ELSEVIER

Advances in Water Resources 23 (2000) 653–665

Advances in Water Resources

A multiscale approach for estimating solute travel time distributions

Michael M. Daniel [a,*], Alan S. Willsky [a], Dennis McLaughlin [b]

[a] *Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[b] *Ralph M. Parsons Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

Received 23 October 1998; received in revised form 3 September 1999; accepted 7 September 1999

Abstract

This paper uses a multiscale statistical framework to estimate groundwater travel times and to derive conditional travel time probability densities. In the applications of interest here travel time uncertainties depend primarily on uncertainties in hydraulic conductivity. These uncertainties can be reduced if the travel times are conditioned on scattered measurements of hydraulic conductivity and/or hydraulic head. In our approach the spatially discretized log hydraulic conductivity is modeled as a multiscale stochastic process, where each scale describes the process at a different spatial resolution. Related dependent variables such as hydraulic head and travel time are approximated by discrete linear functions of the log conductivity. The linearization makes it possible to incorporate these variables into efficient multiscale estimation and conditional simulation algorithms. We illustrate the application of these algorithms by considering two options for estimating travel time densities: (1) a Monte Carlo technique which only requires linearization of the groundwater flow equation and (2) a Gaussian approximation which also requires linearization of Darcy's law and an implicit particle tracking equation. Both options provide reasonable estimates of the travel time probability density in a synthetic experiment if the underlying log hydraulic conductivity variance is small (0.5). When this variance is increased (to 5.0), the Monte Carlo result is still quite good but the Gaussian approximation is unsatisfactory. The multiscale Monte Carlo option is a very competitive approach for estimating travel time since it provides accurate results over a wide range of conditions and it is more computationally efficient than competing alternatives.

1. Introduction

Many hydrologic analyses require information on spatially distributed variables such as rainfall, evapotranspiration, groundwater recharge, and soil hydraulic conductivity. Values for these variables may need to be specified at each cell of a regular computational grid or as averages over larger irregular regions such as watersheds. Field measurements are seldom available in either of these forms. Most in situ measurements of hydrologic variables are recorded only at scattered points, such as meteorological stations or groundwater wells. Remote sensing measurements generally cover larger areas but are only indirectly related to the variables needed for hydrologic analysis. In either case, some sort of data processing or retrieval algorithm must be used to convert raw measurements into more readily usable information. This algorithm may be as simple as Thiessen polygon interpolation or as complex as a variational data assimilation procedure.

Most of the algorithms used to derive estimates of spatially distributed hydrologic variables are based on least-squares principles. Examples include geostatistical techniques such as kriging and Bayesian inverse estimation procedures. These algorithms typically presume that the quantity to be estimated is a random field which can be described in terms of its mean and covariance functions. Although such descriptions are conceptually appealing they lead to computationally demanding algorithms if the random field is finely discretized. For example, a relatively modest 100 by 100 pixel two-dimensional grid yields 10 000 discrete values which are related by a covariance matrix with 10^8 elements. Least-squares computations based on such large matrices are inefficient and undesirable.

The computational difficulties encountered in spatial least-squares estimation do not usually arise in temporal estimation problems. The sizes of the covariance matrices used in temporal estimation algorithms such as the Kalman filter do not depend on the total number of measurements processed [7]. These algorithms have a recursive form which takes advantage of the Markovian structure of the underlying time series model. The

* Corresponding author.

0309-1708/00/$ - see front matter
PII: S0309-1708(99)00039-1

updated estimate computed at any given time depends only on measurements obtained at this time and a prior estimate derived from previously processed measurements. Consequently, the recursive algorithm is much less computationally demanding than equivalent batch processing algorithms.

The above discussion suggests that spatial estimation could be performed more efficiently if it were implemented in a recursive fashion. In order to pursue this point further, suppose that we discretize a continuous random field over a number of nested regular grids with progressively finer resolutions (i.e. smaller pixels). The discretization operation generally assigns a vector of 'states' to each pixel in a given grid. For example, the states for a given pixel could be samples of the continuous field at several specified points inside the pixel boundary. In any case, each of the grids used in a multigrid discretization can be characterized by a particular spatial scale (e.g. the pixel size). Note that here we use the term 'scale' to refer to the spatial discretization process rather than to an intrinsic property of the original random field.

Once we have defined a multiscale discretization we can apply the concept of recursive estimation over scale, in a manner which is similar but not identical to temporal recursion [4,5]. This is feasible if the values of the discretized variable at different scales are related by a Markov statistical model. When the statistical model is appropriately selected a scale-recursive estimator can be much more efficient than a conventional batch estimator operating only at the finest scale.

Multiscale spatial estimation concepts have proven to be useful in a number of image processing and remote sensing problems [6,10–12,15,16]. In this paper we show how such concepts can also be applied to groundwater problems. As an example, we describe a multiscale algorithm which estimates groundwater travel times and generates travel time probability distributions. Our example is a simplified version of the waste isolation performance assessment problem considered in a recent comparison of popular groundwater inverse procedures [19]. In the remaining sections of this paper we review relevant multiscale estimation concepts, describe the travel time estimation problem, and present some typical results. We also discuss some of the advantages and limitations of the multiscale approach.

2. Multiscale estimation

2.1. Background

It is helpful for estimation purposes to distinguish dependent variables (such as hydraulic head and travel time) from independent variables (such as hydraulic conductivity) which are more difficult to observe or model [14]. Independent variables are the primary sources of uncertainty in hydrologic estimation problems and are often characterized as random fields with specified statistical properties. For simplicity, we restrict our attention to time-invariant problems and consider two dependent variables, denoted $\delta h(\mathbf{x})$ and $\delta\tau$, respectively. These dependent variables are related to the independent variable $\delta f(\mathbf{x})$ by the following transformations:

$$\begin{bmatrix} \delta h(\mathbf{x}) \\ \delta\tau \end{bmatrix} = \begin{bmatrix} \mathscr{G}_h(\delta f) \\ \mathscr{G}_\tau(\delta f) \end{bmatrix}, \tag{1}$$

where the $\mathscr{G}(\cdot)$ operators represent solutions to physically based equations which depend on $\delta f(\mathbf{x})$. In the application discussed in this paper, $\delta h(\mathbf{x})$ is the deviation of the random hydraulic head $h(\mathbf{x})$ from the specified nominal head $h_0(\mathbf{x})$, $\delta\tau$ is the deviation of the solute travel time τ from the nominal value τ_0, and $\delta f(\mathbf{x})$ is the deviation of the random log hydraulic conductivity $f(\mathbf{x})$ from the nominal $f_0(\mathbf{x})$. These definitions facilitate linearizations required later in our discussion. Since the nominal values are all known the hydrologic variables of primary interest can always be derived from the deviations as follows:

$$h(\mathbf{x}) = h_0(\mathbf{x}) + \delta h(\mathbf{x}), \tag{2}$$

$$\tau = \tau_0 + \delta_\tau, \tag{3}$$

$$f(\mathbf{x}) = f_0(\mathbf{x}) + \delta f(\mathbf{x}). \tag{4}$$

We assume that $f_0(\mathbf{x})$ is chosen to be the mean of $f(\mathbf{x})$ so that the independent variable $\delta f(\mathbf{x})$ is zero mean. The operators $\mathscr{G}_{\delta h}$ and $\mathscr{G}_{\delta\tau}$ are derived from the groundwater flow equation and the implicit integral equation which relates travel time to log conductivity. Detailed definitions of these operators are provided later in the paper.

