# ASSESSMENT AND FORMULATION OF DATA ASSIMILATION TECHNIQUES FOR A 3D RICHARDS EQUATION-BASED HYDROLOGICAL MODEL

**Marino Marrocu and Claudio Paniconi**

*Center for Advanced Studies, Research and Development in Sardinia (CRS4)*
*Cagliari, Italy*

**May, 2001**

# Contents

# 1 Introduction

## 1.1 Overview

Water-related directives emanating from the European Union increasingly emphasize the hydrological catchment as the fundamental organizational unit for integrated planning and management of surface and subsurface freshwater resources. Continued progress in our scientific understanding of hydrological processes at the catchment scale relies on making the best possible use of advanced simulation models and the large amounts of environmental data that are increasingly being made available. Processes at the interface between the land surface and the atmosphere, for instance, determine the partitioning of rainfall into infiltration and runoff and the redistribution of water between the surface, soil, underlying aquifers, and streams. Understanding and predicting these exchanges is important to agriculture (irrigation planning and vegetation and crop growth), climate studies (weather prediction and global change), natural hazards prevention and mitigation (floods, droughts, erosion, landslides), and water quality management (point and nonpoint source pollutants in catchment and stream waters). Demographic and land use changes, in turn, can effect the rainfall-runoff response of a river basin, and these impacts, which can have grave socio-economic consequences, are difficult to assess without reliable data and simulation models.

A wide variety of distributed hydrological models has been developed over the past decades, ranging from simple empirical equations that can be solved analytically to complex systems of partial differential equations that require sophisticated numerical algorithms and powerful computers. The common feature of distributed models is that they can incorporate the spatial distribution of various inputs and boundary conditions, such as topography, vegetation, land use, soil characteristics, rainfall, and evaporation, and produce spatially detailed outputs such as soil moisture fields, water table positions, groundwater fluxes, and surface saturation patterns. A major factor contributing to the popularity of the distributed modeling approach is the availability of digital terrain data, and GIS-based algorithms for extraction of hydrologically relevant information from this data. One of the major problems plaguing distributed modeling is parameter identifiability, owing to a mismatch between model complexity and the level of data which is available to parameterize, initialize, and calibrate models, and to uncertainty and error in both models and observation data. One outcome of this is that most models have not yet been validated in all their detail.

New data sources for observation of hydrological processes however can alleviate some of the problems facing the validation and operational use of hydrological models. In situ or ground-based measurement has become more feasible with the advent of simpler and cheaper sensors, gauges, and loggers, while satellite and airborne remote sensing has begun to fulfill some of its potential for hydrological applications, allowing monitoring and measurement of rainfall, snow, soil moisture, vegetation, surface roughness, and land cover over large areas. In situ and remote measurement techniques are complementary, the one offering high temporal detail and the other fine spatial resolution.

Modeling and observation of soil moisture will be a particular focus of the study undertaken here as part of the EU-financed DAUFIN project (*Data assimilation within a unifying modeling framework for improved river basin water resources management).* The importance of surface soil moisture in hydrometeorology and agriculture is such that the study of its spatial and temporal variability continues to receive a lot of attention. One of the main developments to this end has been the deployment of active and passive microwave remote sensing instruments to measure soil moisture at basin and regional scales, although additional progress is needed before accurate moisture mapping becomes a reality. As this goal is neared, however, the combined use of models and remotely sensed soil moisture data is being proposed to address the important problem of inferring soil moisture information for the deeper layers of the soil profile, beyond the 5-20 centimeters directly detectable by remote sensors.

## 1.2 Objectives of the study

The main objectives of the DAUFIN project are:

- To develop a unifying modeling framework applicable at the catchment scale and based on rigorous conservation equations for the study of hydrological processes in the stream channel, land surface, soil, and groundwater components of a river basin;
- To implement data assimilation methodologies within this modeling framework and for other control models to enable the optimal use of remote sensing, ground-based, and simulation data;
- To test and apply the models and the data assimilation methods at various catchment scales, including hillslopes and subcatchments of the of the Ourthe watershed in Belgium and the entire Meuse river basin, one of the major basins in Europe with a drainage area of 33 000 km$^2$ that comprises the Ourthe.

In general terms, geophysical data assimilation is a quantitative, objective method to infer the state of the earth-atmosphere-ocean system from heterogeneous, irregularly distributed, and temporally inconsistent observational data with differing accuracies [*U. S. National Research Council* 1991]. It represents a formal methodology to integrate these data with simulation models to provide physically consistent estimates of spatially distributed environmental variables, providing at the same time more reliable information about prediction uncertainty in model forecasts. In operational systems where observation data is available on a routine basis at regular intervals, data assimilation is an important tool in assessing data quality, identifying for instance any biases or systematic errors in satellite-based sensors.

In terms of the DAUFIN project, where some first attempts at adopting data assimilation techniques for catchment scale hydrological applications are being undertaken, one aim is to show how these techniques are able to add value with respect to stand-alone data and model predictions, for instance as applied to the problem of soil moisture profile estimation. In particular in our study a physically-based catchment hydrologic model, CATHY [*Bixio et al.* 2000], will form the basis for the formulation and implementation of a simple data assimilation scheme. In the longer term it is hoped that data assimilation will lead to, for this restricted application and in more general cases, improved models and parameterizations (including initial and surface boundary conditions), a more effective framework for hypothesis testing and scenario analyses, better data sampling strategies as the characteristics of different data sources become incorporated into the modeling framework, and ultimately improved predictions and model predictability.

## 1.3 Data assimilation in hydrology

Data assimilation is by now routinely used in research and operational meteorology, although many scientific challenges remain for improving and extending existing methodologies [*U. S. National Research Council* 1991]. More recently, data assimilation is being introduced in the oceanographical and hydrological sciences, owing to the trend towards better and more regular observation of a wide range of parameters of interest to the Earth sciences, beyond those traditionally used in numerical weather prediction, and to the need, in addressing global change and other environmental problems, for both longer-range and more local forecasting. This need arises where ocean – land surface – atmosphere exchange processes play an important role and where there are inadequacies in the simple spatial integration or upscaling methods currently used to derive, from hydrological models representative of processes at the small scale of a vertical soil column, field plot, or small watershed, land surface parameterizations for climate models operating at scales of several hundred kilometers.

Although there are many spatially distributed models in hydrology which could provide a basis for data assimilation [*McLaughlin* 1995], the most advanced domain of application, to the extent that global model-assimilated datasets are currently being generated, is atmospheric hydrology, due to the important role this "fast component" of the global hydrological cycle plays in weather forecasting. In other domains, early applications of data assimilation in hydrology, reviewed in a later section, have concentrated on how to incorporate into models information from some of the new remote sensors, often based on synthetic experiments, and on assessing available data assimilation methodologies. It is still premature to see "results" (in terms of models being validated or observations being corrected) out of current implementations of data assimilation in hydrological models.

One aspect that may hinder initial attempts at producing such results is incomplete knowledge of the spatial and temporal variability of important hydrological processes and state variables such as soil moisture, rainfall, evapotranspiration, and hydraulic conductivity. Adequate characterization of this variability is needed in the measurement equations or interpolation and extrapolation formulas used in many data assimilation schemes. For soil moisture, for example, widely varying correlation lengths have been reported in the literature, ranging from 1 to 1000 m, and it is questionable whether even at the hillslope scale one can speak of a unique and fixed (in space and time) correlation length, due to diurnal and seasonal cycles and topographic and geomorphologic factors that influence the wetness of a soil. On the other hand, with improvements in model and data quality arising from data assimilation and other advances, progress can be expected in turn in our understanding of the dynamics and interactions responsible for the spatial patterns we observe in runoff, water table levels, and other components of the catchment scale water balance and, at larger scales, for the generation and persistence of floods, droughts, etc on seasonal and interannual bases.

## 1.4 Data assimilation in the CATHY model

The CATHY (CATchment HYdrological) model, a coupled overland and subsurface flow model, is physically-based (or process-based) in the sense that its underlying equations are derived from first principles and represent as complete a description as possible of the underlying physics of water flow, within the limits of the processes and observations of interest. The main processes not accounted for in the current version of the model are preferential (macropore) flow, a separate air phase in the unsaturated zone, hysteresis in the soil hydraulic properties, and explicit modeling of vegetation (transpiration and root-water uptake). The model simulates the dynamics of catchment flow processes in a consistent manner based on conservation principles, and so is a good candidate for data assimilation. Indeed data assimilation is one means of addressing an oft-cited drawback of distributed, data-intensive models such as CATHY, that of requiring more data than is readily or accurately available. Until this data limitation, related to the high degree of heterogeneity found in catchment properties, is overcome, and the computational costs of running detailed three-dimensional models for large basins become less prohibitive, a model such as CATHY will normally be restricted to hillslopes and small catchments.

Other features that make the model appealing for data assimilation include its coupled nature, handling in a unified manner all the mechanisms of rainfall-runoff partititioning and streamflow generation where other models treat these processes as separate components and in a more ad hoc manner. This applies, for instance, to the distinction between infiltration excess and saturation excess overland flow and to the dynamics of storm vs interstorm catchment response. The model readily produces detailed primary (pressure head) and derived (moisture content, integrated measures of soil water, surface saturations, water table positions, groundwater velocities, surface water fluxes) output fields at selected times that can be used for comparison against, and integration with, observation data in an assimilation context. In addition, hydrograph output from the model (typically at the catchment outlet node) gives the spatially integrated response of the basin to potential and actual atmospheric forcings; this is the time series traditionally used for calibration of hydrologic models. One consideration to bear in mind concerning the CATHY model is possible additional complexities in implementing sophisticated data assimilation algorithms for a three-dimensional finite element model such as this, as opposed to the one-dimensional finite difference implementations most commonly found in the hydrological literature to date (see Section 2.4).

