The continuing evolution of land surface parameterizations

Randal Koster

Hydrological Sciences Branch, Laboratory for Hydrospheric Science, NASA/Goddard
Space Flight Center, Greenbelt, MD, 20771, US
randal.koster@gsfc.nasa.gov

1. Introduction

Land surface models (LSMs) play a critical role in the simulation of climate, for they determine the character of a large fraction of the atmosphere's lower boundary. The LSM partitions the net radiative energy at the land surface into sensible heat, latent heat, and energy storage, and it partitions incident precipitation water into evaporation, runoff, and water storage. Numerous modeling experiments and the existing (though very scant) observational evidence suggest that variations in these partitionings can feed back on the atmospheric processes that induce them. This land-atmosphere feedback can in turn have a significant impact on the generation of continental precipitation. For this and other reasons (including the role of the land surface in converting various atmospheric quantities, such as precipitation, into quantities of perhaps higher societal relevance, such as runoff), many modeling groups are placing a high emphasis on improving the treatment of land surface processes in their models.

LSMs have evolved substantially from the original bucket model of Manabe et al. (1969). This evolution, which is still ongoing, has been documented considerably [e.g., Avissar and Verstraete, 1990; Garratt, 1993; Sellers et al., 1997]. The present paper also takes a look at the evolution of LSMs. The perspective here, though, is different -- the evolution is considered strictly in terms of the "balance" between the formulations of evaporation and runoff processes. The paper will argue that a proper balance is currently missing, largely due to difficulties in treating subgrid variability in soil moisture and its impact on the generation of runoff.

2. Evaporation versus Runoff in Land Surface Models

2.1. "Effective" Evaporation and Runoff Functions

Given the tremendous complexity built into state-of-the-art LSMs, a careful analysis of evaporation and runoff response to meteorological forcing is an enormous undertaking, and a complete explanation of the different behaviors of different LSMs can be prohibitively difficult. To sidestep this difficulty, Koster and Milly [1997] propose the use of simple, empirically derived relationships between an LSM's soil moisture state variable and its simulated evaporation and runoff ratios. These relationships essentially distill the explicitly coded, complex model parameterizations into simple, "implicit" linear equations.

Some examples are provided in Figure 1. Plotted on the x-axis is the average degree of saturation in the Mosaic LSM's [Koster and Suarez, 1992, 1996] soil moisture profile. In the top plot, the y-axis shows the evaporation ratio, i.e., the ratio of evaporation to net radiation, with the latter scaled by the latent heat of vaporization so that the ratio is dimensionless. In the bottom plot, the y-axis shows the runoff ratio, or the ratio of total runoff to precipitation. Each plotted point represents July-averaged data from an individual year of a multi-decade GCM simulation using the Mosaic LSM. A line has been fitted through the points using linear regression.
Clearly, the linear fits are not perfect; significant scatter is seen around the fitted lines. Nevertheless, the lines do show in a gross sense how evaporation and runoff increase with soil moisture. Koster and Milly [1997] fit lines like these (including an additional one for gravitational drainage) for sixteen LSMs as part of a subproject for PILPS (Process for the Intercomparison of Landsurface Parameterization Schemes; see Henderson-Sellers et al. [1993]). A monthly water balance model that explicitly used these fits was able to reproduce the annual mean and seasonal cycle of evaporation and runoff generated by each full LSM in the study. In other words, simple as they are, the fitted lines capture the key relationships between soil moisture and surface fluxes in each LSM.

2.2. Annual Mean Water Balance

Koster and Milly [1997] used the fitted relationships to derive two parameters that together describe the control of land surface parameterization over annual evaporation. The definition of these parameters, $<\beta>$ and $f$, are illustrated in Figure 2. The rectangle shows, for a hypothetical LSM, the variation of evaporation fraction (solid line) and runoff fraction (dotted line) with the degree of saturation in the root zone ($w_r$). Note that the dynamic range of soil moisture is bounded at the low end (at $w_s$) by the evaporation function and at the high end (at $w_r$) by the runoff function; due to the actions of evaporation and runoff, soil moisture in this model can never go beyond these bounds. The parameter $<\beta>$ is defined as the average value of the evaporation fraction within the dynamic range. The parameter $f$ is defined as the fraction of the range over which runoff occurs.