For computational purposes all the variables and operators appearing in (1) are generally discretized over space. The discretization procedure replaces the scalar spatial functions $\delta f(\mathbf{x})$ and $\delta h(\mathbf{x})$ with N_f-dimensional vectors, where N_f usually corresponds to the number of cells, pixels, or nodes in the computational grid. To simplify notation, the symbols δf and δh appearing without $\mathbf{x}$ arguments will henceforth be used to indicate these discretized vectors. When the problem is discretized the functional operators $\mathscr{G}_{\delta h}$ and $\mathscr{G}_{\delta\tau}$ are replaced with the discrete operators $G_{\delta h}$ and $g_{\delta\tau}$, which act on the vector δf. Consequently, the discrete version of (1) is

$$\begin{bmatrix} \delta h \\ \delta\tau \end{bmatrix} = \begin{bmatrix} G_h(\delta f) \\ g_\tau(\delta f) \end{bmatrix}. \tag{5}$$

The discretization operation also replaces the scalar covariance function of $\delta f(\mathbf{x})$ with the discrete N_f-dimensional covariance matrix P_f. The complete set of discretized variables can be assembled in a single state vector z

$$z = \begin{bmatrix} \delta f \\ \delta h \\ \delta \tau \end{bmatrix}. \quad (6)$$

The discretized operators and statistics which characterize this state vector constitute a 'finest scale' model of the hydrologic problem of interest.

The unknown variables of the finest scale model may be estimated from measurements of the state variables included in z. In many applications the measurement process can be described by the following linear equation:

$$y = Cz + v, \quad (7)$$

where y is a vector of measurements, C a matrix that identifies the measured elements of z, and v is a random measurement error vector with specified statistics. The objective of the estimation procedure is to derive an 'optimal' estimate of z from y. A reasonable and widely used choice for the optimal estimate is the linear least-squares estimate. This estimate $\hat{z}$ is the conditional mean of z given y, written as $E[z|y]$, when the $G(\cdot)$ operators are linear and z and v are normally distributed [4,14]. Moreover, in this case the estimation error covariance is equal to the conditional covariance of z given y. When the $G(\cdot)$ operators are nonlinear they can sometimes be replaced by linear approximations. In such cases, linear least-squares estimates and error covariances derived from the linearized transformations can be viewed as approximations to the conditional moments.

Although it is possible to formulate a classical least-squares solution to the finest scale estimation problem, the computational requirements are formidable when N_f is large. For this reason, we seek an equivalent representation of the finest scale problem which is generated by a more efficient multiscale model. The structure and derivation of this model are discussed in the next section.

2.2. Multiscale models

The objective of the multiscale modeling discussed here is to provide an efficient method for generating and estimating the random variables included in the state vector z. It is convenient to begin by viewing the model as a recursive algorithm for generating replicates of the random field δf. Then we can extend (or augment) this model so that it can also generate other variables which depend on δf. Once the structure of the multiscale model has been established it is relatively easy to develop a recursive least-squares estimation algorithm.

Multiscale models are typically arranged on inverted trees such as the one shown in Fig. 1(a) [5]. The tree in this figure describes a one-dimensional random field at four different spatial scales, each corresponding to a particular discrete representation of the random field. In practice, the scales frequently correspond to grids of different resolutions. Each scale includes a number of nodes (indicated by filled circles) which are characterized by vectors of state variables. Each node is identified by an index s, with the corresponding state vector denoted by $z(s)$.

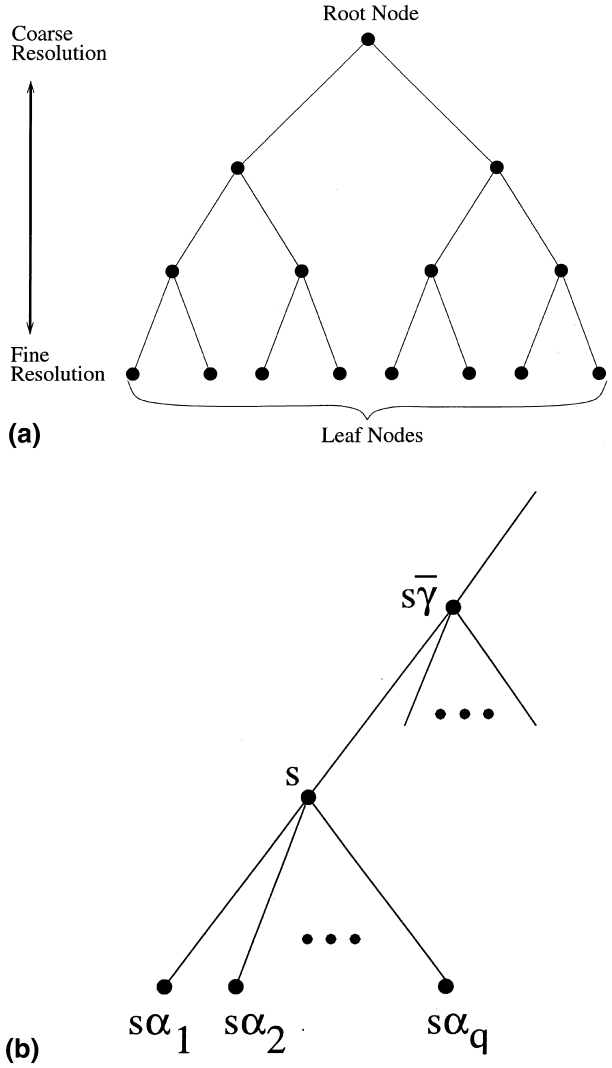

Fig. 1. (a) A simple multiscale tree with nodes defined at four scales and (b) notation used to identify the parent and children of node s.

Fig. 1(b) introduces some standard terminology associated with multiscale models [5]. The nodes at the bottom of the inverted tree are called the 'leaf nodes' while the single node at the top is called the 'root node'. Each node s (except the root node) has a unique parent $s\bar{\gamma}$. Also, each node s (except the leaf nodes) has q_s children $s\alpha_1, s\alpha_2, \ldots, s\alpha_{q_s}$. Any node s partitions the tree into q_{s+1} nodal subsets $S_{s\alpha_1}, S_{s\alpha_2}, \ldots, S_{s\alpha_{q_s+1}}$. The first q_s subsets contain the nodes descended from each of the children of s while the final subset contains the nodes not descended from s. This multiscale formalism is quite

flexible since it can be applied to multidimensional random fields with varying numbers of children and state variables at each node [2,3].

The simple example of Fig. 2 shows how the states of a multiscale tree could be selected for the application of interest here [5]. In this example nine pixels with different log conductivity deviations ($\delta f_k,\ k = 0,\ldots,8$) are defined along a line (so $N_f = 9$). We model this discrete log conductivity field with a simple three-scale tree containing seven nodes (shown by ellipses), each with a three-component state vector. Each node's state vector contains the subset of log conductivity values enclosed by the corresponding ellipse. So the states of the first three nodes are given by:

$$z(0) = \begin{bmatrix} \delta f(0) \\ \delta f(4) \\ \delta f(8) \end{bmatrix}, \quad z(1) = \begin{bmatrix} \delta f(0) \\ \delta f(2) \\ \delta f(4) \end{bmatrix},$$
$$z(2) = \begin{bmatrix} \delta f(4) \\ \delta f(6) \\ \delta f(8) \end{bmatrix}. \tag{8}$$

As we move down the tree the collection of states at each scale provides a progressively higher resolution description of the δf process. Note that the δf process originally defined on nine pixels along the line now appears in the state vectors of the four leaf nodes (with some redundancy). The state vector at each leaf node completely characterizes the finest scale conductivity process in a particular quadrant of the line. Consequently, the entire δf vector can always be extracted from the four states $z(3),\ z(4),\ z(5)$, and $z(6)$.

It is easy to verify that each state vector in the model of Fig. 2 can be expressed as a linear function of δf [5]:

$$z(s) = V_s \delta f. \tag{9}$$

A collection of states produced in this way is called an 'internal realization' and the matrix V_s is called an 'internal matrix' [8]. In our example the V_s matrix for each node is a 3 by 9 array of zeros and ones which may be readily identified from Fig. 2. Internal matrices of this type produce state vectors which are simple subsets of the set of all finest scale log conductivity values. Note that the finest scale δf values generated in an internal realization do not all need to be included in the states at the leaf nodes (as they are in Fig. 2). They can be associated with nodes at different scales, so long as (9) holds.

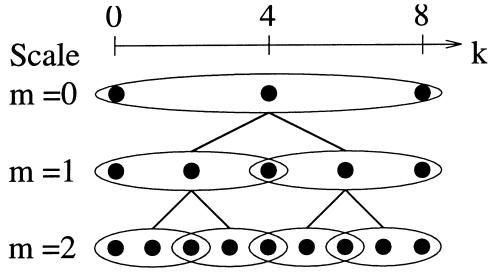

Fig. 2. Three-scale example used to generate a one-dimensional Gauss–Markov process over nine pixels.