The CATHY model is one of the control models to be used in the DAUFIN project for testing and validating the various algorithms and hypotheses within the unified modeling and data assimilation framework to be developed. Issues of consistency, accuracy, and computational efficiency are especially important given the limited possibility there will be to conduct extensive field tests in the pilot phase of the project. Outputs and implementation details to be intercompared and assessed will concentrate on distributed water table and soil moisture response and the representation and handling of exchange terms (mass fluxes) between soil and aquifer and between subsurface (saturated and unsaturated) and surface (overland and channel flow).

Some fairly simple data assimilation algorithms will be implemented as first trials for the CATHY model, with more advanced methods that allow incorporation of model and data uncertainty in an optimal sense reserved for future research. Given the data scarcity in hydrology described earlier and the flexibility of the CATHY model to generate various output fields, the more immediate interest is for a data assimilation formulation that can systematically combine information from different observation sources, both satellite and ground-based, and is not restricted to a single data source by the measurement error models embedded in more advanced assimilation methodologies. The primary focus, as already mentioned, will be on soil moisture, soil profile estimation, and, given the three-dimensional nature of the model, incorporation of oft-neglected lateral subsurface flow and other effects to yield reliable estimates of four-dimensional soil moisture distributions (in space and time).

# 2 Review of data assimilation techniques and applications

Data assimilation in the geophysical sciences refers to a methodology to estimate in a physically consistent way the state of a given physical system using observations. In general this is accomplished, in a optimal way, using also some prior knowledge of the system such as climatology and error covariance of the model and of the observations. With reference to the aims of the DAUFIN project we are interested in all those methodologies that are able to improve the predictive skill of a hydrological model "including", in the physical state of the simulation model, scattered measures of various kinds (satellite, in situ, indirect measures, etc.). Besides providing a better estimate of the initial condition of the system state, a general procedure of data assimilation should be able to assimilate measurements during model simulation every time new observations are available (four-dimensional data assimilation). In this section we will describe the more commonly used data assimilation techniques starting from the simplest one, and following a brief review of data assimilation in meteorology and hydrology we will describe applications to the problem of the estimation of soil moisture using a catchment scale hydrological model.

## 2.1 Sub-optimal data assimilation methods

### 2.1.1 Direct insertion

If during the model integration some measure or indirect estimate of the state variable $s_o$ is available at position $r_i$, this estimate is simply substituted for the corresponding variable $s$ of the model:

$$s(r_i) = s_o(r_i) \tag{1}$$

Although of trivial implementation this procedure is of little, if any, utility. In fact, due to dynamical inconsistency with the model solution, the introduction of observations into the model generates as a side effect non-physical noise that quickly propagates from the insertion point to the entire integration domain, usually rendering the solutions physically inconsistent and sometimes even producing model instabilities. Another major drawback is that observations must be given at the same location as model nodes so the spreading of information can be achieved only via model physics advection, and therefore slowly.

### 2.1.2 Statistical correction

Since, in general, model estimates are affected by systematic and not negligible errors, prior knowledge of the system statistics obtained by observations can be used to adjust the mean value and standard deviation of the model $(\overline{s}, \sigma)$ to that of the measures $(\overline{s_o}, \sigma_o)$.

First the standard deviation is adjusted re-defining "a posteriori" the model state estimation on all the domain grid points $r_i$:

$$s'(r_i) = \frac{\sigma_o}{\sigma} s(r_i) \tag{2}$$

Mean values of the model estimates can be then adjusted

$$s''(r_i) = s'(r_i) - (\overline{s'} - \overline{s_o}) \tag{3}$$

where $\overline{s'}$ is the mean value of the estimate (2). This technique, although simple to implement, makes use of an important concept common to all the more sophisticated data assimilation procedures, that is, the use of the observation statistics to assimilate data in a more consistent and physical way. On the other hand the method implicitly assumes that the statistics have zero bias and that spatial patterns of model estimates are correct but biased. Also, as with direct insertion, advection of information can be accomplished only via model physics.

### 2.1.3 Statistical and optimal interpolation or analysis

Contemporaneous use of data and model statistics can be accomplished, and problems related to non-correspondence of model and observation points avoided, using the technique known as statistical interpolation (SI).

Suppose we want to assimilate the observations $s_o$ to establish a better initial condition for our model, knowing an initial estimate of the same field $s_b$ (known also as first guess or background field). $s_b$ reflects our prior and imperfect knowledge of the system and depending on the problem at hand, can be identified, for example, with the climatology or with a prior integration of the model. The fundamental idea behind SI is the estimate of the "optimal state" $s_a$ (generally known as "analysis") that is at a minimum distance both from observation and first guess. This can be done in a mathematical way minimizing a quadratic cost function that in matricial form reads as

$$J(s_a) = \frac{1}{2}\left\{(s_a - s_o)^T O^{-1}(s_a - s_o) + (s_a - s_b)^T B^{-1}(s_a - s_o)\right\} \tag{4}$$

where $O$ and $B$ are the covariance matrices of the observation and background error, superscript $T$ denotes transposition, and the $s$ are column vectors. Minimizing the first variation of this functional with respect to the analysis, we obtain

$$0 = \frac{\partial J}{\partial s_a} = O^{-1}(s_a - s_o) + B^{-1}(s_a - s_b) \tag{5}$$

In explicit matrix form, solutions of the previous problem are of the form

$$s_a(r_i) = s_o(r_i) + \sum_{k=1}^{L} W_{ik}[s_o(r_k) - s_b(r_k)] \tag{6}$$

where $L$ is the number of measures and the weight matrix $W$ is obtained by solving the system of linear equations

$$W_i = B_i(B + O)^{-1} \tag{7}$$

where $W_i$ is the $i$th column of the matrix $W$.

If the estimates of covariance matrices $B$ and $O$ are exact then the interpolation is said to be optimal, meaning that in this case the variance of the analysis is really a minimum. More reasonably we can think that only estimates of $B$ and $O$ are known and for this reason this technique is known as statistical interpolation.

The minimization of the functional (4) has been performed here taking implicitly the number and position of measures equal to that of the grid model nodes. If this is not the case equation (6) remains formally identical but in this case $B_i$ are the column vectors containing the covariance between the background error at the $i$th analysis grid point and the background error estimate at every observation point.

Observations information is in this case propagated (advected) implicity through the covariances matrices $B, O$ and benefits of the inizialitions, also where observations are not available, can be obtained from the beginning of the model integration.

### 2.1.4 Newtonian relaxation or nudging

The procedures of assimilation considered to this point do not have any explicit time dependence (for the SI procedure however if the background state is derived by a model simulation this time dependence exists implicitly). Depending on the problem at hand, the analysis of the system obtained via these procedures is not suitable to be used as initial conditions of the model (see for example the discussion on meteorology in Section 2.3). In these cases space and time dependence must be considered together and explicitly in the assimilation process (four-dimensional data assimilation).

Nudging is the simplest 4D data assimilation procedure. In this procedure model variables are driven toward observations adding to the forcing $F$ of the model equation

$$\frac{\partial s}{\partial t} = F(s) \tag{8}$$

an additional term with the aim to relax the actual model state to the observed one. This relaxation term is taken to be active for a certain period of time $t_a$, called assimilation time, preceding the observation time $t_o$. After observation time is reached this term is relaxed (made zero) and the model equation is integrated in its original form[1]. The main objective is to obtain, at least partially, a dynamical consistency between measured data and numerical solution of the model. This should avoid introduction of non-physical noise, typical of direct insertion, and assure at the same time an improvement of the variable estimate.

In practice during the assimilation time the model equation is integrated adding a term proportional to the difference between the model solution and the observed state of the system (the so-called "nudging term") to the physical forcing term $F(s)$. In general we can write

$$\frac{\partial s}{\partial t} = F(s) + GW(r,t)\epsilon(r)(s'_o - s) \tag{9}$$

where $s'_o$ are the observation interpolated to the model grid, $G$ determines the relative strength of the nudging term with respect to the physical forcing term, $W(r,t)$ are weights to be specified (see Section 4.2), and $\epsilon \leq 1$ is a factor measuring the accuracy of the observation that if we assume perfect measures should be taken equal to 1.

To understand some of the features of the nudging technique let us consider a limiting case that can be solved analytically. We assume the physical forcing term of the system equation to be zero, $\epsilon = 1$, and the observed state $s_o$ and the product of the nudging factor with the weighting function to be constant. In this case a general solution of equation (9) is

$$s(t) = s_o + e^{-GWt}(s(t_0) - s_o) \tag{10}$$

with $t_0 = t_o - t_a$. This equation shows that even in the case where the analysis quality factor $\epsilon$ is equal to 1 the "nudged" solution reaches the observed state only if the assimilation time is infinite.

[1] Some authors argue that the assimilation time must be centered with respect to the time of observation, thus retaining the nudging term also after the observation time is reached. In this case $t_a$ can be interpreted as a time of influence or correlation between the observations and the model state.

Since measures are taken at discrete points, to be of practical use the continuum form of the nudging term in equation (9) must be discretized. This can be done in a general way allowing the system equation to be relaxed to individual scattered observations. To do this we first interpolate the model grid values to the observation points $r_i$ (forward interpolation), then backward to the model grid including this last step in the nudging term. A useful form that is generally given to equation (9) is

$$\frac{\partial s}{\partial t} = F(s,r,t) + GW_M(r,t)\frac{\sum_{i=1}^{L} W(r_i,t)\epsilon(r_i)(s'_o(r_i) - s(r_i))}{\sum_{i=1}^{L} W(r_i,t)} \tag{11}$$

where $W_M(r_i,t)$ is the maximum weight for any single observation. Factorization of this term prevents the nudging algorithm from giving too much relative weight to closely located observations (not much new information is really added in this case) with respect to measures that are farther apart. To further reduce this kind of problem backward and forward interpolation required by the algorithm should be performed taking steps similar to those used in the SI procedure for trying to obtain an "optimal a posteriori estimate" of weight coefficients. In operational practice however the weight matrix is generally set a priori. This consideration suggests that particular care must be paid to the choice of weighting function in order for the nudging technique to be of real effectiveness.

