THEORETICAL HYDROLOGY
WITH RESPECT TO SCALING AND HETEROGENEITY
OVER
SPARSELY-GAGED OR UNGAGED
REGIONS AND BASINS

M.L.Kavvas¹, Z.Q.Chen¹, J.Y.Yoon¹ and J. Yoshitani²

¹Hydrologic Research Lab., Civil & Envr. Engineering, U.C.Davis, CA 95616, U.S.A.

²Indep. Adm. Inst. Public Works Research Institute, Ministry of Land, Infrastructure and Transport, Japan
For the quantification of hydrologic water balances over sparsely gaged or ungaged regions/watersheds

A Central Issue:

How to model the hydrologic processes at the scale of the grid areas of ungaged small-mesoscale watersheds (grid size~1km), and of sparsely gaged/ungaged regions (grid size~10km).
Since there may be very sparse or no precipitation/runoff data over an ungaged/sparsely gaged 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 be able to utilize remotely sensed observations as areally-averaged quantities over such grid areas.

Then in order to have scale-consistent description of the hydrologic processes with respect to both numerical modeling and remotely sensed observations over such grid areas of an ungaged/sparsely gaged watershed, it becomes necessary to develop upscaled hydrologic conservation equations for the hydrologic processes of interest over such grid areas.
spatial scale = observational/computational grid size

time scale = observational/computational time interval
Current state of hydrologic science:

The hydrologic conservation equations are generally known at "point-scale".

Point-scale = scale of differential control volume

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 ~10m to ~100km depending upon the domain being modeled.

In order to utilize these hydrologic conservation equations for modeling the hydrologic processes at the particular scale of a grid area one makes the assumption that

the conservation equation (usually a PDE) at the node represents the whole hydrologic process evolving over the area that surrounds that node.
This amounts to assuming:

**Homogeneity** of soils, vegetation, geology, topography, atmospheric inputs.

over the area (volume) that surrounds any nodal point of the computational grid network.
However, soils, vegetation, geology, topography, atmospheric inputs over an area (volume) that surrounds any nodal point of the computational grid network are heterogeneous.

Therefore, a **hydrologic conservation equation** which 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 behavior of the hydrologic process which is taking place over the grid-scale area that surrounds that node.
Fundamental problem:

How to upscale the existing point-scale hydrologic conservation equations

for

mass, momentum (and/or energy) to the increasingly larger spatial scales,

in order to have

the conservation equations to be consistent with

the scale of the grid areas over which they will describe

the hydrologic processes over ungaged/sparsely gaged watersheds.
Various Approaches to Upscaling of Hydrologic Conservation Equations:

I. Averaging approaches:

A. Volume/areal averaging approaches

B. Ensemble averaging approaches

1. Numerical probabilistic averaging approaches

2. Analytical ensemble averaging approaches

   a. Averaging based on analytical solutions to realizations
   b. Averaging based on the regular perturbations approach
   c. Averaging based upon decomposition theory of Adomian
   d. Averaging based on projector-operator approach
   e. Averaging based on cumulant expansion approach
      i. Cumulant expansion combined with spectral theory
      ii. Cumulant expansion combined with Lie group theory
II. Approaches based upon similarity

   A. Upscaling based upon dimensional similarity theory

   B. Fractional Fokker-Planck equation
A. Volume/Areal Averaging Approach

The point-scale hydrologic conservation equation is integrated over a volumetric or areal domain, and then the resulting integrals are divided by the size of the domain.

Was possible to derive Darcy's Equation from the microscopic Navier-Stokes equations under many simplifying assumptions in order to obtain closure (Whittaker, 1999).

Was used in hydrology to reduce the hydrologic conservation equations from their original PDE forms at point scale to ODE forms at larger spatial scales:

a) Duffy (1996) reduced the unsaturated-saturated subsurface flow conservation equation from its original PDE form to a set of ODEs by means of volume averaging;

b) Tayfur and Kavvas (1994, 1998) reduced rill and interrill overland flow equations from a 2-D PDE at point scale to an ODE at hillslope scale by volume averaging.