Multiscale trees such as Fig. 2 implicitly relate the state at each node to the state at its parent. This relationship can be expressed recursively with the following state transition equation [5]:

$$z(s) = A_s z(s\bar{\gamma}) + w(s), \tag{10}$$

where A_s is the state transition matrix for node s and $w(s)$ is a random 'process noise' vector. The process noise accounts for the additional variability which is revealed when the discretization is refined from the scale of the parent to the scale of the child. We can use this recursion to generate random realizations of δf if we initialize with a random root node vector $z(0)$. The states at all the children of the root node are then derived from (10), with a different random $w(s)$ value added for each child. This process is repeated at each scale.

Since we are configuring the tree to suit our own modeling needs we can select the statistics of the root zone state and the process noise to facilitate subsequent derivations. The moments of $w(s)$ and $z(0)$ are defined as follows:

$$E[w(s)] = 0, \quad E\left[w(s)w(r)^{\mathrm{T}}\right] = Q_s \delta_{rs}, \quad E[z(0)] = 0,$$

$$E\left[w(s)z(0)^{\mathrm{T}}\right] = 0, \quad E\left[z(0)z(0)^{\mathrm{T}}\right] = P_0 \quad \forall s, r.$$

Since all states are linearly related to $z(0)$ and $w(s)$, they are all zero mean. Moreover, since the $w(s)$ vectors are uncorrelated with one another ('white') the statistical properties of the states at nodes $s > 0$ depend only on the root node covariance P_0 and the autoregression parameters A_s and Q_s.

We can now consider the problem of identifying a multiscale model which generates a δf field with a specified covariance P_f. This is the multiscale 'realization problem' [8,9]. When the multiscale model is internal, (6) can be used to relate P_f to the covariance $P_{z(s)z(s\bar{\gamma})}$ between the state at any node s and the state at the parent of s

$$P_{z(s)z(s\bar{\gamma})} = P^{\mathrm{T}}_{z(s\bar{\gamma})z(s)} = V_s P_f V^{\mathrm{T}}_{s\bar{\gamma}}. \tag{11}$$

If the state covariances are consistent with this relationship, the δf values generated on the tree will have a covariance P_f. Arguments from least-squares estimation theory can be used to show that the recursion will satisfy (11) if the error covariances and transition matrix are given by:

$$P_0 = V_0 P_f V_0^{\mathrm{T}}, \tag{12}$$

$$Q_s = P_{z(s)} - P_{z(s)z(s\bar{\gamma})} P_{z(s\bar{\gamma})}^{-1} P_{z(s\bar{\gamma})z(s)}, \tag{13}$$

$$A_s = P_{z(s)z(s\bar{\gamma})} P_{z(s\bar{\gamma})}^{-1}, \tag{14}$$

where $P_{z(s)} = P_{z(s)z(s)}$ is the autocovariance of $z(s)$ [4,5]. If we apply these equations to the example of Fig. 2 it is relatively easy to show that the series of nine δf values contained in the leaf node state vectors will have a covariance P_f.

The concepts described above apply to any internal multiscale model which satisfies (9). For example, we could use a model which includes four rather than three of the finest scale conductivity values in each state vector. Although V_s and the matrices defined in (12)–(14) would change, the covariance of the δf values generated by the recursion would still be P_f. Generally speaking, many different multiscale models can be used to generate a δf field with specified statistics. Some of these may be more convenient or more computationally efficient than others. In practical applications multiscale realization is somewhat of an art, although several investigators provide helpful guidelines [5,8,13].

It can be shown that the whiteness of the process noise $w(s)$ imparts a useful multiscale Markov property to the tree [3]. This property states that the conditional covariance between any of the $q_s + 1$ random vectors partitioned by s is 0:

$$E\left[z(r)z^{\mathrm{T}}(p) \mid z(s)\right] = 0$$
$$\text{for } r \in S_{s\alpha_i},\ p \in S_{s\alpha_j},\ i \neq j,\ i,\ j = 1, \ldots, q_s + 1.$$

The multiscale Markov property decouples information from the nodes above and below s in the tree, in much the same way that the well-known temporal Markov property decouples information before and after the current time. The decoupling enables state estimates to be computed recursively in successive upward and downward sweeps through the tree.

In many cases the multiscale characterization $[z(s),\ P_0,\ Q_s,\ A_s,\ s \in S_0]$ can provide a more efficient description of the finest scale process than the conventional discrete characterization $[\delta f,\ P_f]$. If the state dimension d is the same for all nodes the multiscale storage requirement is $O(d^2 N_f)$ while the storage required by P_f is $O(N_f^2)$. The difference can be dramatic for large N_f, so long as d does not depend on N_f. This is the case when the tree describes a one-dimensional Gauss–Markov process, as in the example discussed above. It is also the case for certain two-dimensional random field models, such as the $1/f$ fractal model used in [6] to process satellite altimetry data. When the process is two-dimensional Gauss–Markov, the state vector needs to include enough finest scale values to partition the nodes on the tree according to the multiscale Markov criterion cited above [4]. In practice, this means that d must be $O(N_f^{0.5})$, giving a multiscale storage requirement of $O(N_f^2)$, the same order as the storage required by P_f. We return to this topic at the end of the paper.

With the multiscale model of δf in place we can consider how discretized dependent variables and associated physical relationships can be included. This is accomplished by augmenting the state $z(s)$ at each node of the tree [5]. Since the augmentation process requires $G_h(\delta f)$ and $g_\tau(\delta f)$ to be linear in δf these transformations must be derived from linear approximations to the applicable flow and travel time equations. We discuss the linearization process in more detail in the next section and focus for now on the methods used to incorporate dependent variables that are linear combinations of independent variables.

The state augmentation procedure must: (1) allow the desired dependent variable to be constructed from the states defined at a particular node and (2) preserve the multiscale Markov property throughout the tree, so that the desired finest scale log conductivity covariance is preserved. This process is discussed in detail in [4,5]. Here we summarize the general concepts and then illustrate them with a simple example.

The first step in the state augmentation procedure is to assign the dependent variable (say $\delta\tau$) to a particular node r. This means that information available in the augmented state vector at r will be used to construct $\delta\tau$. We require the function $g_\tau(\delta f)$ to be linear (or to be approximated by a linear function) so that we can write it as $g_\tau(\delta f) = g_\tau^{\mathrm{T}} \delta f$, where the vector g_τ is a set of N_f known weighting coefficients. The second step of the augmentation procedure is to expand this linear combination into a series of partial linear combinations as follows:

$$g_\tau^{\mathrm{T}} \delta f = g_{\tau,r\alpha_1}^{\mathrm{T}} \delta f_{r\alpha_1} + g_{\tau,r\alpha_2}^{\mathrm{T}} \delta f_{r\alpha_2} + \cdots + g_{\tau,r\alpha_{q_{r+1}}}^{\mathrm{T}} \delta f_{r\alpha_{q_{r+1}}}, \tag{15}$$

where $\delta f_{r\alpha_k}$ is the vector of independent variables contained in the states of nodes in $S_{r\alpha_k}$. Recall that $S_{r\alpha_k}$ is the set containing all nodes descended from $r\alpha_k$ (when $k \leqslant q_r$) or all nodes not descended from $r\alpha_k$ (when $k = q_{r+1}$). The augmentation is completed by adding each of the partial linear combinations in this series to the state vector at node r. After the state vector at r is augmented, each of the descendants of r is augmented in a similar way, down to the leaf nodes, which are not augmented. If a partial linear combination adds to new information about the descendants of a given node, it is not included. The procedure may be repeated in a recursive fashion when several dependent variables need to be incorporated into the multiscale model. If redundancies arise the size of the augmented state can be reduced to eliminate them [4,5].