When Koster and Milly [1997] determined the values of $<\beta>$ and $f$ for each LSM participating in the PILPS experiment and plugged these values into the equation shown at the top of the figure, they found strong agreement between the resulting estimates of annual evaporation and the values actually computed by the LSMs. (Note that in the equation, $D$ describes the character of the local climate, and $E_i$ represents the interception loss, which was indeed removed from all evaporation rates in the PILPS study before determining the fitted functions. See Koster and Milly [1997] for details.) Understanding the gross aspects of
an LSM that control the annual water balance thus amounts to understanding what controls \( \langle \beta \rangle \) and \( f \). A study of Figure 2 shows that \( \langle \beta \rangle \) and \( f \) are essentially controlled by the relative positions of the evaporation and runoff functions -- if either function changes its slope or its position relative to the other, both \( \langle \beta \rangle \) and \( f \) change, and the annual evaporation changes accordingly.

\[
\frac{E}{P - E} = \frac{2D \langle \beta \rangle}{1 + 2D \langle \beta \rangle f_R}
\]

where

\[
D = \frac{E_p - E_i}{P - E_i}
\]

"index of dryness"

\( \langle \beta \rangle \) = Average of beta function across soil moisture range

\( f_R \) = Fraction of soil moisture range over which runoff occurs

Figure 2. Summary of results from Koster and Milly [1997] study. The top equation relates the evaporation from the soil to a climatic index (\( D \)) and two parameters related to the structure of the LSM.

This is illustrated in Figure 3. The top left plot shows four possible runoff formulations (dotted lines) in an idealized model in combination with a single evaporation formulation (solid line). The bottom left plot shows the same evaporation formulation in combination with a different set of runoff formulations. The imposed variations in the positions of the runoff lines relative to the evaporation function are, if anything, much smaller than the differences seen amongst the PILPS LSMS [Koster and Milly, 1997]. The evaporation formulation is somewhat more complex than that assumed in Figure 2, but this does not affect the main result: variations in runoff formulation have a distinct impact on annual evaporation rates.

This exercise illustrates quantitatively what should, in fact, be intuitively clear -- a realistic annual evaporation rate depends on both a realistic evaporation formulation and a realistic runoff formulation. Model development should be aimed at a balanced representation of both. Model development focused mostly on evaporation processes (i.e., on effectively giving the solid line in Figure 2 a more realistic slope and position) is inadequate, since model performance will always be limited by inaccuracies in the runoff formulation (i.e., in the slope and position of the dotted line).
2.3. Timescales of Persistence

Another example of the joint roles played by evaporation and runoff formulations is afforded by a recent analysis of soil moisture memory in climate models [Koster and Suarez, in press]. Using the linearizations illustrated in Figure 1, these authors transformed the standard water balance equation into an equation that relates the autocorrelation of soil moisture to four physical controls: one associated with seasonality in the forcing, one associated with evaporation, one associated with runoff, and the last associated with the correlation between the forcing and antecedent soil moisture. Figure 4 shows the derived equation along with a scatter plot that compares the one-month-lagged soil moisture autocorrelation (July 1 to August 1) simulated by the NSIPP climate modeling system to the autocorrelation estimated with the derived equation. Though not perfect, the agreement is clearly strong. On a global basis, the equation explains about 3/4 of the geographical variation in simulated autocorrelation.

A salient feature of the derived equation is that the evaporation and runoff terms have exactly the same form and enter the equation in exactly the same way. The evaporation term depends on the slope of the fitted line for evaporation fraction (Figure 1), the average net radiation, and the water holding capacity of the soil. Similarly, the runoff term depends on the slope of the fitted line for runoff fraction, the average precipitation, and the water holding capacity. The equation itself gives no hint that the evaporation formulation has a stronger influence on soil moisture persistence than the runoff fraction. Indeed, when the evaporation and runoff terms are plotted side by side, the evaporation and runoff formulations are both seen to be important, though usually in different places. The evaporation formulation has a dominant impact on persistence in the western and central United States, for example, whereas runoff formulation is most important in southern Mexico and Central America. Such analysis shows that for the NSIPP GCM system, the evaporation and
runoff formulations are dominant over roughly 2/3 and 1/3 of the global land area, respectively. Thus, when considering persistence (and hydrological behavior in general over seasonal timescales), the runoff formulation cannot be ignored.

\[ p = \frac{\sigma_{w_0}}{\sigma_{w_{n+1}}} \left( 2 - \frac{\sigma_{R_n}}{C_s} + \frac{\sigma_{P_n}}{C_e} + \frac{\text{cov}(w_0, P_n)}{\sigma_{w_0}} \right) \]