## 2.2 Optimal data assimilation methods

### 2.2.1 4D variational data assimilation

The idea beyond these methods is that of minimizing, using the variational method, the difference between a temporal series of objective analysis of measured data and the corresponding time series simulated by the model. The aim is to obtain a state of the system closest as possible to observation but also dynamically consistent with the model equation over the assimilation time. Since the technique is quite general and can be implemented in many ways for the same problem (imposing the model as a strong or weak constraint, for example) we will introduce here the technique in its basic form without reference to any specific implementation. A more general formulation that has been applied to a case study of hydrological problems at large scale can be found in *Reichle et al.* [*2000*].

To preserve generality and to show the applicability of the procedure also to nonlinear problems we will use here explicit functional notation. Suppose $M$ is a differentiable operator (in general nonlinear) relating the state variable $s$ to an observation $o$

$$o(t) = M[s(t)] \tag{12}$$

$M$, known as the measurement prediction operator, will also provide, in general, interpolation from the model grid to that of the measures. Define $R_t$ as the matrix of observation error covariance that can be thought of as the sum of the covariance of instrument errors (matrix $O$ of equation (4)) and the covariance of representiveness error due both to errors in interpolation and errors of the measurement prediction operator itself.

We introduce the following functional sum of a least-squares performance measure plus a term introducing the Lagrange multipliers:

$$I(s) + J(s,\lambda) = \int_0^{t_a} \{[o(t) - M[s(t)]]^T R_t^{-1}[o(t) - M[s(t)]] + 2\lambda(\frac{\partial s}{\partial t} - F)\}dt \tag{13}$$

In this expression the first term is proportional to the difference between the time series of measurements and the prediction of the model during the entire assimilation time $t_a$, normalized with the covariance of observation error, and the second term assures dynamical consistence of the solution.

Minimizing the first variation of this functional we have

$$-\frac{\partial \lambda}{\partial t} = S_t^T \lambda + M_t^T R^{-1}[o - M(s)] \qquad \lambda(t_a), \lambda(0) = 0 \tag{14}$$

$$\frac{\partial s}{\partial t} = F(s) \qquad s(0) = s_a \tag{15}$$

where $S_t = [\frac{\partial F}{\partial s}]$ and $M_t = [\frac{\partial M}{\partial s}]$ are the Jacobian matrices of partial derivatives of the system equation and of the measurement operator. The explicit time index for these linear operators (matrices) is mandatory because they are no longer independent of $s$ and must be re-evaluated after each time step during integration.

The algorithm defined by equations (14)–(15) works as follows: first the system equation is solved starting from one reasonable, but arbitrary, initial condition $s(0)$. The time series $s(t)$ is then used to solve the "adjoint equation" for $\lambda$ using reverse time to obtain $\lambda(0)$ subject to initial value $\lambda(t_a) = 0$. In general, this "first guess" value of $\lambda$ will not satisfy the natural boundary condition $\lambda(0) = 0$. Nevertheless, it can be used iteratively in a descent algorithm to find the initial condition $s(0) = s_a$ that satisfies the natural boundary condition $\lambda(0) = 0$ and minimizes the whole functional (13).

#### 2.2.2 Extended Kalman filtering

With the development in the last two decades of various remote sensing techniques (from satellites but also from ground instruments), increasing amounts of data, inhomogenous in both type and spatio-temporal resolution, are available to be assimilated in models related to enviromental sciences. To meet this challenge new approaches have been devised in which data can be assimilated continuously as they become available, and at any moment of the operational chain an analysis usable as optimal initial conditions for the simulation can be released.

One technique by means of which continuous data assimilation can be implemented in an elegant and powerful manner is the Kalman filter (KF) procedure [*Daley* 1991]. KF has been variously applied to signal processing and other fields and is used for example to determine, and continuously correct, orbital parameters of satellites. To present the methodology as simply as possible we will suppose the model and measurement operators to be linear. It should be stressed however that the procedure is powerful enough to be applied to the nonlinear case making considerations similar to those used in 4D variational assimilation.

One fundamental assumption to derive the KF algorithm is to admit our model to be randomly perturbed. We write then the stochastic version of the linearized system equation discretized in time as:

$$\hat{s}_{t+1} = S_t \hat{s}_t + \epsilon_t^m \tag{16}$$

which evolves the true status $\hat{s}$ of the system at time $t$ to time $t+1$ except for an unbiased and uncorrelated error in time due to model imperfection. Defining the error of the analysis (observation error) $\epsilon_t^a = s_t^a - \hat{s}_t$ and because by definition

$$s_{t+1} = S_t s_t^a \tag{17}$$

we have for the error of the model estimates

$$\epsilon_{t+1} = S_t \epsilon_t^a - \epsilon_t^m. \tag{18}$$

Multiplying on the right this last expression by $\epsilon_{t+1}^T$ and applying the expectation operator we obtain

$$P_{t+1} = S_t P_t^a S_t^T + Q_t \tag{19}$$

where $P_{t+1}$ is the covariance error matrix of the model, $P_t^a$ is the covariance of the analysis error, and $Q_t$ is the covariance matrix of the model error. Equations (17) and (19) are the predictive part of the algorithm. The other three equations needed to close the algorithm are obtained minimizing a cost function such as that used to obtain the SI equations. In particular we can introduce the functional

$$J = \frac{1}{2}\{[o_t - M_t s_t^a]^T R^{-1}[o_t - M_t s_t^a] + [s_t - s_t^a]^T P_t^{-1}[s_t - s_t^a]\} \tag{20}$$

where, as before, $R$ is the error covariance of the measurement operator. Minimizing with respect to the analysis we get

$$s_t^a - s_t = K_t[o - M_t s_t] \tag{21}$$
$$K_t = P_t M_t^T [R_t + M_t P_t M_t^T]^{-1} \tag{22}$$
$$P_t^a = [I - K_t M_t] P_t \tag{23}$$

where $[o_t - M_t s_t]$ is the observational increment or innovation vector, $K_t$ is the gain matrix, $[s_t^a - s_t]$ is the analysis increment, $R_t$ the covariance error matrix of observations, and the other symbols are as defined in a previous section.

KF contains as particular cases direct insertion when only equation (17) is used, $s_t^a$ being in this case the interpolated field, and statistical interpolation when $P$ is estimated a priori in some way and only the three equations (22), (21), and (17) are used. In this last case equation (23) can be used to diagnose the covariance error of the analysis from that of the estimate.

## 2.3 Meteorological perspective

Until the middle of the 20th century the state of the atmosphere was diagnosed using manual and subjective techniques. Drastic changes were introduced with the rapid increase of computational resources and with the complexity and the amount of atmospheric data collected every day all over the globe and from various and inhomogeneous sources (radiosoundings, satellites, radars, in situ measurements, etc). Since then, techniques of data assimilation have had, in meteorology more than in other enviromental sciences, great developments. This is due to the fact that meteorological forecasts are subject to a continuous validation by direct comparison with data from the global meteorological observation network, and by virtue of the fact that such forecasts have a great social and economic impact and thus national authorities have allocated substantial resources aimed at improving the accuracy and skill of meteorological forecasts.

Besides the great spatial and temporal variability of the observational network, to understand data assimilation in meteorology and its peculiarities it is necessary to give some details about the atmospheric system and the equations used to simulate it. Almost all of the numerical weather prediction centres in the world use, to solve the atmospheric problem, the so-called primitive equations which include the two-component $(u, v)$ equations of the horizontal motion, the hydrostatic law for the vertical component, the first law of thermodynamics for the potential temperature $(\theta)$, the gas equation state for density and pressure $(\rho, p)$, and the conservation of mass and moisture equations for specific humidity $(q)$. The solution of these equations are characterized by two different time scales:

$$\tau_1 = f^{-1} \approx 10^4 \text{ s} < 3 \text{ h} \qquad\qquad \tau_2 = \frac{L_H}{V_H} \approx \frac{10^6 m}{10 m/s} = 10^5 \text{ s} > 1 \text{ day} \tag{24}$$

where $f$ is the Coriolis parameter and $V_H$ and $L_H$ are the charateristic horizontal velocity and length scales. Motions of primary importance in meteorology are the midlatitude perturbations that have horizontal scales of order $10^6$ m, last 1 day or more, and are clearly forced by the diurnal cycle (Rossby wave scale). At the spatial scale of $10^4$ m the motion field $(u, v)$ and the mass field $(\rho, q, \theta)$ can be in not perfect balance. If this happens perturbations that last a few hours, and propagate much faster than Rossby waves, can be excited. These motions, with time scale $< O(\tau_1)$, are known as inertial gravity waves. Power spectra analysis for meteorological observation shows that the bulk of the energy in the atmosphere is confined to the Rossby wave scale, indicating that most of the time the atmosphere is in a "balanced state". So, if initial conditions (analysis) used to forecast the future state of the atmosphere is not dynamically consistent with the model equations, spurious gravity waves are excited that can strongly affect the quality of the simulation. To eliminate these non-physical high frequency waves from the model integration two different approaches were proposed by *Charney* [*1947*]:

- Integrate a model that doesn't permit propagation of inertia-gravity waves. This solution is clearly the simplest and far less CPU-intensive, but gives forecasts of limited potential skill.
- Integrate the primitive equations model but using as initial conditions a state that doesn't excite gravity waves.

For the first numerical weather prediction (NWP) models, in the late 40s, the first solution was the only feasible one, and as initialization procedure the optimal analysis (OA) method was used. The old data assimilation cycle could be schematized with the following three components:

- collection and validation of observations;
- short forecast to obtain background or first guess fields;
- analysis (OA).

When computer power increased, in the mid 50s, and operational forecasts based on the primitive equations were made possible, the problem of data assimilation gradually moved to the more complex component cited in the second item above. With the use of the primitive equations another step had to be added, known as initialization. Initialization is a numerical process by means of which high frequency noise in the analysis is filtered out. Various procedures have been used for this purpose.

The simplest initialization procedure consists of disregarding the first few hours (typically 6 hours) of time of forecast, during which non-physical noise is dissipated by the model dynamics itself. This solution is generally adopted by national and regional meteorological services to initialize limited area models (LAMs). To obtain optimal forecasts with general circulation models (GCMs), on the other hand, increasingly complex initialization procedures have been devised, such as QG initialization, linear and nonlinear balance equations, procedures based on the solution of some variational problem subject to various dynamical constraints, and the most widely used, normal model initialization.

