analytical solution. *Water Resour. Res*. **28**, 1801–1808.

Tayfur, G. & Kavvas, M. L. (1994) Spatially averaged conservation equations for interacting rill-interrill area overland flows. *J. Hydraul. Engng, ASCE* **120**, 1426–1448.

Tayfur, G. & Kavvas, M. L. (1998) Areally-averaged overland flow equations at hillslope scale. *Hydrol. Sci. J*. **43**(3), 361–378.

Wood. B. D. & Kavvas, M. L. (1999) Stochastic solute transport under unsteady flow conditions: comparison of theory, Monte Carlo simulations, and field data. *Water Resour. Res*. **35**, 2069–2084.

Zwanzig, R. (1960) Ensemble method in the theory of irreversibility. *J. Chem. Phys*. **33**(5), 1338–1341.