Figure 4. Top: Equation relating the autocorrelation of soil moisture \( (\rho) \) between times \( n \) and \( n+1 \) to the standard deviation of soil moisture (\( \sigma_w \)), the slope of the runoff function (\( a_0 \)), the slope of the evaporation function (\( c_e \)), the average net radiation in the time period (\( R_n \)), the average precipitation in the time period (\( P_n \)), the water holding capacity of the soil (\( C_s \)), and the covariance between the initial soil moisture and the forcing term \( F \) (a combination of the radiation and precipitation anomalies during the time period). See Koster and Suarez [in press] for details. Bottom: Scatter plot showing the degree to which the equation reproduces the one-month-lagged autocorrelation of soil moisture simulated by the NSIPP GCM.

3. Evolution of Evaporation Formulations

The discussion above argues that the land surface component of a GCM requires realistic formulations for both evaporation and runoff. This section focuses on how evaporation formulations have evolved from those used in the earliest, simplest models. The discussion is not meant to be comprehensive; only a "broad brush" is provided here, to give the reader a flavor for the priorities of LSM developers.

3.1. Changes in Form

LSM developers have long recognized that moisture availability is the key control on evaporation and its variations. The earliest LSMS simply prescribed soil moisture conditions, imposing, for example, dry conditions in deserts and wet conditions in tropical forests. Because such prescription, however, necessarily precluded important interactions with the atmosphere (such as higher evaporation rates following heavy rainfall), interactive land surface models were introduced. The simplest is the "bucket" model of Manabe et al. [1969], which allows the water level in a soil moisture reservoir to increase during precipitation events and to decrease as the water evaporates. Evaporation efficiency varies with the water level in the reservoir, and as a result, rainy periods do lead to high evaporation rates, and droughts do lead to low rates. The bucket model is still finding use in climate studies [e.g., Milly and Dunne, 1994].
In the mid-1980's, Sellers et al. [1986] and Dickinson et al. [1986] introduced the "SVAT" (soil-vegetation-atmosphere-transfer) model, a fundamentally different type of LSM. The motivation for the SVAT approach is that vegetation has a tremendous impact on the heat and moisture balances of a region. In addition to an improved treatment of radiation and momentum transfer, the SVAT model typically allows stomatal conductance (a plant property describing the ease with which water travels out the leaves) to reduce transpiration rates during times of environmental stress, including water stress. The evaporation formulation in most SVAT schemes can be described with the aid of a "resistance diagram", an example of which is presented in Figure 5. It is analogous to the resistance diagrams of electrical engineering, but with: (a) vapor pressure replacing voltage, (b) plant, soil and air resistances replacing electrical resistance, and (c) evaporation or sensible heat flux replacing current. The resistance diagram of simpler SVAT schemes is equivalent to that of the Penman-Monteith evaporation formulation [Monteith, 1965].

The roster of LSMs participating in PILPS suggests that the basic SVAT framework is currently very popular. Recently, however, various groups have extended the approach to the next level. The physics of photosynthesis is now explicitly included in some models [e.g., Bonan, 1995; Sellers et al., 1996], using parameters that are strongly tied to satellite-based land surface data. The idea is that because plants and trees open their stomata to maximize carbon uptake while minimizing water loss, an accurate representation of the physics of this uptake is needed to ensure realistic transpiration rates. This approach allows the modeling of carbon budgets in addition to energy and water budgets. Such emphasis on the carbon cycle is indeed noted by Sellers et al. [1997] as the next logical step (after the SVAT) in the evolution of LSMs.

Related to carbon cycle modeling are many recent efforts to model interactive vegetation phenology and/or species distribution [e.g., Foley et al., 1996; Dickinson et al., 1998; Cox et al., 2000]. By allowing vegetation to become "leafier" in response to improved climatic conditions, for example, and by letting the leafier vegetation affect albedo and evaporation, such models allow an additional feedback to the atmosphere.
Pielke et al. [1998] note that terrestrial vegetation dynamics can be as important a climate forcing signal as changes in CO$_2$.

3.2. Treatment of Subgrid Variability

As indicated above, much of the evolution of evaporation formulations in past years has focused on improving the one-dimensional structure of the soil-canopy system. Further indications of this include the presence in some LSMs of a number of soil layers, each with a different root density (e.g., [Desborough, 1997]), detailed treatments of the radiation budget within the canopy (e.g., [Sellers, 1985]), the development of multi-layer snowpack thermodynamics models [e.g., Lynch-Stieglitz, 1994], and careful tests of land surface model behavior against point evaporation measurements [e.g., Sellers et al., 1989; Chen et al., 1997].