We can illustrate the state augmentation procedure using the simple multiscale model of Fig. 2. For the moment, suppose that $\delta_\tau = \sum_{k=0}^{8} g_{\tau k} \delta f(k)$ is defined to

be a weighted sum of the nine finest scale log conductivity deviations, where the weights $g_{\tau k}$ have been derived from an appropriate physical model. This dependent variable can be assigned to the root node (i.e. $r = 0$). An acceptable augmentation can then be obtained by including partial weighted sums of the finest scale process values in the nodal state vectors, as prescribed above. The complete state vectors at nodes 0, 1, and 2 are then:

$$z(0) = \begin{bmatrix} \delta f(0) \\ \delta f(4) \\ \delta f(8) \\ S_{0,4} \\ S_{4,8} \end{bmatrix}, \quad z(1) = \begin{bmatrix} \delta f(0) \\ \delta f(2) \\ \delta f(4) \\ S_{0,2} \\ S_{2,4} \end{bmatrix},$$

$$z(2) = \begin{bmatrix} \delta f(4) \\ \delta f(6) \\ \delta f(8) \\ S_{4,6} \\ S_{6,8} \end{bmatrix}, \tag{16}$$

where the partial sum $S_{i,j} = \sum_{k=i}^{j} g_{\tau k} \delta f(k)$. The desired value of $\delta\tau$ can be obtained from an appropriate linear combination of the augmented root node states.

The augmented state definitions of (16) can be used to define new somewhat larger V_s matrices. Then (12)–(14) can be used to derive the augmented covariances and transition matrix needed to generate both δf and $\delta\tau$. Other augmentation strategies could achieve the same result as (16). For example, $\delta\tau$ can be assigned to a node further down the tree than the root node. Since some choices will give a lower total number of states than others, it is best to tailor the augmentation strategy to the particular problem to be solved [5].

2.3. Multiscale estimation and conditional simulation

The multiscale modeling concepts discussed above can be used to develop efficient estimation and conditional simulation algorithms [3–6]. In particular, suppose that we wish to estimate the log hydraulic conductivity and travel time deviations δf and $\delta\tau$ from scattered measurements of log conductivity and hydraulic head. We can describe the log conductivity field and linear approximations to the head and travel time with an augmented multiscale model similar to the one presented in the previous section. The vector of measurements $y(s)$ available at node s on the tree is given by the following measurement equation, which has the same form as (7)

$$y(s) = C_s z(s) + v(s), \tag{17}$$

where C_s is a specified measurement matrix and $v(s)$ is a random measurement error with the following specified statistical properties:

$$E[v(s)] = 0, \quad E\left[v(s)\delta f^{\mathrm{T}}\right] = 0, \quad E\left[v(r)v^{\mathrm{T}}(s)\right] = R\delta_{rs} \quad \forall s, r.$$

Eqs. (10) and (17) together contribute three sources of uncertainty to the estimation problem: (1) the random root node state, (2) the random process noise vectors defined at all nodes below the root node, and (3) the random measurement errors. Optimal (linear least-squares) estimates of the state $z(s)$ given the measurements $y(s)$ can be computed with a multiscale recursive algorithm which is based on (10) and (17). This algorithm is executed in two passes. The first pass starts at the root node and moves downward to progressively finer scales, updating estimates at nodes where measurements are assigned. The second pass moves upward from the leaf nodes, merging estimates from the children of each node and performing a second update at each measurement node. The complete estimation algorithm is described in [3,6]. It is a generalization of the Rauch-Tung-Striebel smoother used to estimate the states of time series models. This multiscale estimation algorithm is derived using least-squares estimation concepts. However, the effort needed to compute state estimates on a tree with N_f leaf nodes and a fixed state vector size of d is $\mathrm{O}(d^3 N_f)$, as compared to $\mathrm{O}(N_f^3)$ to solve the classical least-squares normal equations. So the efficiency of the multiscale approach depends strongly on the ability to keep d small and, ideally, independent of N_f. Some efficiency gain is obtained even for the two-dimensional Gauss–Markov case, where d is proportional to $N_f^{0.5}$. In this case, the multiscale effort is $\mathrm{O}(N_f^{2.5})$ rather than $\mathrm{O}(N_f^3)$.

Like the Kalman filter, the multiscale estimator provides valuable information on the accuracy of its estimates. The estimation error $e(s)$ at any node is $e(s) = z(s) - \hat{z}(s)$, where $\hat{z}(s)$ is the final estimate computed after both passes of the algorithm. On the second pass the estimator computes the covariance $P_{e(s)}$ of $e(s)$ [4]. If all error sources are normally distributed and the linearity assumptions (or approximations) adopted in the state augmentation discussion hold, the estimation error is also normally distributed and $\hat{z}(s)$ and $P_{e(s)}$ are the conditional mean and covariance of $z(s)$. In this case the conditional covariance may be used to construct a confidence region around the estimate $\hat{z}$. We adopt the normality assumption in the remainder of this paper.

The multiscale estimation equations may be manipulated to provide a scale-recursive model for the error $e(s)$ [4]. This model has the same general form as (10)

$$e(s) = J_s e(s\overline{\gamma}) + \tilde{w}(s), \tag{18}$$

where $J(s)$ is a coefficient matrix and $\tilde{w}(s)$ is a zero mean white random variable. The recursion is initialized with a zero mean random root node error $e(0)$. Given the normality assumption made earlier, both $\tilde{w}(s)$ and $e(0)$

are normally distributed. The matrix $J(s)$, the covariance $\tilde{Q}_s$ of $\tilde{w}(s)$, and the covariance $P_e(0)$ of $e(0)$ may all be derived from conditional covariances computed by the estimation algorithm [4].

The error model of (18) provides a convenient way to generate log conductivity realizations which are conditioned on the measurements used by the estimator. A given realization is obtained by generating normal random variables $e(0)$ and $\tilde{w}(s)$ with the appropriate covariances and propagating the error downwards from the root node to the leaf nodes. The conditional realization of the state at node s is obtained from $z(s) = \hat{z}(s) + e(s)$. The desired conditional log conductivity realization can then be extracted in the usual way from the collection of conditional leaf node states. Each realization derived in this way requires $O(d^2 N_f)$ computations. This can be considerably more efficient than alternative conditional simulation methods (such as those described in [19]), even when the log conductivity field is Gauss–Markov and d is $O(N_f^{0.5})$. The travel time example discussed in the next section illustrates both the estimation and conditional simulation aspects of the multiscale approach.

3. The travel time estimation problem

Solute travel times between specified points or boundaries can frequently provide useful information about the aggregate effects of heterogeneous subsurface flow processes [18]. The travel times of natural tracers, such as certain radioisotopes, can be used to identify areas and rates of groundwater recharge. The travel times of wastes released from underground storage facilities to the accessible environment provide valuable information for assessing possible exposure risks. Travel time is particularly important in radioactive waste disposal applications, since the exposure level for any given constituent depends on the ratio of travel time to half-life [19]. When the factors controlling solute transport are highly variable and uncertain it is best to use probabilistic approaches to characterize travel times and related risks. This viewpoint has influenced repository licensing criteria, which have sometimes been expressed in terms of the probability that the travel time will be less than some specified threshold [19]. In order to work with such criteria we need methods for deriving travel time probability densities.

The problem of deriving solute travel time densities is well suited for the multiscale modeling approach described in this paper. If the log hydraulic conductivity field is a normally distributed random variable and we have measurements of log conductivity and other variables such as hydraulic head, we can use multiscale concepts to derive a travel time probability density which is conditioned on the measurements. This can be done at two different levels of approximation.

The first (Monte Carlo) option is to generate conditional realizations of a normally distributed log conductivity at the nodes of a multiscale model, following the method discussed earlier. Each log conductivity realization yields distinct velocity and travel time realizations, which may be derived from the equations governing groundwater flow and advective transport along a streamline. The desired travel time probability density is constructed from the resulting ensemble of simulated travel times. The key approximation made in this approach is linearization of the log conductivity-to-head transformation, which is required to derive the conditional log conductivity realization. Since the log conductivity–travel time transformation is not linearized the Monte Carlo travel time density may not be normal.

The second (Gaussian) option is based on the observation that the travel time is normally distributed if the log conductivity is normal and the conductivity-to-travel time transformation is linear. In this case, the conditional travel time density is completely characterized by its mean and variance. These conditional moments may be derived from a single run of the multiscale model. This option is less computationally demanding than the Monte Carlo approach but requires linearization of both the log conductivity-head and the log conductivity-to-travel time transformations.