The explanation of these procedures is beyond the scope of this brief review although some of these are not very complex to understand. For example in the dynamic initialization procedure the model itself is used for the initialization, executing alternatively forward and backward time stepping with high frequency damping properties. In this way an initial condition free of gravity waves can be obtained that is fully consistent with model dynamics.

All of these procedures, however, suffer in principle a serious limitation because are all based on a static principle. When asynoptic satellite remote sensing data became available to be assimilated in a meteorological model, in the late 60s, a very different approach to the problem of analysis and initialization became possible/necessary: continuos data assimilation. Since then, assimilation methods based on 3D or hopefully 4D variational procedures are used or are going to be used. The ideal technique for meteorological applications would be the Kalman filter procedure but the great amount of computational effort needed to multiply and invert matrices in equation (22) makes this technique currently not feasible.

## 2.4 Recent applications of data assimilation to soil moisture and catchment hydrology

Although relatively new to hydrology, a handful of papers related to data assimilation go back as far as 20 years. *Newton et al.* [*1983*] combined a hydrologic model and passive microwave soil moisture observations to derive deeper (root-zone) soil moisture estimates from near-surface measurements but were only partially successful due to complications from temperature (diurnal variations), surface roughness, vegetation, and soil texture on the signal response. Active microwave soil moisture observations from a C band scatterometer were used by *Prevot et al.* [*1984*] in another 1D model application, without data assimilation, using observed surface soil moisture data as an upper boundary condition for a Richards equation soil water transfer model to improve the accuracy of simulated actual evaporation rates. *Bruckler and Witono* [*1989*] later applied similar techniques to soil water balance estimation.

*Milly* [*1986*] and *Milly and Kabala* [*1986*] introduced the extended Kalman filter, and hence a data assimilation context, to some of these earlier efforts to integrate remote sensing data and hydrologic models for soil moisture profile estimation. *Entekhabi et al.* [*1994*] took this approach further, integrating passive microwave and infrared emitted radiation observations into a coupled heat transport and moisture flow model based on Kalman filter data assimilation algorithms. A similar approach but with an application to active microwave soil moisture data, and so simpler from a modeling point of view since there is no need for a heat transport model and a radiative transfer inversion scheme, has been recently presented by *Hoeben and Troch* [*2000*].

Away from 1D soil moisture profile estimation, hydrologic applications of 3D and 4D data assimilation at the larger scales of catchments and river basins are only very recently appearing. *Hostetler and Giorgi* [*1993*] and *Georgakakos and Baumer* [*1996*] suggested some of the links with climate modeling and the potential benefits (as well as problems to be overcome) for hydrology. In coupling a regional climate model (RCM) to a lake model and to a streamflow model, using the output from the RCM (surface temperature, evaporation, precipitation) to drive the landscape-scale hydrologic models (LSHMs), *Hostetler and Giorgi* [*1993*] reported that the RCM scale (60 km in this application) is still not fine enough for hydrologic modeling, especially in complex mountainous terrain. This echoes the problems faced with assimilating many types of remote sensing data which are still too coarse-scale for most hydrologic needs. *Georgakakos and Baumer* [*1996*], in an overview of past regional and national U.S. soil moisture measurement campaigns, propose a data assimilation technique for ground-based soil moisture and discharge data together with remotely sensed data for estimation of soil water content aggregated over large areas. They discuss at length aspects concerning the characterization of the spatial and temporal variability of soil water at these scales. The authors suggest that using basin-average soil water, computed by integrating over drainage basins and over depth in several layers, is a viable approach for large scale hydrometeorologic studies. At the smaller SVATS scale (soil-vegetation-atmosphere transfer scheme) *Wigneron et al.* [*1999*] applied statistical interpolation and a radiative transfer model together with coupled moisture and heat diffusion equations to assimilate ground-measured surface soil moisture data in an effort to define the requirements for the eventual use of passive remotely sensed microwave observation data.

The 6 years since the publication of a review paper on nascent developments in hydrologic data assimilation [*McLaughlin* 1995] have seen increasing research activity on various aspects of the topic, investigating both appropriate methodologies and potential applications. *Houser et al.* [*1998*] provide an assessment of several sub-optimal sequential data assimilation algorithms for a distributed conceptual catchment scale water and energy balance model. *Li and Islam* [*1999*] present a simple "hard-update" (direct insertion) method for assimilation of passive microwave data using a 1D 4-layer land surface moisture and heat balance model applied to a $15 \times 15$ km Kansas prairie field site. Perhaps the most ambitious effort so far at implementing data assimilation for hydrologic models is the work of *Reichle et al.* [*2000*] and *Reichle* [*2000*] where a weak constraint variational scheme is used in a Richards equation-based coupled moisture and heat transport model intended for passive microwave remote sensing soil moisture and temperature data. The authors claim that their variational algorithm is more efficient than other optimal data assimilation techniques such as KF.

A number of recent papers pertaining more to inverse problems for model calibration and parameter estimation suggest that there are strong similarities between this problem and that of data assimilation, for affinity of ends (improving the reliability or prediction accuracy of models) as well as means (methodologies that share a number of traits). *Cahill et al.* [*1999*] address the upscaling and inverse problem of estimating a large scale hydraulic conductivity function from measurements obtained on a small scale, applying an extended Kalman filter to a 1D Richards equation. In particular the authors explore issues related to estimation of the covariance matrix, which is "the most difficult parameter of the KF algorithm to supply", and other fine points concerning practical implementation and performance of this algorithm. *Senarath et al.* [*2000*] move to a larger scale watershed application and employ the shuffled complex evolution method for calibration on a "continuous basis" of a simple hydrologic model using soil moisture data.

# 3 The CATHY model

## 3.1 General description

Precipitation fluxes during storm events and potential evapotranspiration during interstorm periods are the driving forces of catchment dynamics. The catchment partitions this atmospheric forcing into surface runoff, groundwater flow, actual evapotranspiration, and changes in storage. Surface runoff involves different phenomena such as hillslope and channel flow and retardation and storage effects due to pools and lakes, while groundwater flow includes infiltration to and exfiltration from the vadose zone. The CATHY model simulates these various processes based on a coupling of the Richards equation for variably saturated porous media and a diffusion wave approximation for surface water dynamics. It combines a three-dimensional finite element subsurface flow module, FLOW3D [*Paniconi and Wood* 1993; *Paniconi and Putti* 1994], with a one-dimensional finite difference surface routing module, SURF_ROUTE [*Orlandini and Rosso* 1996]. Hillslope flow is assumed to concentrate in rills or rivulets, allowing both channel and hillslope flow to be described by a one-dimensional convection-diffusion equation. Retardation and storage effects due to lakes or depressions are also implemented, giving a complete description of the catchment flow dynamics.

Starting from a DEM (digital elevation model) discretization of the catchment surface and a corresponding three-dimensional grid of the underlying aquifer, atmospheric input (precipitation and evaporation data) is partitioned into surface and subsurface components by the FLOW3D module. The overland flux values calculated by FLOW3D at the grid nodes are transferred to the DEM cells and implemented as sink or source terms in the SURF_ROUTE module, which routes this surface water and calculates the resulting ponding head values that are in turn used as boundary conditions in FLOW3D.

The mathematical model is described by the system of partial differential equations [*Bixio et al.* 2000]

$$\sigma(S_w)\frac{\partial \psi}{\partial t} = \nabla \cdot [K_s K_r(S_w)(\nabla\psi + \eta_z)] + q_s(h) \tag{25}$$

$$\frac{\partial Q}{\partial t} + c_k\frac{\partial Q}{\partial s} = D_h\frac{\partial^2 Q}{\partial s^2} + c_k q_L(h,\psi) \tag{26}$$

where $\sigma(S_w) = S_w S_s + \phi\frac{\partial S_w}{\partial \psi}$, $S_w(\psi)$ is water saturation, $S_s$ is the aquifer specific storage coefficient, $\phi$ is porosity, $\psi$ is pressure head, $t$ is time, $\nabla$ is the gradient operator, $K_s$ is the saturated hydraulic conductivity tensor, $K_r(S_w)$ is the relative hydraulic conductivity function, $\eta_z = (0,0,1)^T$, $z$ is the vertical coordinate directed upward, and $q_s$ represents distributed source (positive) or sink (negative) terms (volumetric flow rate per unit volume). The surface water is routed using (26) along each single hillslope or channel link using a one-dimensional coordinate system $s$ defined on the drainage network. In this equation, $Q$ is the discharge along the channel link, $c_k$ is the kinematic wave celerity, $D_h$ is the hydraulic diffusivity, and $q_L$ is the inflow (positive) or outflow (negative) rate from the subsurface into the cell, i.e., the overland flow rate. We note that $q_s$ [$L^3/L^3T$] and $q_L$ [$L^3/LT$] are both functions of the ponding head $h$, and that $h$ can be easily derived from the discharge $Q$ via mass balance calculations.

This system of equations must be solved simultaneously for the unknown vector $(Q,\psi)$ or $(h,\psi)$. Nonlinearities arise in the $S_w(\psi)$ and $K_r(S_w)$ characteristic curves in the Richards equation, in the nonlinear dependence of $q_s$ on the ponding head, and in the nonlinear dependence of $q_L$ on $\psi$.