This approach has closure problems.
B. **Ensemble averaging approaches**:

Recognize that the point-scale hydrologic 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/behavior of the original point-scale conservation equation (as a stochastic PDE)

- which will represent its upscaled form at the scale of the modeling grid area.
Schematic description of averaging soil water profiles in a computational mesh

Stochastic approach

Integration of non-linear processes (averaging)

Heterogeneity in porosity, unsaturated hydraulic conductivity, and other parameters

Areally-averaged single conservation equation

Soil moisture profile areally-averaged over the mesh derived from single equation
1. **Numerical probabilistic averaging approaches:**


Assign 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, numerically average the point-scale conservation equations over the grid area in order to obtain the grid-area-scale behavior of the corresponding hydrologic process.
2. **Analytical ensemble averaging approaches:**

   a. **Averaging based on analytical solutions to realizations:**
      (Serrano, 1992, 1993, 1995; Chen et al. 1994a, b; etc.)

<table>
<thead>
<tr>
<th><strong>Approach:</strong></th>
<th>Obtain the pathwise analytical solution to the conservation equation, and then take its ensemble average</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advantage:</strong></td>
<td>Possible to obtain exact analytical closures even in nonlinear problems</td>
</tr>
<tr>
<td></td>
<td>Successfully applied to the ensemble averaging of nonlinear unsaturated soil water flow and nonlinear Boussinesq equations</td>
</tr>
<tr>
<td><strong>Drawback:</strong></td>
<td>Solutions are cumbersome and difficult to understand/use by third parties.</td>
</tr>
</tbody>
</table>
b. **Averaging based on the regular perturbations approach:**

The most often used approach in hydrology (Gelhar and Axness, 1983; Dagan, 1982, 1984; Rubin, 1990, 1991; Graham and McLaughlin, 1989; Mantoglou and Gelhar, 1987a,b; Mantoglou, 1992; Tayfur and Kavvas, 1994; Horne and Kavvas, 1997; etc.)

**Approach:** Express each stochastic parameter and each state variable in the conservation equations by a sum of their corresponding mean and a small perturbation term. Then substitute this perturbation expression in place of the original parameter/state variable within the conservation equation. Then take the expectation of the resulting conservation equation to obtain an ensemble average equation for the considered hydrologic process.

**Advantage:** Straightforward to apply even in nonlinear cases.

**Drawbacks:** Immediately results in a closure problem where the equation for the mean requires information about the behavior 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 behavior of even higher moments. Hence, one can close the system of equations only by means of some ad hoc assumption. Small perturbation is often invalid in highly heterogeneous media.
Spatial Horizontally Averaged Richards' Equation

(SHARE) Model

Areal averaging of point-scale Richards' equation:

\[
\frac{\partial}{\partial t} \left( \frac{1}{A} \iint_A \theta \, dx \right) = \frac{\partial}{\partial z} \left[ \frac{1}{A} \iint_A \left( \frac{\partial \Phi}{\partial z} - K \right) \, dx \right] + \frac{1}{A} \iint_A \left( \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} \right) \, dx
\]

\[
\Psi
\]

\[
\Phi(\Psi) = \int K(h) \, dh
\]

\[
<\Phi> (z,t) = \frac{1}{A} \iint_A \Phi(z,t; \gamma) \, dx
\]

\[
<\Phi> (z,t) = \frac{1}{A} \iint_A \Phi(z,t; \gamma) \, dx
\]
The system of equations for \( <s> \) and \( \text{Cov}[s,K_S] \) in SHARE model

\[
(\theta_s-\theta_r) \frac{\partial <s>}{\partial t} = - \frac{\partial <q_z>}{\partial z}
\]