In order to put the travel time problem into a multiscale framework we need to specify the statistics of the log conductivity field and the physical relationships between log conductivity, head, and travel time. It is convenient to begin with a non-discretized description of the governing equations. We then carry out the required linearizations and perform a finest scale spatial discretization. Lastly, we formulate and apply the multiscale model. To illustrate the general concepts, we consider two-dimensional transport in a time-invariant groundwater flow field with a known recharge term and an uncertain hydraulic conductivity. The relevant non-discretized flow equation is:

$$\begin{aligned} &\nabla \cdot \left[\mathrm{e}^{f(\mathbf{x})} \nabla h(\mathbf{x}) \right] - Q(\mathbf{x}) = 0, \quad \mathbf{x} \in \Omega, \\ &h(\mathbf{x}) = h_b(\mathbf{x}), \quad \mathbf{x} \in \partial\Omega_D, \\ &-\mathrm{e}^{f(\mathbf{x})} \nabla h(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) = q_b(\mathbf{x}), \quad \mathbf{x} \in \partial\Omega_N, \end{aligned} \tag{19}$$

where $h(\mathbf{x})$ and $f(\mathbf{x})$ are the head and log-conductivity functions and $Q(\mathbf{x})$ is a known recharge rate. Dirichlet (specified head) conditions apply on the boundary $\partial\Omega_D$ of the computational domain Ω while Neumann (specified flux) conditions apply on $\partial\Omega_N$. The groundwater velocity is given by Darcy's law:

$$\mathbf{u}(\mathbf{x}) = -\frac{1}{\theta(\mathbf{x})} \mathrm{e}^{f(\mathbf{x})} \nabla h(\mathbf{x}), \tag{20}$$

where θ is the effective porosity.

We now consider a solute particle moving without dispersion along a path starting at $\mathbf{x}(0)$ at time 0 and ending at time τ somewhere on a control plane located at $x_1(\tau) = L$. The geometry is illustrated in Fig. 3. We assume for simplicity that the dominant velocity component is in the x_1 direction so that the probability that a particle does not exit through the control plane is negligible. The particle path is defined implicitly by the following velocity integral:

$$\mathbf{x}(t) = \mathbf{x}(0) + \int_{t'=0}^{t} \mathbf{u}[\mathbf{x}(t')]\,\mathrm{d}t' \tag{21}$$

and the travel time τ is obtained by imposing the requirement that $x_1(\tau) = L$. Eqs. (19)–(21) provide the information we need to relate head and travel time to log conductivity. Since the resulting relationships are nonlinear and our state augmentation is based on linear transformations, the governing equations need to be linearized about appropriate nominals.

The linearization process is well documented in the stochastic groundwater literature [14,17,18]. A simple approach is to assume that the deviations $\delta f(\mathbf{x})$, $\delta h(\mathbf{x})$, and $\delta\tau$ are small. We then replace $f(\mathbf{x})$, $h(\mathbf{x})$ and τ in (19)–(21) by the perturbation expressions given in (2)–(4). In a similar way, the velocity $\mathbf{u}(\mathbf{x})$ is replaced by $\mathbf{u}(\mathbf{x}) = \mathbf{u}_0(\mathbf{x}) + \delta\mathbf{u}(\mathbf{x})$. The nominal values $h_0(\mathbf{x}), \mathbf{u}_0(\mathbf{x})$, and τ_0 used in the perturbation expansions are the exact head, velocity, and travel time solutions obtained when $f(\mathbf{x})$ is set equal to $f_0(\mathbf{x})$ in (19)–(21). For simplicity, we suppose that the nominal (or mean) log conductivity $f_0(\mathbf{x}) = f_0$ is constant. Also, we assume that the coordinates have been defined so that $\nabla h_0(\mathbf{x})$ and $\mathbf{u}_0 = u_{10}$ point in the x_1 direction (normal to the control plane). These assumptions simplify notation but can be readily relaxed since they are not essential to the approach.

If the expansions of (2)–(4) are substituted into (19)–(21) and products of perturbation terms are neglected, the following linearized equations result:

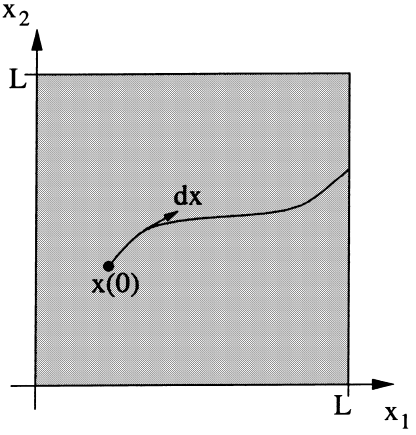

Fig. 3. Solute source and control plane locations for travel time calculation.

$$\begin{aligned}
&\nabla^2\delta h(\mathbf{x}) - \nabla\cdot[\delta f(\mathbf{x})\nabla h_0(\mathbf{x})] = 0, \quad \mathbf{x}\in\Omega,\\
&\delta h = 0, \quad \mathbf{x}\in\partial\Omega_D,\\
&\nabla\delta h(\mathbf{x})\cdot\mathbf{n}(\mathbf{x}) = \delta f[\nabla h_0(\mathbf{x})\cdot\mathbf{n}(\mathbf{x})], \quad \mathbf{x}\in\partial\Omega_N,
\end{aligned} \tag{22}$$

$$\delta\mathbf{u}(\mathbf{x}) = -\frac{1}{\theta}\mathrm{e}^{f_0}[\delta f(\mathbf{x})\nabla h_0(\mathbf{x}) + \nabla\delta h(\mathbf{x})], \tag{23}$$

$$\delta\tau = -\frac{1}{u_{10}}\int_{t'=0}^{\tau_0}\delta u_1[x_{10}(t')]\,\mathrm{d}t', \tag{24}$$

where $x_{10}(t)$ is the path traced by a particle advected by the known nominal velocity u_{10}. The first of these expressions is a partial differential equation which has a solution of the following form [4]:

$$\begin{aligned}
\delta h(\mathbf{x}) = &\int_{\Omega}[\nabla_{x'}G(\mathbf{x},\mathbf{x}')\cdot\nabla_{x'}h_0(\mathbf{x}')]\delta f(\mathbf{x}')\,\mathrm{d}\mathbf{x}'\\
&+\int_{\Omega_N}G(\mathbf{x},\mathbf{x}')[\nabla_{x'}h_0(\mathbf{x}')\cdot\mathbf{n}(\mathbf{x}')]\delta f(\mathbf{x}')\,\mathrm{d}\mathbf{x}',
\end{aligned} \tag{25}$$

where $G(\mathbf{x},\mathbf{x}')$ is the Green's function for (22). Eq. (25) can be substituted into (23) and then (23) substituted into (24) to give a set of explicit integral expressions which relate $\delta h(\mathbf{x})$, $\delta\mathbf{u}(\mathbf{x})$, and $\delta\tau$ to $\delta f(\mathbf{x})$. The expressions for $\delta h(\mathbf{x})$ and $\delta\tau$ are linearized versions of (1)

$$\begin{bmatrix}\delta h(\mathbf{x})\\ \delta\tau\end{bmatrix} = \begin{bmatrix}\mathscr{G}_h\delta f\\ \mathscr{G}_\tau\delta f\end{bmatrix}, \tag{26}$$

where the linear $\mathscr{G}_h\delta f$ and $\mathscr{G}_\tau\delta f$ operators represent convolutions of known functions over the unknown perturbation $\delta f(\mathbf{x})$.