### 3.2 Subsurface flow module

FLOW3D is a three-dimensional finite element model for flow in variably saturated porous media, applicable to both the unsaturated and saturated zones. It handles temporally and spatially variable boundary conditions, including seepage faces and atmospheric inputs. The soil hydraulic properties are specified by $K_s$ and by families of characteristic (constitutive) relationships $S_w(\psi)$ and $K_r(\psi)$ such as those of *van Genuchten and Nielsen* [*1985*] or *Brooks and Corey* [*1964*]. Those of *van Genuchten and Nielsen* [*1985*] can be written as

$$\begin{array}{rcll} \theta(\psi) & = & \theta_r + (\theta_s - \theta_r)[1+\beta]^{-\gamma} & \psi < 0 \\ \theta(\psi) & = & \theta_s & \psi \geq 0 \end{array} \tag{27}$$

$$\begin{array}{rcll} K_r(\psi) & = & (1+\beta)^{-5\gamma/2}\left[(1+\beta)^{\gamma} - \beta^{\gamma}\right]^2 & \psi < 0 \\ K_r(\psi) & = & 1 & \psi \geq 0 \end{array} \tag{28}$$

where $\theta$ is the volumetric moisture content, $\theta_r$ is the residual moisture content, $\theta_s$ is the saturated moisture content (generally equal to the porosity $\phi$), $\beta = (\psi/\psi_s)^n$, $\psi_s$ is the capillary or air entry pressure head value, $n$ is a constant, and $\gamma = 1 - 1/n$ for $n$ approximately in the range $1.25 < n < 6$. The corresponding general storage term is

$$\sigma = S_w S_s + \phi \frac{dS_w}{d\psi} \tag{29}$$

where $S_w = \theta/\theta_s$ and $S_s$ is the specific storage.

For the treatment of the atmospheric boundary conditions, the input flux values are considered "potential" rainfall or evaporation rates, and the "actual" rates, which depend on the prevailing flux and pressure head values at the surface, are dynamically calculated by the code during the simulation. This automatic "switching" of surface boundary conditions from a specified flux (Neumann) to a constant head (Dirichlet) condition, and vice versa, is implemented to correctly reproduce the physical phenomena occurring at the surface.

For example, in the case of precipitation, if a surface node becomes saturated because of infiltration excess, the fraction of precipitation that does not infiltrate and remains at the surface (ponding head) becomes the overland flow to be routed via the surface module. The boundary conditions in this case switch from Neumann (atmosphere-controlled) to Dirichlet (soil-controlled) type. If precipitation intensity decreases, so that the magnitude of actual (computed) flux across the soil surface exceeds the magnitude of the atmospheric flux, the boundary condition switches back to a Neumann type. If a surface node becomes saturated because of saturation excess (the water table reaches the surface), and there is an upward flux across the soil surface (return flow), the overland flow is calculated as the sum of precipitation and return flux. The entire amount of water that remains at the surface or exfiltrates from the subsurface is then transferred for routing to the DEM-based surface runoff module (see next section), which in turn returns, after surface propagation, the ponding head distribution to FLOW3D. In essence overland flow, defined as the flow rate that is present at the surface and that can be routed via the surface model, is calculated at every time step from the balance between potential and actual fluxes.

## 3.3 Surface routing module

The surface hydrologic response of a catchment is considered as determined by the two processes of hillslope and channel transport, operating across all the hillslopes and stream channels forming a watershed and including storage and retardation effects of pools or lakes and infiltration/evapotranspiration and exfiltration effects from subsurface soils.

### 3.3.1 Hillslope and channel processes

It is assumed that hillslope flow concentrates in rills or rivulets that form because of topographic irregularities or differences in soil erodibility and that deepen and widen during the runoff event as a function of slope, runoff characteristics and soil erodibility. To minimize the computational effort and economize on the number of model parameters, the rill formations are lumped at the DEM elemental scale into a single conceptual channel. Each elemental hillslope rill and network channel is assumed to have bed slope and length that depend on location within the extracted transport network, and a rectangular cross section whose width varies dynamically with discharge according to the scaling properties of stream geometry as described by the "at-a-station" and "downstream" relationships first introduced by *Leopold and Maddock* [*1953*]. The distinction between hillslope and channel flow is based on the "constant critical support area" concept as described by *Montgomery and Foufoula-Georgiou* [*1993*]. Rill flow is assumed to occur for all those cells for which the upstream drainage area A does not exceed the constant threshold value $A^*$, while channel flow is assumed to occur for all those cells for which $A$ equals or exceeds $A^*$.

A routing scheme developed on the basis of the Muskingum-Cunge method with variable parameters is used to describe both hillslope rill and network channel flows, with different distributions of the Gauckler-Strickler roughness coefficients to take into account the different processes that characterize the two physical phenomena [*Orlandini and Rosso* 1998]. The model sequentially routes surface runoff downstream, from the uppermost DEM cell in the basin to the outlet, following the previously determined drainage network. A given grid cell will receive water from its upslope neighbor and discharge it to its downslope neighbor, with the inflow or outflow rate $q_L$ at any catchment cell given by

$$q_L = q\Delta x \Delta y / \Delta s \tag{30}$$

where $q$ is the local contribution to surface runoff, as calculated by FLOW3D, $\Delta x$ and $\Delta y$ are the cell sizes, and $\Delta s$ is the channel length within the cell.

### 3.3.2 Topographic depressions

Isolated topographic depressions ("pits") in the catchment DEM can be attributed to the presence of pools or lakes, or can be interpreted as erroneous or missing data. Depressions cannot be handled by automatic drainage network extraction procedures, and depitting techniques are generally used to modify the elevation values and to regularize the DEM. When depressions play an important role in the formation of surface and subsurface fluxes these procedures introduce inconsistent flow directions and do not correctly reproduce the storage and retardation/attenuation effects of pools and lakes on the catchment response. This typically happens in relatively flat areas where flow patterns are strongly influenced by small slope

| | | | | | |
|---|---|---|---|---|---|
| 12 | 11 | 10 | 11 | 12 | 13 |
| 12 | 12 | 12 | 12 | 12 | 13 |
| 13 | 13 | 13 | 13 | 13 | 14 |
| 14 | 14 | 14 | 14 | 14 | 15 |
| 14 | 13 | 13 | 13 | 14 | 15 |
| 13 | 12 | 12 | 12 | 13 | 14 |
| 13 | 12 | 11 | 12 | 13 | 14 |
| 13 | 12 | 12 | 12 | 13 | 14 |
| 14 | 13 | 13 | 13 | 14 | 15 |
| 14 | 14 | 14 | 14 | 14 | 15 |
| 15 | 15 | 15 | 15 | 15 | 15 |

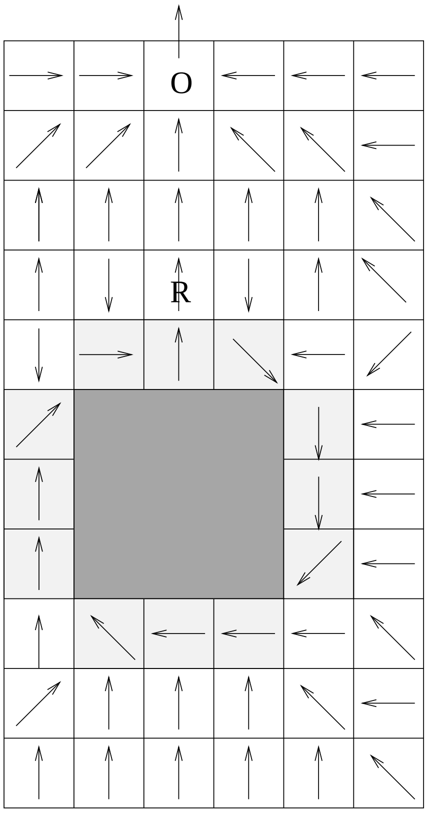


Figure 1: A catchment DEM with elevations (m a.s.l.) (left) and a schematized representation of the catchment with flow paths as calculated by the "depitting" procedure (right). The interior area of the depression is displayed in dark grey and the buffer cells with forced flow directions in light grey. The reservoir cell is identified by the letter "R", while "O" is the outlet cell.

changes. The model therefore incorporates the "lake boundary-following" procedure [*Mackay and Band* 1998] to isolate and correct for potential breakdown in drainage network extraction when natural depressions are present in the DEM.

### 3.4 Surface – subsurface coupling

The explicit in time nature of the Muskingum-Cunge discretization scheme allows the construction of the following noniterative algorithm for the solution of equations (25) and (26):

for $t_k = 0$ to $t_{max}$ with step $\Delta t$ do:

- solve (26) using $q_L^k$ as input to the SURF_ROUTE model, obtaining $Q^{k+1}$ and from this the distribution of ponding heads $h^{k+1}$;
- use $h^{k+1}$ and precipitation/evaporation input at time $t^{k+1}$ to set up boundary and initial conditions for FLOW3D, and solve (25) for $\psi^{k+1}$
- calculate (again with FLOW3D) the overland flux $q_L^{k+1}$ using $\psi^{k+1}$ and the balance between atmospheric inputs and actual fluxes.

The algorithm needs to be initialized, and this is done by setting an initial condition in terms of $q_L$ for equation (26). If this condition is not known a priori, it can be calculated from an initial run of FLOW3D that will evaluate a first guess for the overland flow based on the actual atmospheric input. In this case an initial distribution of $\psi$ needs to be specified.

Different time stepping regimes can be used for the surface and subsurface modules, with SURF_ROUTE normally taking several time steps within each FLOW3D step. This is reasonable physically since the characteristic time scales of surface runoff and channel flow processes are normally much shorter than those for infiltration, redistribution, and groundwater flow. It is also necessary numerically because the explicit time discretization used in SURF_ROUTE requires small time steps for stability.

The "bookkeeping" for the coupled model is done by the FLOW3D module, which determines the current status of each surface node (ponded, saturated, below saturation, air dry) and, knowing for each of these nodes whether the potential atmospheric forcing is positive (rainfall) or negative (evaporation) and, in the case of a Dirichlet boundary condition whether the actual, back-calculated flux represents infiltration or exfiltration and also its magnitude relative to the potential flux, it calculates the overland fluxes to be passed to SURF_ROUTE, partitions the atmospheric and soil surface components of the hydrograph into its various contributions (infiltration, actual evaporation, return flow, direct runoff), and flags any anomalous events (e.g., infiltration at a saturated node with evaporative potential flux, surface runoff at an air dry node). The ability to separate this bookkeeping in the CATHY model is convenient in that it makes it possible to avoid complications due to fact that the FLOW3D module is node-based whereas SURF_ROUTE is cell-based. This structural difference between the two modules can also be exploited to perform simple grid upscaling when warranted by computational or physical considerations. More importantly, for implementation of data assimilation in CATHY, the fact that the FLOW3D module handles the logistics implies that adding the forcing term of the nudging technique (or, in future work, implementing other assimilation algorithms) will not affect the structure of the SURF_ROUTE module. Thus in the next section it suffices to describe in detail only the numerical discretization of the FLOW3D component, which will need to be interfaced with the data assimilation module.