\[
(\theta_s-\theta_r) \frac{\partial \text{Cov}[s,K_S]}{\partial t} = - \frac{\partial}{\partial t} \text{Cov}[q_z,K_S]
\]

\[
\langle q_z \rangle = -\langle K_S \rangle \left[ \frac{\partial \text{Cov}[s,K_S]}{\partial z} - K_r(<s>) \right]
\]

\[
-\frac{\partial}{\partial z} \left[ \Phi_r(<s>) \text{Cov}[s,K_S] \right] + K'_r(<s>) \text{Cov}[s,K_S]
\]

\[
\text{Cov}[q_z,K_S] = \text{Var}(K_S) \left( \frac{\partial \Phi_r(<s>)}{\partial z} - K_r(<s>) \right)
\]

\[
+ \langle K_S \rangle (1+\rho_k^2) \left( \frac{\partial}{\partial z} \left[ \Phi_r(<s>) \text{Cov}[s,K_S] \right] - K'_r(<s>) \text{Cov}[s,K_S] \right)
\]

\[
\rho_k^2 = \frac{\langle [K_S(x) - <K_S>]^2 \rangle}{\langle K_S \rangle \langle K_S \rangle}
\]
Chen et al. (WRR, April 1994) have developed exact analytical expressions for ensemble averaged Green-Ampt soil water flow conservation equations under ponded, infiltration and evapotranspiration boundary conditions (Rectangular Profile Model) where saturated hydraulic conductivity is taken to be a random field.
Rectangular Profile Model Solutions:
Areally-Averaged Green-Ampt Model

Surface held at saturation (Ponded condition)
\[ s(0,t)=1, \quad <s>(0,t)=1, \quad \text{Cov}[s,K_s](0,t)=0 \]

Exact analytical solutions:
\[ <s>(z,t) = \frac{1}{A} \int_A s[t,z;K_s(x)] \, dx \]
\[ = s_i \, P_r[ \, K_s < K_{sz}(z,t) \, ] + \, P_r[ \, K_s > K_{sz}(z,t) \, ] \, s_i \]
where \( K_{sz} = [z - \alpha \, \ln(1+z/\alpha)]/(\mu t) \)
\[ \alpha = \frac{\Phi_s(1) - \Phi_s(s_i)}{1 - K_s(s_i)}, \quad \mu = \frac{1 - K_s(s_i)}{(\theta_s - \theta_s)(1-s_i)} \]

Using the lognormal distribution of \( K_s \)
with given parameters of \( K_m \) and \( \sigma_y \), i.e.
\[ f_{K_s}(K_s) = \frac{1}{K_s \sigma_y \sqrt{2\pi}} \exp \left[ -\frac{\ln^2(K_s/K_m)}{2\sigma_y^2} \right] \]
then,
\[ <s>(z,t) = s_i + \frac{1-s_i}{2} \text{erfc} \left\{ \ln \left( \frac{z - \alpha \cdot \ln(1+z/\alpha)}{\mu \cdot t \cdot K_m} \right) \right\} / 2\sigma_y \]
where \( \text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} \, dt \)
Numerical model of 3-d Richards equation with

Grid Network = 10m x 10m x 2m
Horizontal grid size = 1 m
Vertical grid size = variable (~cm)

was used (Chen et al., WRR, 1994b) to simulate a 3-d soil moisture field under infiltration boundary conditions. Then the simulated soil moisture field was horizontally averaged in order to obtain areally-averaged soil moisture content profiles at different times.

These areally-averaged soil moisture profiles were compared with those, predicted by areally-averaged Green-Ampt model (Rect.Prof.) and by a second-order regular perturbation closure to areally-averaged 1-d Richards equation.
c. **Averaging based upon decomposition theory of Adomian:**
   (Adomian, 1983; Serrano, 1993; 1995a,b)

**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.

**Advantages:** can accommodate any size of fluctuation; can be applied both to linear and nonlinear problems; avoids closure problems by adding successively smaller magnitudes to the solution.