The non-discretized description of the travel time problem presented above can be put in a discrete form appropriate for multiscale modeling if we construct a regular finest scale computational grid with the discrete values of $\delta f(\mathbf{x})$, $\delta h(\mathbf{x})$, and $\delta\mathbf{u}(\mathbf{x})$ defined at the center of each pixel. The elements of the covariance matrix of the discrete vector δf may be obtained from the relationship $P_{f,ij} = P_f(\mathbf{x}_i,\mathbf{x}_j)$, where $P_f(\mathbf{x},\mathbf{x}')$ is the covariance function of the non-discretized $\delta f(\mathbf{x})$ process and $\mathbf{x} = \mathbf{x}_i$ and $\mathbf{x}' = \mathbf{x}_j$ are the coordinates of the center of pixel (i,j). The convolution integrals derived from (22)–(25) can be expressed in a compatible discrete form if trapezoidal integration is used to write all integrals as summations. The discretized flow equation Green's function which appears inside the summations can be obtained by solving the discretized adjoint of (22) at locations where the head is required (head measurement locations and locations along the particle path where velocity must be computed).

After the discretization procedure is completed, the equations relating the discrete dependent variables δh and $\delta\tau$ to the discrete independent variable δf may be written in the form

$$\begin{bmatrix}\delta h\\ \delta\tau\end{bmatrix} = \begin{bmatrix}G_h\delta f\\ g_\tau^{\mathrm{T}}\delta f\end{bmatrix}, \tag{27}$$

where G_h is a matrix and g_τ^T is a vector of weighting factors derived from the discretized convolution integrals. The weighting factors specify the influence that a discrete log conductivity perturbation at any given pixel has on the head perturbation at another pixel (in the case of G_h) or on the travel time from the source to the control plane (in the case of g_τ). In this sense, they can be viewed as sensitivity derivatives [14].

We now have a discrete problem formulation which fits into the multiscale framework. Consequently, we can construct an augmented multiscale model which generates finest scale log conductivity perturbations as well as head and travel time perturbations. In order to complete the multiscale model we need to assign the measured heads and the travel time to particular nodes on the multiscale tree. Since this process is application-specific the details are discussed in the next section where we present an example.

The procedure outlined above is quite general. It can be applied to a wide range of hydrologic problems described by differential equations derived from physical principles. The primary limitation imposed is the need to linearize some or all of the independent-to-dependent variable relationships. In our case, this was accomplished by linearizing the governing differential equations, expressing the solutions as convolution integrals, and then discretizing. Other approaches which give the same final outcome could be used. The linearization operation introduces approximations which will affect the estimates and conditional simulations generated by the multiscale model. In any given application care must be taken to establish the range of validity of this operation. This topic is investigated in more detail in the next section.

4. Examples of multiscale travel time estimation

In this section we illustrate multiscale travel time estimation with a two-dimensional example based on synthetically generated measurements. A synthetic experiment gives a good feeling for the effect of linearizations and other assumptions since the state variables which generate the synthetic measurements are known perfectly. Of course, the merits of the approach cannot be completely assessed until it is tested with real field data since the synthetic experiment inevitably relies on assumptions that may not be valid in practice (e.g. time-invariant two-dimensional flow, perfectly known boundary conditions, conservative transport, etc.). For present purposes, we are concerned primarily with computational efficiency and the effects of linearization assumptions. Both of these can be conveniently tested with a synthetic experiment.

In the synthetic experiment we generate a single unconditional log hydraulic conductivity field with the multiscale model. The head, velocity, and travel time corresponding to this field and the specified boundary conditions are derived from (19)–(21). Log conductivity and head values at specified measurement locations are identified and corrupted with zero mean random errors having the appropriate covariance, to produce a set of synthetic measurements. Conditional moments and multiple conditional realizations of the log conductivity, estimation error, and travel time are then derived from these measurements.

The following specifications define the synthetic experiment:

- The region of interest is the unit square $\mathbf{x} \in [0,1] \times [0,1]$.
- Travel time is computed from a source at $x(0) = (0.25, 0.5)$ to a control plane at $x_1 = 0.75$.
- The non-discretized log hydraulic conductivity is a zero mean two-dimensional Gauss–Markov process with an exponential covariance function. The log conductivity variance is either 0.5 or 5.0 and the correlation distance in each direction is 1.5.
- The time-invariant groundwater head satisfies (19) with $Q = 0$, no-flux conditions across the $x_2 = 0$ and $x_2 = 1$ boundaries, and specified heads of $h(0, x_2) = 1$ and $h(1, x_2) = 0$.
- Head and travel time are linearized about the constant nominal (mean) log conductivity value $f_0(x) = 0$. The corresponding nominal travel time is 0.1.
- The finest scale numerical grid has 33-by-33 ($nx = ny = 33$) pixels regularly spaced on the unit square.
- Synthetic head and conductivity measurements are generated at each of the 20 clustered locations shown in Fig. 4.
- The log conductivity measurement errors are zero mean with a variance of 0.1.

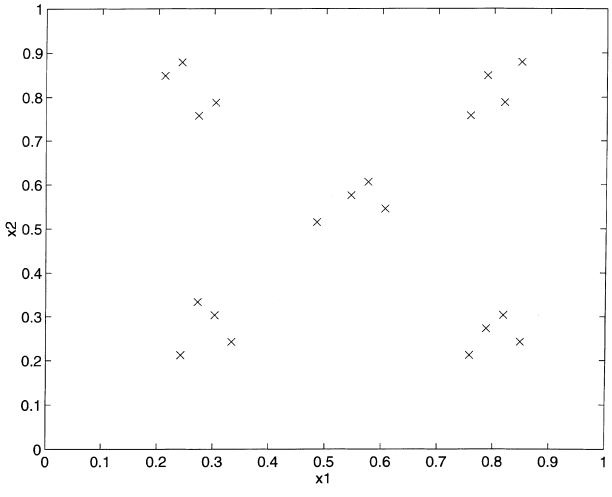

Fig. 4. Log conductivity and head measurement locations.

- The head measurement errors are zero mean with a variance of 0.001.
- The number of Monte Carlo realizations for computing the travel time probability density is 400.

We derive travel time probability densities using both the Monte Carlo and Gaussian approaches. The multiscale model is implemented on a quadtree with four children assigned to each node except the leaf nodes. The tree has five scales which describe the log conductivity field at progressively finer resolutions. The number and arrangement of states at each scale are selected to meet the requirements of the multiscale Markov property.

The linearized travel time is assigned to the root node while the linearized heads at measurement locations are assigned to various nodes, depending on the location of their cluster. Measured heads in the central of Fig. 4 are assigned to the root node. Measured heads in each of the remaining clusters are assigned to the Scale 1 node associated with the spatial quadrant containing that cluster. This assignment strategy increases the size of each of the Scale 0 and Scale 1 state vectors to about 100. The sizes of the augmented state vectors at the remaining scales are smaller.

Fig. 5(a) shows the synthetic experiment log hydraulic conductivity realization generated for a variance of 0.5. This realization yields a relatively smooth hydraulic head profile along the x_1 axis. The resulting 'true' travel time is 0.149. The 20 log conductivity and 20 head measurements derived from the realization of Fig. 5(a) can be used to estimate the states at every node of the multiscale tree. The finest scale log conductivity values extracted from these state estimates are plotted in Fig. 5(b) and the corresponding estimation error variance is plotted in Fig. 5(c). The log conductivity estimate is considerably smoother than the underlying random replicate while the estimation error variance decreases modestly in the vicinity of the measurements. Both results indicate that the measured data provide relatively little information about the true log conductivity. This behavior is typical of groundwater flow inverse problems which are not forced by an internal specified flux (such as a pumping or injection well). In such problems, log conductivity measurements provide only local information and head measurements are relatively insensitive to moderate variations in the log conductivity. Nevertheless, these measurements contain useful information about more global quantities such as travel time.