### 3.5 Numerical discretization of the subsurface flow module

We express equation (25) with a generic source/sink term $q$.

$$\sigma \frac{\partial \psi}{\partial t} = \nabla \cdot \left[ K_s K_r \left( \nabla \psi + \eta_z \right) \right] + q \tag{31}$$

Initial conditions and Dirichlet, Neumann, or Cauchy boundary conditions are added to complete the mathematical formulation of the flow problem.

$$\psi(\mathbf{x}, 0) = \psi_o(\mathbf{x}) \tag{32}$$

$$\psi(\mathbf{x}, t) = \psi_p(\mathbf{x}, t) \qquad \text{on } \Gamma_1 \tag{33}$$

$$\mathbf{v} \cdot \mathbf{n} = -q_n(\mathbf{x}, t) \qquad \text{on } \Gamma_2 \tag{34}$$

where $\mathbf{x} = (x, y, z)^T$ is the Cartesian spatial coordinate vector, superscript $T$ is the transpose operator, $\psi_o$ is the pressure head at time 0, $\psi_p$ is the prescribed pressure head (Dirichlet condition) on boundary $\Gamma_1$, $\mathbf{n}$ is the outward normal unit vector, and $q_n$ is the prescribed flux (Neumann condition) across boundary $\Gamma_2$. We use the sign convention of $q_n$ positive for an inward flux and negative for an outward flux, consistent with the convention used for $q_s$ in equation (25).

The finite element solution approximates the exact solution $\psi$ by $\hat{\psi}$ using linear basis functions $w(\mathbf{x})$ defined over a domain $\Omega$ discretized by $E$ tetrahedral elements and $N$ nodes:

$$\psi \approx \hat{\psi} = \sum_{j=1}^{N} \hat{\psi}_j(t) w_j(\mathbf{x}) \tag{35}$$

where $\hat{\psi}_j$ are the components of the nodal solution vector $\hat{\mathbf{\Psi}}$.

Recasting equation (31) in operator notation

$$L(\psi) = \nabla \cdot [K_s K_r (\nabla\psi + \eta_z)] - \sigma\frac{\partial\psi}{\partial t} + q = 0 \tag{36}$$

the error, or residual, represented by the finite element approximation (35) is given as $L(\hat{\psi}) - L(\psi)$, or simply $L(\hat{\psi})$. This error is minimized by imposing an orthogonality constraint between the residual and the basis functions, which yields the Galerkin integral

$$\int_\Omega L(\hat{\psi}) w_i(\mathbf{x}) d\Omega = 0 \qquad i = 1, \ldots, N \tag{37}$$

We assume that the coordinate directions are parallel to the principal directions of hydraulic anisotropy, so that the off-diagonal components of the conductivity tensor $K$ are zero. Expanding equation (37) and applying Green's lemma to the spatial derivative term we get, for $i = 1, \ldots, N$

$$\begin{aligned} -\int_\Omega K_r \left[K_s \left(\nabla\hat{\psi} + \eta_z\right) \cdot \nabla w_i\right] d\Omega + \int_\Gamma K_r \left[K_s \left(\nabla\hat{\psi} + \eta_z\right) \cdot \mathbf{n}\right] w_i d\Gamma \\ -\int_\Omega \sigma\frac{\partial\hat{\psi}}{\partial t} w_i d\Omega + \int_\Omega q w_i d\Omega = 0 \end{aligned} \tag{38}$$

Substituting equation (35), changing sign, and making use of boundary condition (34) to replace the boundary integral term above, we obtain the following system of ordinary differential equations

$$H(\hat{\mathbf{\Psi}})\hat{\mathbf{\Psi}} + P(\hat{\mathbf{\Psi}})\frac{d\hat{\mathbf{\Psi}}}{dt} + \mathbf{q}^*(\hat{\mathbf{\Psi}}) = \mathbf{0} \tag{39}$$

where

$$h_{ij} = \sum_{e=1}^{E} \int_{V^e} K_r^e \left(K_s^e \nabla w_j^e \cdot \nabla w_i^e\right) dV \tag{40}$$

$$p_{ij} = \sum_{e=1}^{E} \int_{V^e} \sigma^e w_j^e w_i^e dV \tag{41}$$

$$q_i^* = \sum_{e=1}^{E} \left[\int_{V^e} K_r^e K_{sz}^e \frac{\partial w_i^e}{\partial z} dV - \int_{V^e} q^e w_i^e dV - \int_{\Gamma_2^e} q_n^e w_i^e d\Gamma\right] \tag{42}$$

In the above equations, $H = \{h_{ij}\}$ is the flow stiffness matrix, $P = \{p_{ij}\}$ is the flow mass (or capacity) matrix, $\mathbf{q}^* = \{q_i^*\}$ accounts for the prescribed boundary flux, the withdrawal or injection rate, and the gravitational gradient term, and $K_{sz}$ is the vertical component of the saturated conductivity tensor. Model parameters that are spatially dependent are considered constant for each element. Parameters that depend on pressure head are evaluated using $\psi$ values averaged over each element and are also elementwise constant. Dirichlet boundary conditions are imposed after the discretized system has been completely assembled.

Equation (39) is integrated in time by the weighted finite difference scheme

$$\left(\nu H^{k+\nu} + \frac{P^{k+\nu}}{\Delta t_k}\right)\hat{\mathbf{\Psi}}^{k+1} = \left(\frac{P^{k+\nu}}{\Delta t_k} - (1-\nu)H^{k+\nu}\right)\hat{\mathbf{\Psi}}^{k} - \mathbf{q}^{*^{k+\nu}} \tag{43}$$

where $k$ and $k+1$ denote the previous and current time levels, $\Delta t_k$ is the time step size, and $H$, $P$, and $\mathbf{q}^*$ are evaluated at pressure head $\hat{\mathbf{\Psi}}^{k+\nu} = \nu\hat{\mathbf{\Psi}}^{k+1} + (1-\nu)\hat{\mathbf{\Psi}}^{k}$. For numerical stability, parameter $\nu$ must satisfy the condition $0.5 \le \nu \le 1$.

Equation (25) is highly nonlinear due to the pressure head dependencies in the storage and conductivity terms, and is linearized in the code using either Picard or Newton iteration [*Paniconi and Putti* 1994]. Expressing the discretized flow equation (43) as

$$\mathbf{g}\left(\hat{\mathbf{\Psi}}^{k+1}\right) = H^{k+\nu}\hat{\mathbf{\Psi}}^{k+\nu} + \frac{1}{\Delta t_k}P^{k+\nu}\left(\hat{\mathbf{\Psi}}^{k+1} - \hat{\mathbf{\Psi}}^{k}\right) + \mathbf{q}^{*^{k+\nu}} = \mathbf{0} \tag{44}$$

Newton's method can be written as

$$J(\hat{\mathbf{\Psi}}^{k+1,m})\mathbf{s}^m = -\mathbf{g}(\hat{\mathbf{\Psi}}^{k+1,m}) \tag{45}$$

where $m$ is the iteration level, $\mathbf{s}^m = \hat{\mathbf{\Psi}}^{k+1,m+1} - \hat{\mathbf{\Psi}}^{k+1,m}$, and the Jacobian matrix is

$$\begin{aligned} J_{ij} = \nu H_{ij} &+ \frac{1}{\Delta t_k}P_{ij} + \sum_s \frac{\partial H_{is}}{\partial \hat{\psi}_j^{k+1}}\hat{\psi}_s^{k+\nu} \\ &+ \frac{1}{\Delta t_k}\sum_s \frac{\partial P_{is}}{\partial \hat{\psi}_j^{k+1}}(\hat{\psi}_s^{k+1} - \hat{\psi}_s^{k}) + \frac{\partial q_i^*}{\partial \hat{\psi}_j^{k+1}} \end{aligned} \tag{46}$$

The Picard method is usually arrived at by evaluating all nonlinear terms in equation (43) at the previous iteration level, $m$, and the linear terms at $m+1$. This yields

$$\begin{aligned} &\left(\nu H^{k+\nu,m} + \frac{1}{\Delta t_k}P^{k+\nu,m}\right)\hat{\mathbf{\Psi}}^{k+1,m+1} \\ &= \left(\frac{1}{\Delta t_k}P^{k+\nu,m} - (1-\nu)H^{k+\nu,m}\right)\hat{\mathbf{\Psi}}^{k} - \mathbf{q}^{*^{k+\nu,m}} \end{aligned} \tag{47}$$

By simple algebraic manipulation, the above equation can be rearranged to give

$$\left(\nu H^{k+\nu,m} + \frac{1}{\Delta t_k}P^{k+\nu,m}\right)\mathbf{s}^m = -\mathbf{g}(\hat{\mathbf{\Psi}}^{k+1,m}) \tag{48}$$

Comparing (45) and (48), it is apparent that the Picard scheme can be viewed as an approximate Newton method. An important difference between the two schemes is that Newton linearization generates a nonsymmetric system matrix, whereas Picard preserves the symmetry of the original discretization of the flow equation. Convergence of an iterative scheme can be enhanced by introducing a relaxation (or damping) parameter $\omega$ of the form $\hat{\mathbf{\Psi}}^{m+1} = \hat{\mathbf{\Psi}}^{m} + \omega\mathbf{s}^m$.