**Drawbacks:** requires a pathwise analytical solution to the conservation equation in order to develop the corresponding ensemble average equation; however, such analytical solutions are unattainable in many nonlinear hydrologic processes.
d. **Averaging based on projector-operator approach**:  

(Nakajima, 1958; Zwanzig, 1960; Cushman, 1991; Cushman and Ginn, 1993)

**Approach:** Considers an operator which projects quantities onto their averages \( (Pu = <u>) \). Then applying this operator together with an operator that represents the difference between the actual variable and its mean \( (Du = u - <u>) \), derives an exact integro-differential equation for the ensemble average. This integro-differential equation is nonlocal.

**Advantages:** Avoids the closure problem.

**Drawbacks:** Applicable only to linear problems. The obtained integro-differential equation is implicit in the state variable. Therefore, it requires further approximations for its explicit solution.
e. **Averaging based on cumulant expansion approach**:


**Approach:** Express the original conservation equation in terms of an operator equation with an average component and a fluctuating dynamic component. Solve 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.
Two approaches for explicit expressions:

i. Cumulant expansion combined with spectral theory:
   (Kabala and Sposito, 1991, 1994)

   Takes the Fourier transform of the cumulant expansion expression in order to develop an equation for the ensemble average in the Fourier space. Still needs to be inverted to the real time-space for practical applications.

ii. Cumulant expansion combined with Lie group theory:
   (Kavvas and Karakas, 1996; Wood and Kavvas, 1999; Karakas and Kavvas, 2000; Kavvas, 2002; Yoon and Kavvas, 2002)

   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 (Serre, 1965; Olver, 1993) in order to obtain an explicit expression for ensemble average conservation equation in real time-space.
A general formula for the upscaling of linear hydrologic conservation equations from point-scale to next larger spatial scale:

Any linear hydrologic conservation equation may be written in the operator form:

$$\frac{\partial h(x,t)}{\partial t} = A(x,t) \ h(x,t)$$  \hspace{1cm} (1)$$

where h is the state variable and A is the operator coefficient function.
"Master Key" differential equation

for the upscaling of any linear hydrologic conservation equation (1)

from point-scale to next larger scale (Kavvas, ASCE JHE, 2002):

$$\frac{\partial <h(x_t,t)>}{\partial t} = <A(x_t,t)> <h(x_t,t)>$$

$$+ \int_0^t ds \ Cov_o[A(x_t,t) ; A(x_{t-s},t-s)] <h(x_t,t)> \quad (2)$$

to the order of the covariance time of the operator A. (Exact second order.)

In equation (2), the Lagrangian location $x_{t-s}$ is obtained from the known location $x_t$ by

$$x_{t-s} = \hat{e}xp \left( \int_{t-s}^t dt \ A_L(x_t,t) \right) x_t \quad (3)$$
$A_L$ is that portion of $<A>$ which is made up of the linear combination of the first spatial derivatives.

As such, the time-ordered exponential operator $\exp(\cdot)$ on the right-hand-side (rhs) of Eqn.(3) is a Lie operator.

Since this Lie operator is fundamentally a displacement operator, it displaces the spatial location $x_t$ at time $t$ to a location $x_t - s$ at time $t - s$, 
Example: Groundwater solute transport by unsteady, spatially nonstationary stochastic flow (velocity $v$ is a time-space stochastic process) is expressed by the following Darcy-scale conservation equation:

$$\frac{\partial c(x,t)}{\partial t} = - v_i(x,t) \frac{\partial c(x,t)}{\partial x_i} + D_{ji} \frac{\partial^2 c(x,t)}{\partial x_j \partial x_i}$$  \hspace{1cm} (4)