This can be seen in Fig. 6, which shows the travel time probability densities obtained from the Monte Carlo and Gaussian approaches for the 0.5 log conductivity variance case. The Monte Carlo result (solid line) is presented as a histogram constructed from 400 additional random realizations similar to the one shown in Fig. 5(a). The Gaussian result (dashed line) is a smooth density function which depends only on the travel time

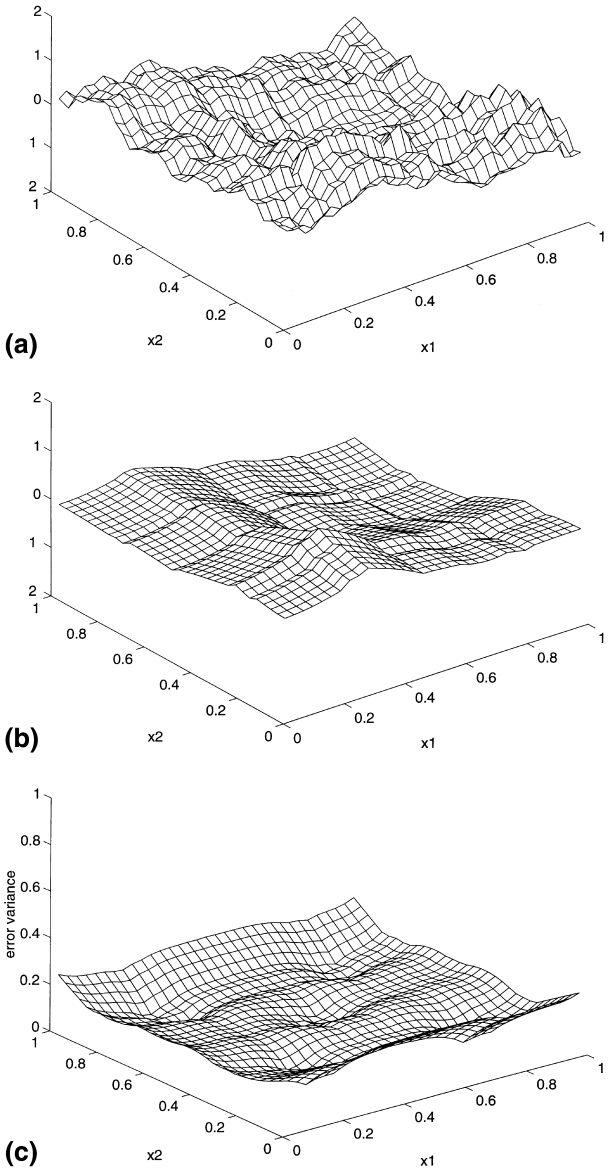

Fig. 5. (a) A typical unconditional log conductivity realization for $\sigma_f^2 = 0.5$. (b) Finest scale log conductivity estimate based on log conductivity and head measurements generated from the realization plotted in (a). (c) The corresponding estimation error variance.

estimate (conditional mean) and estimation error variance (conditional variance) obtained from the two-pass multiscale estimation algorithm. The symmetric Gaussian approximation has a mean of 0.146 while the mean of the slightly skewed Monte Carlo density is 0.155. These both compare well to the true value of 0.149. These results are typical of experiments performed with a log conductivity variance of 0.5. The accuracy of the Gaussian and Monte Carlo travel time estimates and the similarity of the two densities suggests that the linearizations adopted in the multiscale model are good approximations for this low variance case.

The situation changes considerably when the variance is raised. Fig. 7 shows the same information as Fig. 6 for

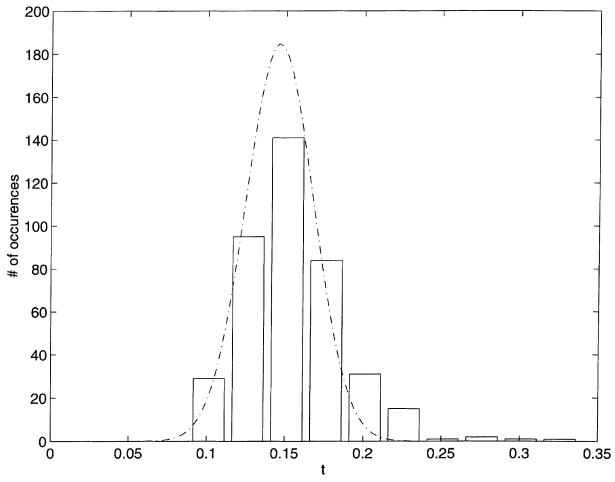

Fig. 6. Travel time histogram from Monte Carlo method and probability density function from Gaussian approximation for $\sigma_f^2 = 0.5$.

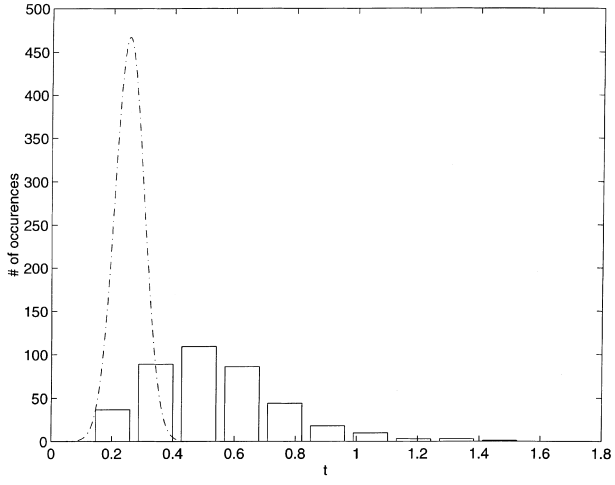

Fig. 7. Travel time histogram from Monte Carlo method and probability density function from Gaussian approxiamtion for $\sigma_f^2 = 5.0$.

a log conductivity variance of 5.0, an order of magnitude larger. Variances of this magnitude are quite high but can be observed at particularly heterogeneous sites. For present purposes, the large variance case is of interest primarily because it provides a demanding test of the linearizations used in both the Gaussian and Monte Carlo approaches. The true travel time obtained for the 5.0 variance case shown in Fig. 7 is 0.567. The Monte Carlo mean of 0.536 compares well to the true value but the Gaussian mean of 0.246 is less than half of the true travel time. Moreover, the symmetric and relatively narrow Gaussian density differs significantly from the Monte Carlo density, which has a more log normal shape. In some large variance cases simulated in our experiments the Gaussian density extended well into the region of negative travel times. The Monte Carlo approach cannot produce such unrealistic results since it accurately simulates the movement of the solute particle along the streamline from the source to the control plane. The Gaussian approach replaces this simulation with a linear approximation which leads to problems when the log conductivity variance is large.

The results presented in Figs. 6 and 7 suggest that the log conductivity-to-head linearization used in the Monte Carlo density may be an acceptable approximation even at log conductivity variances as large as 5.0 (because the Monte Carlo approach still gives a good estimate of the true travel time). On the other hand, the additional log conductivity-to-travel time linearization used in the Gaussian approach does not appear to be acceptable at such large variances (because the travel time estimate is inaccurate and the shape of the distribution is unrealistic). Both methods and their associated linearizations appear to give good results at log conductivity variances of the order 0.5. These conclusions are consistent with other studies, which have found that head statistics based on a small perturbation assumption remain valid over a wider range of log conductivity variances than velocity statistics derived from the same assumption [1].

Much of the motivation of the multiscale approach is to provide an efficient method for estimating hydrologic variables and for deriving their probability densities. Consequently, we need to examine the computational performance of the algorithm used in our example. As we have seen, the computational advantages of the multiscale approach depend on the sizes of the nodal state vectors. In the two-dimensional example considered here the unaugmented state vectors at Scales 0 and 1 are relatively large (O(100)) compared to the number of fine scale variables (O(1000)). Note that these large state dimensions are needed to properly model the specified two-dimensional Gauss–Markov log conductivity covariance. Relatively speaking, the contribution from augmentation is minor.

For state vectors of this order the multiscale and classical least-squares Gaussian options require comparable amounts of computational effort to derive the travel time density in our example. However, the effort required by the multiscale Monte Carlo option is much less than required with cokriging, the most competitive least-squares conditional simulation alternative. Both the multiscale and cokriging approaches only need to carry out time-consuming covariance calculations once, at the beginning of the Monte Carlo simulation. But individual random realizations can be computed much more quickly with the multiscale approach. This is because the cokriging method needs to generate a new correlated random field of dimension N_f for each realization while the multiscale approach only needs to generate the uncorrelated random variables used in the recursion of (18). The combination of accuracy and efficiency provided by the multiscale Monte Carlo approach make it the best choice for computing travel times, especially for problems where the log conductivity variance is high.