Treatments of subgrid variability have received relatively less attention. Many models implicitly assume a mixture of vegetation types or at least a mixture of vegetation and bare soil within a grid cell. In the SiB model [Sellers et al., 1986] and many of its derivatives, the “canopy air” covers all types and is well mixed, so that it can serve as the single point of connection between the atmosphere and the underlying land surface. The mosaic approach is a popular alternative (e.g., Avisar and Pielke [1989], Koster and Suarez [1992], Decoursey et al. [1993], Seth et al. [1994]). With the “mosaic” approach, the land surface grid element is separated into subgrid tiles based on surface characteristics, and each tile interacts independently with the GCM atmosphere. Usually, each tile maintains its own set of surface prognostic variables.

Notice that the mosaic approach essentially accounts for subgrid variability by transforming a one-dimensional evaporation calculation at the large scale into several parallel one-dimensional calculations at the smaller scale. They are aided considerably by the recent wealth of satellite-derived surface properties [e.g., Los, 2000] at high spatial resolution. The subgrid areas, however, are typically fixed, and thus the LSMs have difficulty representing the dynamic spatial variation of soil moisture and its impact on evaporation. “Statistical-dynamical” approaches do attempt this; they relate evaporation to the joint probability distribution functions of soil moisture and meteorological forcing [e.g., Entekhabi and Eagleson, 1989; Famiglietti and Wood, 1990; Avisar, 1992]. They come with their own limitations, however, and are not widely used in operational LSMs.

4. Evolution of Runoff Formulations

4.1. Changes in Form

As with evaporation, runoff production in LSMs is typically a function of soil moisture content. In the bucket model described above, runoff is in fact nonexistent until the water prognostic variable reaches its maximum value, associated with what hydrologists call the “field capacity”. More recent runoff formulations allow runoff to take different forms, such as overland flow, drainage out the bottom of the soil column, and lateral flow out the sides of the soil column. With these formulations, runoff may be generated even when the soil column is relatively dry.

The multitude of runoff and soil moisture transport formulations used by various LSMs (highlighted, for example, by Wetzel et al. [1996]) makes generalizing them difficult. One feature, however, seems almost universal – runoff (and associated changes in soil moisture storage) in most LSMs is determined, at least in large part, from calculations of soil moisture transport in the vertical dimension. The soil column is divided into a set of vertically stacked layers, with each layer maintaining its own soil moisture prognostic variable.
The upward or downward flux of water through the soil column (and eventually out of the column) is a function of the amount of water in each vertical layer.

An example transport calculation, along the lines of that used by Sellers et al. [1996] and Dickinson et al. [1986], is shown in Figure 6. The term $\psi$ is the soil water potential, which becomes large and negative as soil moisture decreases. The vertical gradient of $\psi$ appears in the diffusion equation because water tends to flow from wetter to drier layers, whereas the $l$ appears because water also wants to drain downwards. The term $K$ is the hydraulic conductivity, which also varies strongly with soil moisture. Of course, this particular approach isn’t universal; many LSMs, for example, simply apply timescales to the transport of moisture between layers, and some also compute lateral flow out of each soil layer. This lateral flow, however, is not generally keyed to explicitly resolved horizontal moisture gradients.

The vertical transport calculations typically lead to estimates of drainage. In some LSMs, similar diffusion considerations also lead to estimates of overland flow – first, a maximum possible rate of infiltration at the surface is computed from the soil properties and water content, and then this maximum rate is compared to the precipitation rate. If the precipitation rate is larger, the excess is assumed to run off the surface. Some LSMs instead relate the fraction of precipitation diverted into overland flow to simple, empirical functions of soil moisture content in the topmost layer.

4.2. Treatment of Subgrid Variability

Arguably, the “soil layer” framework of most LSMs acts as an impassable barrier to the evolution of runoff formulations. As emphasized in Figure 7, the soil layer approach implicitly assumes uniform soil moisture conditions over areas spanning many thousands of square kilometers. For many LSMs, even the soil moisture in a thin surface layer, say a couple of centimeters thick, is assumed to be uniform over these vast distances. Again, transports between soil layers are computed from one-dimensional equations, typically using hydraulic conductivities and soil moisture potentials derived with point scale equations. The hydraulic conductivity equation shown in Figure 7 (which relates the conductivity to $w$, the degree of saturation) is typical; its key characteristic is its tremendous nonlinearity, since the parameter $b$ may have values of 4 or higher. This nonlinearity makes the application of such an equation to large-scale average soil moisture somewhat ludicrous.
TYPICAL SWAT MODEL.