Eqn.(4) may be expressed as the operator equation

$$\frac{\partial c(x,t)}{\partial t} = A(x,t) \cdot c(x,t)$$  \hspace{1cm} (5)

where the operator $A(x,t)$ is defined by

$$A(x,t) = - v_i(x,t) \frac{\partial}{\partial x_i} + D_{ji} \frac{\partial^2}{\partial x_j \partial x_i}$$

Then substituting this definition of $A(x,t)$ into the "Master Key" upscaling equation (2) one obtains
Upscaled conservation equation for solute transport at a spatial scale one step larger than the Darcy scale:

(Kavvas and Karakas, J. of Hydrol., 1996)

\[
\frac{\partial <c(x_t,t)>}{\partial t} = \left\{ -<v_i(x_t,t)> + \int_0^t ds \text{Cov}_o \left[ v_j(x_t,t) ; \frac{\partial v_i(x_t - s,t - s)}{\partial x_j} \right] \right\} \frac{\partial <c(x_t,t)>}{\partial x_i} \\
+ \left\{ D_{ji} + \int_0^t ds \text{ Cov}_o[v_j(x_t,t) ; v_i(x_t - s,t - s)] \right\} \frac{\partial^2 <c(x_t,t)>}{\partial x_j \partial x_i}
\]

to the order of the covariance time of A. In this case

\[ A_L(x_t,t) = -<v_i(x_t,t)> \frac{\partial}{\partial x_i} \]

and

\[ x_{t-s} = \exp \left[ -\int_{t-s}^t dt <v_l(x,t)> \frac{\partial}{\partial x_l} \right] x_t \]
APPLICATION TO SOLUTE TRANSPORT OBSERVATIONS
AT BORDEN AQUIFER

(Wood, B.D. and M.L. Kavvas, WRR, 35(7), 1999)
Comparison of the first spatial moments of the ensemble-averaged concentration field with data from the Borden site.
Second spatial moments of the ensemble-averaged concentration field as determined from the upscaled transport equation and from the Monte Carlo simulation. The second moments calculated directly from the Borden aquifer data appear as points.
Plate 1. Ensemble-averaged concentration fields plotted for 11 output times (times are in days). The fields represent the ensemble-averaged concentration determined by the numerical solution of (2) and by the ensemble average of the Monte Carlo simulations. The effect of the transient field can be seen as a rotation of the principal axis of the plume.
APPLICATION TO THE OVERLAND FLOW PROCESS

Point-scale one-dimensional kinematic wave equation as conservation equation for overland flow:

\[
\frac{\dot{Z}y(x,t)}{\dot{Z}t} = - \frac{\dot{Z}}{\dot{Z}x} \left[ \alpha(x) \, y(x,t)^m \right] + q(x,t) \tag{9}
\]

\[
\eta(y,\alpha,m,q;x,t) = - \frac{\partial}{\partial x} \left[ \alpha(x) \, y(x,t)^m \right] + q(x,t) \tag{10}
\]

\[
\frac{\partial y}{\partial t} = \eta(y,\alpha,m,q;x,t) \tag{11}
\]
Fokker-Planck Equation of the upscaled overland flow kinematic wave equation

\[
\frac{\partial P(y(x_t,t),t)}{\partial t} = - \frac{\partial}{\partial y} \left \{ P(y(x_t,t),t) \left [ \alpha(x_t) y(x_t,t)^m + q(x_t,t) \right ] \right \} \\
+ \int_0^1 ds \text{Cov}_o \left [ - \frac{\partial}{\partial x} \left [ \alpha(x_t) y(x_t,t)^m \right ] - \frac{\partial}{\partial x} \left [ \alpha(x_{t-s}) y(x_{t-s},t-s)^m \right ] + q(x_{t-s},t-s) \right ] \\
+ \frac{1}{2} \frac{\partial^2}{\partial y^2} \left \{ 2P(y(x_t,t),t) \cdot \int_0^1 ds \text{Cov}_o \left [ - \frac{\partial}{\partial x} \left [ \alpha(x_t) y(x_t,t)^m \right ] + q(x_t,t) - \frac{\partial}{\partial x} \left [ \alpha(x_{t-s}) y(x_{t-s},t-s)^m \right ] + q(x_{t-s},t-s) \right ] \right \}
\]