The computational results obtained from our example are somewhat misleading in one important respect. The relatively large computational effort needed to calculate the travel time density with the multiscale Gaussian option is a result of our requirement that the log conductivity be a two-dimensional exponentially correlated Gauss–Markov random field. This requirement reflects our desire to use a commonly accepted model of log conductivity variability, rather than any fundamental property of heterogeneous porous media. The performance advantage offered by the multiscale approach would have been more dramatic if we had modeled the log conductivity field as a $1/f$ process, as was done with sea surface elevation data in [6], precipitation data in [16], and soil moisture data in [10]. In this case $d = 4$ at each node, and the storage requirement and number of floating point operations are both only $O(N_f)$. By comparison, recall that the multiscale Gauss–Markov storage and floating point requirements are $O(N_f^3)$ and $O(N_f^{2.5})$, respectively. Somewhat less dramatic computational savings could be obtained by using approximate two-dimensional Gauss–Markov models such as those described in [5,9,13]. Given the advantages of using $1/f$ models and the uncertainties in log conductivity statistics estimated from limited field data, there may be merit to using $1/f$ rather than Gauss–Markov models in practical applications. This would make it possible to fully realize the computational benefits of the multiscale approach.

5. Conclusions

The multiscale methods described in this paper provide a way to apply the powerful concept of recursion to spatially distributed estimation problems. This capability is particularly useful in hydrology, meteorology, and other earth sciences, where high resolution estimates are required over large regions. Multiscale techniques offer a flexible and very efficient estimation option in such situations. The particular multiscale approach described here is able to incorporate physical laws, such as those which govern groundwater flow and transport, directly into the multiscale model. This greatly expands the range of applications that can be addressed with the multiscale approach.

The methods discussed in this paper require all physical relationships included in the multiscale model to be linear. This important restriction makes it possible to apply classical linear estimation concepts in a multiscale framework. When spatial variability is sufficiently small it may be possible to derive acceptable linear approximations to nonlinear relationships between variables such as log hydraulic conductivity, head, and travel time. These approximations must be used with caution when there is a significant amount of variability. In our travel time example, we observed that a linear approximation of the log conductivity–travel time relationship that works well for a log conductivity variance of 0.5 gives unreasonable results when the variance increased to 5.0. On the other hand, linearization of the log conductivity–head relationship appears to be acceptable even at the higher variance. When a simple linearization is not adequate, multiscale estimation and conditional simulation can be carried out iteratively, with each iteration using updated approximations based on linearizations around the most recent estimate of the model states. A Gauss–Newton version of this iterative approach has been successfully applied to a groundwater inverse problem in [4].

The computational advantages of our multiscale estimation technique depend primarily on the size of the state vectors assigned to the nodes of the multiscale tree. If the state vector dimensionality can be kept independent of the number of finest scale variables to be estimated, the multiscale approach is much more efficient than conventional alternatives such as cokriging. This is the case when the independent variable is a one-dimensional Gauss–Markov process or a multidimensional $1/f$ process. It is usually not the case when the independent variable is multidimensional Gauss–Markov. This is because the state vector of multidimensional Gauss–Markov models must be sufficiently large to provide the partitioning required by the multiscale Markov property. In the two-dimensional case we have examined, this implies that the nodal state dimension must increase as $N_f^{0.5}$, where N_f is the total number of finest scale independent variable values. Since the computational penalty incurred by using a multidimensional Gauss–Markov model is significant it is wise to consider replacing this model with a much more efficient $1/f$ alternative in practical applications, especially if the $1/f$ model appears to be equally compatible with available field data.

Our travel time example indicates that the multiscale conditional simulation (Monte Carlo) technique provides good estimates of the travel time probability density over a wide range of log conductivity variances. Morevoer, this approach is more efficient than competing conditional simulation alternatives, such as cokriging, which need to generate computationally demanding correlated random fields. The computational advantages of the multiscale Monte Carlo approach apply even when the log conductivity is modeled as a two-dimensional Gauss–Markov random field. The multiscale Monte Carlo method for deriving travel time densities appears the best way to simultaneously optimize accuracy and computational efficiency.

There are still many open research questions related to the general field of multiscale estimation and conditional simulation [4]. In particular, there is a need for a more systematic approach to state augmentation, so that the multiscale state vectors can be selected to

maximize computational efficiency. There is a need to develop a wider range of independent variable statistical models which are amenable to the multiscale approach. There is also a need to expand the multiscale framework to include nonlinear and/or time-varying models of relevant physical processes. Finally, we need to investigate ways to apply multiscale estimation methods to the hydrologically important problems of upscaling and downscaling. This may require consideration of a broader class of multiscale models than the internal models examined here. Taken together, these topics constitute a fruitful area for continued research.

Acknowledgements

This research was supported by the Office of Naval Research under grant N00014-91-J-1004, the Air Force Office of Sponsored Research under grant F49620-95-1-0083 and through Boston University under subcontract GC123919NGD, and the National Science Foundation under grant EAR- 9218602.

References

[1] Ababou R. Three dimensional flow through randomly heterogeneous porous media. PhD thesis, MIT, October 1987.
[2] Chou KC. A stochastic modeling approach to multiscale signal processing. PhD thesis, MIT, May 1991.
[3] Chou KC, Willsky AS, Benveniste A. Multiscale recursive estimation, data fusion and regularization. IEEE Trans Automat Contr 1994;39(3):464–78.
[4] Daniel MM. Multi-resolution statistical modeling with application to modeling groundwater flow. PhD thesis, MIT, January 1997.
[5] Daniel MM, Willsky AS. A multiresolution methodology for signal-level fusion and data assimilation with applications to remote sensing. Proc IEEE 1997;85(1):164–8.
[6] Fieguth PW, Karl WC, Willsky AS, Wunsch C. Multiresolution optimal interpolation and statistical analysis of TOPEX/POSEIDON Satellite altimetry. IEEE Trans Geosci Remote Sensing 1995;33(2):280–92.
[7] Gelb A. Applied optimal estimation. New York: Prentice-Hall, 1975.
[8] Irving WW. Multiscale stochastic realization and model identification with applications to large-scale estimation problems. PhD thesis, MIT, August 1995.
[9] Irving WW, Willsky AS. A canonical correlations approach to multiscale stochastic realization. IEEE Trans Image Proc 1996, submitted.
[10] Kumar P. A multiple scale state-space model for characterizing subgrid scale variability of near-surface soil moisture. IEEE Trans Geosci Remote Sensing 1999;37(1):182–97.
[11] Luettgen MR. Image processing with multiscale stochastic models. PhD thesis, MIT, May 1993.
[12] Luettgen MR, Karl WC, Willsky AS. Efficient multiscale regularization with applications to the computation of optical flow. IEEE Trans Image Proc 1994;3(1):41–64.
[13] Luettgen MR, Karl WC, Willsky AS, Tenney RR. Multiscale representations of Markov random fields. IEEE Trans Signal Proc 1993;41(12):3377.
[14] McLaughlin DB, Townley LR. A reassessment of the groundwater inverse problem. Water Resour Res 1996;32(5):1131–61.
[15] Menemenlis D, Fieguth PW, Wunsch C, Willsky AS. A fast optimal interpolation algorithm for mapping hydrographic and other oceanographic data. J Geophys Res 1996, submitted.
[16] Primus I. Scale-recursive estimation of precipitation using remote sensing data. Master's thesis, MIT, June 1996.
[17] Rubin Y. Transport in heterogeneous porous media: prediction and uncertainty. Water Resour Res 1991;27(7):1723–38.
[18] Rubin Y, Dagan G. Conditional estimation of solute travel time in heterogeneous formations: impact of transmissivity measurements. Water Resour Res 1992;28(4):1033–40.
[19] Zimmerman DA, De Marsily G, Gotway CA, Marietta MG, Axness CL, Beauheim RL, Bras RL, Carrera J, Dagan G, Davies PB, Gallegos DP, Galli A, Gomex-Hernandez J, Gringrod P, Gutjahr AL, Kitanidis PK, Lavenue AM, McLaughlin D, Neuman SP, Ramarao BS, Ravenne C, Rubin Y. A comparison of seven geostatistically-based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. Water Resour Res 1998;34(6):1373–413.