We now write the expressions for the finite element matrices of the flow equation, using linear basis functions on tetrahedral elements. The basis function $w_i^e$ for a generic tetrahedron $e$ with vertices $i$, $j$, $k$, and $m$ is $w_i^e = (\varrho_i + \varsigma_i x + \zeta_i y + \xi_i z)/6V^e$ where the volume of the element is given by

$$V^e = \frac{1}{6}\begin{vmatrix} 1 & x_i & y_i & z_i \\ 1 & x_j & y_j & z_j \\ 1 & x_k & y_k & z_k \\ 1 & x_m & y_m & z_m \end{vmatrix} \tag{49}$$

and

$$
\begin{aligned}
\varrho_i &= \begin{vmatrix} x_j & y_j & z_j \\ x_k & y_k & z_k \\ x_m & y_m & z_m \end{vmatrix} & \varsigma_i = - \begin{vmatrix} 1 & y_j & z_j \\ 1 & y_k & z_k \\ 1 & y_m & z_m \end{vmatrix} \\
\zeta_i &= \begin{vmatrix} 1 & x_j & z_j \\ 1 & x_k & z_k \\ 1 & x_m & z_m \end{vmatrix} & \xi_i = - \begin{vmatrix} 1 & x_j & y_j \\ 1 & x_k & y_k \\ 1 & x_m & y_m \end{vmatrix}
\end{aligned} \tag{50}
$$

We evaluate matrices $H$ and $P$ and vector $\mathbf{q}^*$ of equations (40)–(42). The $ij$th element of matrix $H$ is given by $h_{ij} = \sum_{e=1}^{E} h_{ij}^e$ where

$$
\begin{aligned}
h_{ij}^e &= \int_{V^e} K_r^e \left( K_s^e \nabla w_j^e \cdot \nabla w_i^e \right) dV \\
&= \int_{V^e} K_r^e \left( K_{sx}^e \frac{\varsigma_j}{6V^e} \frac{\varsigma_i}{6V^e} + K_{sy}^e \frac{\zeta_j}{6V^e} \frac{\zeta_i}{6V^e} + K_{sz}^e \frac{\xi_j}{6V^e} \frac{\xi_i}{6V^e} \right) dV \\
&= \frac{K_r^e}{36 \mid V^e \mid} (K_{sx}^e \varsigma_j \varsigma_i + K_{sy}^e \zeta_j \zeta_i + K_{sz}^e \xi_j \xi_i)
\end{aligned} \tag{51}
$$

where $K_{sx}$, $K_{sy}$, and $K_{sz}$ are the diagonal components of the saturated conductivity tensor, and the superscript $e$ indicates that the quantity is averaged over the element. The nonlinear coefficient $K_r^e$ is evaluated using the average value of $\hat{\mathbf{\Psi}}$ at the centroid of each tetrahedron. This treatment of nonlinear coefficients is also used for all the other integral terms.

The $ij$th element of $P$ is $p_{ij} = \sum_{e=1}^{E} p_{ij}^e$ where

$$
p_{ij}^e = \int_{V^e} \sigma^e w_j^e w_i^e dV = \sigma^e \frac{\mid V^e \mid}{20} \cdot \begin{cases} 2 & \text{if } i = j \\ 1 & \text{if } i \neq j \end{cases} \tag{52}
$$

Finally, $\mathbf{q}^* = \mathbf{g}_z + \mathbf{b}_f + \mathbf{q}_f$ where the gravity vector $\mathbf{g}_\ell$, $\ell = x, y, z$ is such that $\mathbf{g}_x = \mathbf{g}_y = \mathbf{0}$ and $\mathbf{g}_z = \{g_{z_i}\} = \sum_{e=1}^{E} G_i^e$, $\mathbf{b}_f = \{b_{f_i}\} = \sum_{e=1}^{E} F_i^e$, $\mathbf{q}_f = \{q_{f_i}\} = \sum_{e=1}^{E} L_i^e$. The components of these three vectors are given by

$$
\begin{aligned}
G_i^e &= \int_{V^e} K_r^e K_{sz}^e \frac{\partial w_i^e}{\partial z} dV = \frac{\mid V^e \mid}{6V^e} K_r^e K_{sz}^e \xi_i & (53) \\
F_i^e &= \int_{V^e} q^e w_i^e dV = q^e \frac{\mid V \mid}{4} & (54) \\
L_i^e &= -\int_{\Gamma_2^e} q_n^e w_i^e d\Gamma = -q_n^e \frac{\mid \Delta^e \mid}{3} & (55)
\end{aligned}
$$

The quantity $|\Delta^e|$ denotes the area of the triangular face of the tetrahedron where the boundary condition is imposed.

# 4 Data assimilation for the CATHY model

## 4.1 Selection of a data assimilation technique

Selection of an optimal data assimilation technique for a full 3D model such as CATHY must be done taking into account various and in general conflicting aspects. On the one hand it would be desirable to use the more general and performing methods of 4D data assimilation, while on the other hand numerical complexity and heavy computational burdens to some extent prohibit implementation of these techniques, especially when using higher resolution data sets. It is worth noting that even with currently available computational power, implementation of a full 4D Kalman filter procedure has not yet been achieved in operational meteorological models, a field where many advances have been made since the 1950s in the field of data assimilation. Even if implementation of an assimilation procedure based on the KF method were feasible, serious operational limitations would arise for a catchment scale hydrological model, since the only observations having the necessary spatial resolution for assimilation are remote sensing data from radar instruments such as SAR. Use of realistic observational errors corresponding to current SAR configurations, however, have shown that no benefits could be expected applying a KF-based procedure to the retrieval of vertical soil moisture profiles in a 1D formulation of the Richards equation model [*Hoeben and Troch* 2000].

Although this particular problem could be partially overcome by reducing spatial resolution, and consequently observation noise, the problem of inadequate sampling frequency remains unsolved. A reasonable sampling frequency for a soil moisture assimilation would be daily satellite images, or weekly images for drier climates. This level of sampling has yet to be applied in hydrology, where at best tandem or single-pass monthly SAR data has been used.

From a comparative and extensive analysis between various data assimilation procedures applied to a simplified hydrological model at catchment scale *Houser et al.* [*1998*] have concluded that as the complexity of the data assimilation model increases, the assimilated data sets must necessarily be reduced to maintain computational feasibility. In practice less sophisticated data assimilation methods (such as statistical interpolation) applied to dense data sets can extract similar information.

In consideration of these aspects we selected the nudging technique as a first implementation of data assimilation for the CATHY model, since it possesses some of the advantages of 4D data assimilation while remaining relatively simple to implement numerically.

## 4.2 Nudging module

Implementation of the nudging term to be added to the tendency of CATHY model equations is quite simple and is made here following equation (11). First the observation points are located within tetrahedral elements. The model estimates are then interpolated to the observation points using the P1 (first-order polynomial) basis functions of the finite element discretization of the CATHY model. This interpolation of model estimates, made within a single elemental unit of the finite element grid, is clearly not influenced by the interpolation properties of the weighting function.

Once $G$, $W$, and $\epsilon$ have been specified the local contribution to the nudging term is evaluated as specified in equation (9) for all the nodes of the mesh. In this way information can be advected between adjacent tetrahedra and within the single tetrahedra by the specified weighting functions.

To be used in a finite element formulation of the problem this term must be evaluated in "weak form". Multiplying it by the test function and integrating within single tetrahedras $t_i$

we obtain the contribution to be added to the model equations:

$$\int_{t_i} N(r,t)w(r)dr = \frac{\sum_{k=1}^{4} N_i(r_k,t)}{4} \tag{56}$$

where $r_k$ are the positions of the nodes of tetrahedra $t_i$.

In the absence of detailed field studies on the spatio-temporal correlation structure of soil moisture, as mentioned previously, we have implemented for the nudging module the time and space weighting functions of *Stauffer and Seaman* [*1990*] as used by *Houser et al.* [*1998*]. It will be important to assess the sensitivity of the model to the weighting functions, and, in conjunction with ongoing and future hydrological field campaigns, encourage the determination of weighting functions of a form that is as consistent as possible with soil moisture dynamics.

We assume that the horizontal, vertical, and temporal dependence of the 4D weighting function can be separated: $W(r,t) = W(x,y)W(z)W(t)$. Defining a horizontal radius of influence $R$, the Cressman-type horizontal weighting function is

$$W(x,y) = \frac{R^2 - D^2}{R^2 + D^2} \qquad 0 \le D \le R \tag{57}$$

$$W(x,y) = 0 \qquad D > R \tag{58}$$

while the vertical function is specified as

$$W(z) = 1 - \frac{|z_o - z|}{R_z} \qquad |z_o - z| \le R_z \tag{59}$$

$$W(z) = 0 \qquad |z_o - z| > R_z \tag{60}$$

where $R_z$ is the vertical radius of influence and $z_o$ the vertical coordinate of the observation point. If, for the linear temporal weighting function a centered time of influence is chosen, we have

$$W(t) = 1 \qquad |t - t_o| < \frac{\tau}{2} \tag{61}$$

$$W(t) = \frac{(\tau - |t - t_o|)}{\tau/2} \qquad \frac{\tau}{2} \le |t - t_o| \le \tau \tag{62}$$

$$W(t) = 0 \qquad |t - t_o| > \tau \tag{63}$$

where $\tau$ is the half period of a pre-defined time influence window (i.e., $\tau = t_a/2$ when the assimilation time $t_a$ is centered with respect to the time of observation).

# 5 Future work

The nudging algorithm developed above will be implemented in the next phase of the project, and some first numerical aspects that will be investigated include:

- Whether there are any differences (and pros/cons) of formulating nudging (and the Richards equation itself) in pressure head or moisture content form;

- The handling of the weighting (interpolation) in time and the implications of this for the dynamic time stepping that is an important feature of the model;
- Whether to evaluate the nudging term at time level $k$ or $k+\nu$ (see equation (43)).

Subsequent testing will be devoted to assessing various features of the algorithm and in particular the many factors which can effect the numerical and physical performance of the assimilation model:

- How does the nudging term impact mass conservation?;
- How does the nudging term (in particular the size of the $G$ coefficient?) impact numerical stability?;
- How does this term affect convergence of the linearization (Picard, Newton) scheme? Does it behave as a relaxation term?;
- Is the surface boundary condition "switching" component of the model, already quite sensitive to changes in potential fluxes and surface saturation events, made even more so by inclusion of data assimilation?;
- Does assimilation of observation data such as surface soil moisture make it necessary to alter in some way the handling (switching) of boundary conditions based on atmospheric fluxes and surface pressure and ponding heads?;
- What is the influence of the soil hydraulic properties as represented in the $\theta(\psi)$ and $K_r(\psi)$ families of curves on the performance of data assimilation?;
- Are special considerations needed for the coupling between the surface and subsurface modules?;
- CPU aspects (especially with a view to implementation of more sophisticated or optimal data assimilation schemes).