One may note that the FPE (12) is linear in \( P \).
However, in order to solve for the probability density $P$ for overland flow depth $y$, it is necessary to resolve the spatial gradient $\frac{\partial y}{\partial x}$. 
Areally averaged overland flow equation

Can be obtained from Eq. (9) using the following sine function as an approximation to the flow depth profile along the x-direction

\[ y(x,t) = y(L_x,t) \sin\left( \frac{\pi x}{2L_x} \right) \] (13)

From (13), the following relation can be derived by integrating on both sides with respect to the x-axis

\[ \bar{y}(t) = \frac{2}{\pi} y(L_x,t) \] (14)
Using the sine function approximation and integrating along the x-axis, Eq. (9) can now be transformed to an areally averaged form as follows.

\[
\frac{d\bar{y}(t)}{dt} = - \beta(L_x) \bar{y}(t)^{3/2} + \bar{q}(t)
\]  
(15)

\[
\beta(L_x) = 1.97 \frac{\alpha(L_x)}{L_x}
\]  
(16)
Fokker-Planck equation of the upscaled overland flow kinematic wave equation

\[
\frac{\partial P(\bar{y}(t))}{\partial t} = \frac{\partial}{\partial \bar{y}} \left\{ P(\bar{y}(t)) \left[ \left\langle \beta(L_x)\bar{y}(t)^{1.5} \right\rangle - \langle q(t) \rangle - \int_0^t ds \text{ Cov}_o\left[ -1.5 \beta(L_x)\bar{y}(t)^{0.5} ; - \beta(L_x)\bar{y}(t-s)^{1.5} + \bar{q}(t-s) \right] \right] \right\} \\
+ \frac{1}{2} \frac{\partial^2}{\partial \bar{y}^2} \left\{ 2 P(\bar{y}(t)) \int_0^t ds \text{ Cov}_o\left[ -\beta(L_x)\bar{y}(t)^{1.5} + \bar{q}(t) ; - \beta(L_x)\bar{y}(t-s)^{1.5} + \bar{q}(t-s) \right] \right\}
\]  

(17)
MC vs FPE

t = 0.5 hr

P(h(L_x))

h(L_x)

0 2 4 6 8

MC
FPE

0.00 0.05 0.10 0.15 0.20 0.25 0.30

0 2 4 6 8

p(h(L_x))

h(L_x)

MC
FPE

0.00 0.05 0.10 0.15 0.20 0.25 0.30
MC vs FPE - continued

Graphs showing the comparison between MC and FPE at different time points: 
- At $t = 2$ hr, the distributions of $h(L_x)$ are compared.
- The graphs illustrating the time evolution of $<h(L_x,t)>$ for both methods.

The plots highlight the differences and similarities in the predictions made by the two methods.
Welcome to the new world of upscaled hydrologic conservation equations!

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;

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 modeled (eg. 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;
3) The spatial heterogeneities due to topography, soils, vegetation, land use/land cover, geology are incorporated explicitly into 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 hydrologic dynamics (due to heterogeneity in land conditions and atmospheric boundary conditions) is explicitly modeled in the upscaled equations.
4) The hydrologic models which are based upon point-scale conservation equations with effective parameters may yield significantly incorrect predictions over highly heterogeneous ungaged basins.

In such basins it may be necessary to utilize upscaled hydrologic conservation equations with their upscaled parameters.

**IT IS ESSENTIAL TO**

establish a hydrologic model intercomparison project by which existing models (point-scale or upscaled) are tested for their performance when they are provided no atmospheric/hydrologic data over a large basin for predicting hydrological processes at various spatial scales within that basin.