Local "plot" scale: Soil moisture diffusion equations might make sense.

\[ q = K \frac{dh}{dz} \]
\[ K = K_0 \omega^{2b+3} \]

GCM grid scale: These diffusion equations are inappropriate because soil moisture is never homogeneous over such large areas.

Figure 7. Illustration of how the equations derived for vertical flow at a point are inappropriate to compute the average vertical flow over large areas. The term h is the hydraulic head, the sum of \( \psi \) and z.

The problem, of course, is that soil moisture is highly variable in space. The processes that control soil moisture movement and runoff production in nature are three dimensional and are not amenable to an explicit treatment with a set of vertically-stacked soil layers. Figure 8, for example, shows a very simple schematic of a hillslope with a length scale of tens of meters. Notice that the water table intersects the ground surface above the river. When rain falls on the associated saturated seepage face, water does not infiltrate the soil; it runs directly off. (This runoff mechanism is often termed "Dunne runoff"). Rainwater falling on top of the hill, on the other hand, can infiltrate the soil. Clearly, with a set of vertically layers, a modeler cannot hope to separate these regimes explicitly. The modeler must instead impose a parameterization that somehow tries to relate this subgrid behavior to the average moisture contents held in the layers.

Figure 8 also suggests how "Hortonian runoff", i.e., surface runoff over subsaturated soil resulting from high precipitation intensities, can vary in space. Because the water table is farther beneath the surface at the top of the hill than, say, mid way down the hill, the soil at the top of the hill is drier. Thus, one might expect the infiltration capacity to be higher at the top of the hill, so that the generation of Hortonian runoff there would be less. Again, with a set of vertically-stacked layers, explicitly accounting for these differences is impossible.
Even subterranean baseflow to rivers, which one-dimensional LSMs try to represent with the drainage term, is in fact controlled by the spatial-varying structure of the water table. In general, the physical mechanisms that determine how much of the precipitation is converted to runoff are controlled much more by the spatial variation of soil moisture (as perhaps induced by topography) than by the "mean" soil moisture averaged across a large area. Until LSMs try to account explicitly for this subgrid variability, their ability to produce realistic runoff rates -- in effect, their ability to produce the right position for the runoff line in Figure 1 -- will be profoundly limited.

A few LSMs try to come to grips with this problem. The VIC model [Liang et al., 1994], for example, explicitly represents the spatial variation of infiltration capacity with an empirical function. A few models use a "TOPMODEL" approach [Beven and Kirkby, 1979] to allow topography to control subgrid soil moisture variability [e.g., Famiglietti and Wood, 1994; Stieglitz, 1997; Koster et al., 2000] and its impact on runoff production. Such models, however, are not yet fully mature.

5. Modeling at a Crossroads

The relative degrees to which evaporation and runoff formulations have evolved cannot, of course, be quantified; any pronouncements to that effect are necessarily subjective. With that caveat, it does appear that the complexity incorporated into the evaporation calculation over the years exceeds that incorporated into the runoff calculation. A balance between LSM evaporation and runoff formulations, noted to be critical in section 2 for realistic GCM behavior, seems to be absent. This imbalance stems from the one-dimensional structure of LSMs. The one-dimensional structure ignores critical three-dimensional controls over runoff production but is nevertheless amenable to improvements (stomatal conductance modeling, carbon budgets, dynamic vegetation, etc.) in the evaporation calculation.

Sellers et al. [1997] identify three "generations" of land surface models: (1) the basic model that conserves energy and water (e.g., the bucket model); (2) biophysical models such as SVATs; and (3) models that deal directly with the carbon budget. The importance of modeling carbon budgets is undeniable. Nevertheless, given the arguments above, such a description of land surface model evolution seems overly focused on the "evaporation" aspects of land surface modeling and not focused enough on the aforementioned lack of balance. An alternative definition of a third generation LSM is thus proposed here, one that addresses a critical "weak link" in the simulation of surface energy and water budgets and is thus every bit as defensible.
A third-generation LSM can be alternatively defined as one that treats properly the subgrid variability of soil moisture and its impact on the generation of surface runoff.

Acknowledgments. The author thanks Max Suarez and Chris Milly for many useful discussions.

References