Ultimately these tests will provide some feeling for how to proceed with extensions of the nudging scheme or with more sophisticated data assimilation algorithms for the CATHY model, such as Kalman filtering or a variational method. A simple approach might be to take an existing one-dimensional (vertical) implementation of Kalman filtering, such as the one presented by *Hoeben and Troch* [*2000*] based on a finite difference flow model, and apply it in an "offline" mode at each time and spatial location where a new observation is available. The set of "one-dimensional profiles" thus generated can then be spatially integrated or statistically interpolated to generate an updated three-dimensional field for the CATHY model. Additional issues relevant to future formulations and implementations for the CATHY model include:

- The differences, if any, in assimilating observation data at points corresponding to surface vs internal observation nodes of the three-dimensional model domain;
- Similarly, differences in assimilation of remote sensing vs ground data (including considerations regarding regularly-spaced vs irregular data);

- For assimilation of soil moisture observation data, reliable estimation or mapping of variables beyond soil moisture and water table distributions, such as surface saturation, variable source areas (partial contributing areas), discharge and recharge zones, soil parameters related to texture and permeability, and detection of conditions controlling the switching from atmosphere-controlled (stage-one) to soil-limited (stage-two) evaporation [*Salvucci* 1997];

- For soil moisture from active microwave synthetic aperture radar (SAR) satellites, the possibility of using raw backscattering values directly rather than retrieved soil moisture values in the data assimilation algorithms. That is, avoid the costly and complex soil moisture "inversion" step in the SAR data processing of the *observation* data by using instead the "forward" mode of the soil moisture–backscatter model on the CATHY-*simulated* soil moisture values instead;

- More generally, should hydrologic models evolve so that they are formulated using parameters that can be directly measured or observed from remote sensing instruments?;

- Assimilation of observed state variables other than soil moisture.

As a means of conditioning model performance on available data, data assimilation has links to methodologies currently being advanced for addressing problems of model calibration and parameter estimation (or more generally inverse problems), and these links will also need to be explored in more detail. One notable difference between inverse and assimilation problems is that the former are generally used to provide an estimation of lumped or "effective" model parameters, whereas data assimilation provides a spatial distribution of a state variable. But this and other distinctions are becoming blurrier as more sophisticated, "automatic" calibration algorithms are being introduced based on optimal control theory and estimation and variational principles similar to those used in data assimilation techniques.

Future developments on the model front includes continued evolution of the CATHY model to handle processes not yet accounted for (preferential flow, hysteresis, two-phase flow, vegetation, energy balance) as well as improved outputs, more accurate calculation of groundwater velocities, and faster execution (including parallelization of the model based on subcatchment or hillslope partitioning of a catchment).

**Acknowledgements** This work has been funded by the Energy, Environment, and Sustainable Development Programme of the European Commission (contract EVK1-CT-1999-00022) and by the Sardinia Regional Authorities.

## References

Bixio, A. C., S. Orlandini, C. Paniconi and M. Putti, Physically-based distributed model for coupled surface runoff and subsurface flow simulation at the catchment scale. In: Bentley, L. R. (ed.) *Computational Methods in Water Resources, Vol. 2.* Balkema, Rotterdam, The Netherlands, pp 1115–1122, 2000.

Brooks, R. H. and A. T. Corey, Hydraulic properties of porous media. Hydrology Paper 3, Colorado State University, Fort Collins, CO, 1964.

Bruckler, L. and H. Witono, Use of remotely sensed soil moisture content as boundary conditions in soil-atmosphere water transport modeling, 2, Estimating soil water balance, *Water Resour. Res.* 25(12), 2437–2447, 1989.

Cahill, A. T., F. Ungaro, M. B. Parlange, M. Mata and D. R. Nielsen, Combined spatial and Kalman filter estimation of optimal soil hydraulic properties, *Water Resour. Res.* 35(4), 1079–1088, 1999.

Charney, J. G., Dynamics of long waves in a baroclinic westerly current, *J. Meteor.* 4, 135–162, 1947.

Daley, R., *Atmospheric Data Analysis.* Cambridge University Press, Cambridge, UK, 1991.

Entekhabi, D., H. Nakamura and E. G. Njoku, Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations, *IEEE Trans. Geosci. Remote Sensing* 32(2), 438–447, 1994.

Georgakakos, K. P. and O. W. Baumer, Measurement and utilization of on-site soil moisture data, *J. Hydrol.* 184, 131–152, 1996.

Hoeben, R. and P. A. Troch, Assimilation of active microwave observation data for soil moisture profile estimation, *Water Resour. Res.* 36(10), 2805–2819, 2000.

Hostetler, S. W. and F. Giorgi, Use of output from high-resolution atmospheric models in landscape-scale hydrologic models: An assessment, *Water Resour. Res.* 29(6), 1685–1695, 1993.

Houser, P. R., W. J. Shuttleworth, J. S. Famiglietti, H. V. Gupta, K. H. Syed and D. C. Goodrich, Integration of soil moisture remote sensing and hydrologic modeling using data assimilation, *Water Resour. Res.* 34(12), 3405–3420, 1998.

Leopold, L. B. and T. Maddock Jr., The hydraulic geometry of stream channels and some physiographic implications. Prof. Pap. 252, U.S. Geol. Surv., Washington, DC, 1953.

Li, J. and S. Islam, On the estimation of soil moisture profile and surface fluxes partitioning from sequential assimilation of surface layer soil moisture, *J. Hydrol.* 220, 86–103, 1999.

Mackay, D. S. and L. E. Band, Extraction and representation of nested catchment areas from digital elevation models in lake-dominated topography, *Water Resour. Res.* 34(4), 897–901, 1998.

McLaughlin, D., Recent developments in hydrologic data assimilation. In: Pielke Sr., R. A. and R. M. Vogel (eds.) *U.S. National Report to International Union of Geodesy and Geophysics 1991–1994: Contributions in Hydrology* 977–984, American Geophysical Union, Washington, DC, 1995.

Milly, P. C. D., Integrated remote sensing modeling of soil moisture: sampling frequency, response time, and accuracy of estimates. In: *Integrated Design of Hydrological Networks (Proceedings of the Budapest Symposium, July 1986).* IAHS Publication No. 158, pp 201–211, 1986.

Milly, P. C. D. and Z. J. Kabala, Integrated modeling and remote sensing of soil moisture. In: *Hydrologic Applications of Space Technology (Proceedings of the Cocoa Beach Workshop, Florida, August 1985).* IAHS Publication No. 160, pp 331–339, 1986.

Montgomery, D. R. and E. Foufoula-Georgiou, Channel network source representation using digital elevation models, *Water Resour. Res.* 29(12), 3925–3934, 1993.

Newton, R. W., J. L. Heilman and C. H. M. V. Bavel, Integrating passive microwave measurements with a soil moisture/heat flow model, *Agricultural Water Management* 7, 379–389, 1983.

Orlandini, S. and R. Rosso, Diffusion wave modeling of distributed catchment dynamics, *J. Hydrol. Engrg., ASCE* 1(3), 103–113, 1996.

Orlandini, S. and R. Rosso, Parameterization of stream channel geometry in the distributed modeling of catchment dynamics, *Water Resour. Res.* 34(8), 1971–1985, 1998.

Paniconi, C. and E. F. Wood, A detailed model for simulation of catchment scale subsurface hydrologic processes, *Water Resour. Res.* 29(6), 1601–1620, 1993.

Paniconi, C. and M. Putti, A comparison of Picard and Newton iteration in the numerical solution of multidimensional variably saturated flow problems, *Water Resour. Res.* 30(12), 3357–3374, 1994.

Prevot, L., R. Bernard, O. Taconet, D. Vidal-Madjar and J. L. Thony, Evaporation from a bare soil evaluated using a soil water transfer model and remotely sensed surface soil moisture data, *Water Resour. Res.* 20(2), 311–316, 1984.

Reichle, R. H., *Variational assimilation of remote sensing data for land surface hydrologic applications.* PhD dissertation, Massachusetts Institute of Technology, Department of Civil and Environmental Engineering, 2000.

Reichle, R. H., D. McLaughlin and D. Entekhabi, Variational data assimilation of soil moisture and temperature from remote sensing observations. In: Stauffer, F. (ed.) *Model-CARE 99 Proceedings of the International Conference on Calibration and Reliability in Groundwater Modeling.* IAHS, Wallingford, UK, pp 353–359, 2000.

Salvucci, G. D., Soil and moisture independent estimation of stage-two evaporation from potential evaporation and albedo or surface temperature, *Water Resour. Res.* 33(1), 111–122, 1997.

Senarath, S. U. S., F. L. Ogden, C. W. Downer and H. O. Sharif, On the calibration and verification of two-dimensional, distributed, Hortonian, continuous watershed models, *Water Resour. Res.* 36(6), 1495–1510, 2000.

Stauffer, D. R. and N. L. Seaman, Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: Experiments with synoptic-scale data, *Monthly Weather Review* 118(6), 1250–1277, 1990.

U. S. National Research Council, *Four-Dimensional Model Assimilation of Data: A Strategy for the Earth System Sciences.* National Academy Press, Washington, DC, 1991.

van Genuchten, M. T. and D. R. Nielsen, On describing and predicting the hydraulic properties of unsaturated soils, *Ann. Geophys.* 3(5), 615–628, 1985.

Wigneron, J.-P., A. Olioso, J.-C. Calvet and P. Bertuzzi, Estimating root zone soil moisture from surface soil moisture data and soil-vegetation-atmosphere transfer modeling, *Water Resour. Res.* 35(12), 3735–3745, 1999.