A REVIEW OF PARAMETRIZATION SCHEMES FOR LAND SURFACE PROCESSES

Pedro Viterbo

European Centre for Medium-Range Weather Forecasts
Shinfield Park, Reading, UK

1. Introduction

The parametrization of land surface processes (LSP) in numerical weather prediction (NWP) or climate general circulation models (GCMs) is important for a number of reasons. First of all, the sensible and latent heat fluxes at the surface are the lower boundary conditions for the enthalpy and moisture equations in the atmosphere. The land surface schemes are also largely responsible for the quality of model produced near surface weather parameters, such as screen level temperature and dew point, and low level cloudiness. Furthermore, the surface conditions need to be such as to provide the adequate feedback mechanisms for the other physical processes in the atmosphere: low level cloudiness influences the surface radiative balance, sensible heat and latent heat fluxes influence the boundary layer exchanges and the intensity of the moist convective processes. Finally, the correct partitioning between sensible and latent heat fluxes determines the soil wetness, which acts as one of the forcings of low frequency atmospheric variability (Delworth and Manabe, 1988, 1989, Milly and Dunne, 1994). The soil layer acts - through its water content - as an integrator or low pass filter of the time series of rainfall. Understanding the soil wetness variability in the seasonal scale (e.g., extended drought periods), may lead to a better knowledge of low frequency atmospheric variability. Time series of soil moisture anomalies are primarily controlled by potential evaporation and the ratio of potential evaporation over precipitation (Milly, 1994).

The role played by continental and oceanic surfaces on the global heat budget of the atmosphere is illustrated by Fig. 1. The mean atmospheric energy budget is shown in Fig. 1a for all ECMWF forecasts verifying in August 1993; the x-axis specifies forecast days, the picture shows results from day 1 to day 10 in the forecast. For the whole atmosphere, the net radiative forcing is negative (the sign is reversed in the picture), implying a cooling of around 110 Wm\(^{-2}\), balanced by warming at the surface (sensible heat flux), and latent heat release corresponding to moist processes in the atmosphere. The surface sensible heat flux provides around 20% of the energy to balance the net radiative cooling. The moist processes (separated in the picture into convective processes and large-scale condensation) show a different level of activity at the beginning of the forecast range, because of initial imbalances in the thermal and humidity atmospheric fields (Arpe, 1991): in contrast, the sensible heat flux is almost constant during the forecast. Figs. 1b and 1c show the surface energy budget for the same period, for all land points and for all sea points, respectively. Net solar radiation is the energy input to the surface, balanced by the sum of thermal radiation, latent and sensible heat fluxes, labelled total output in the panels b and c. While the radiative fluxes are of the same order for land and sea, the partitioning of the available energy at the surface between latent and sensible heat flux is markedly different. The figures indicate a global value for the Bowen ratio (ratio of sensible heat flux over latent heat flux) of order 1 over the continents and 0.1 for the oceans, suggesting different physical mechanisms controlling the exchanges at the surface: i) the oceans have a larger thermal inertia, slower variations of
Fig. 1 Mean atmospheric and surface energy budget for August 1993 ECMWF forecasts: 

(a) Atmospheric energy budget for all land points; 
(b) Surface energy budget for all sea points; 
(c) As (b), for all sea points.
the surface temperature, imbalances are allowed on a longer time scale; ii) the land masses have a faster responsive surface, the net heat flux at the surface (sum of the radiative fluxes, latent and sensible heat flux) is generally small. All the above balances are only exact (no residuals) at a global/annual time scale; at any particular period, and over a particular area, no exact balance is achieved, implying heat transport by the atmospheric circulation. The latent heat flux over land decreases 20% from day 1 to day 10 in the forecast, suggesting a drying of the surface throughout the forecast range, possibly due to either too much radiative solar energy at the surface or too much evaporation in the presence of correct solar fluxes.

Processes at the soil-vegetation-atmosphere interface, and their impact on GCMs are reviewed in Mintz (1984) and Garratt (1993). First studies on the role of soil water (Namias, 1958) led to the development of the so-called "bucket model" for evaporation and computation of surface runoff (Manabe, 1969). With the work of Deardorff (1978), the attention has somehow been switched from the role of soil water as a slow variable in the climatic system, to the contribution of the vegetation to the latent heat flux (evapotranspiration). Many of the schemes used today in GCMs (e.g. Dickinson et al., 1986; Sellers et al., 1986; Abramopoulos et al., 1988; Noilhan and Planton, 1989) mimic the effect of plant physiology in using the amount of photosynthetic active solar radiation to regulate the opening and closing of leaf stomata, thus controlling the flow of water from the soil into the atmosphere and defining the transpiration rate. The concept of stomatal conductance as a product of different stress functions (Jarvis, 1976) is central to all the above models. On the other hand, many GCMs incorporate an interception reservoir, collecting rain and re-evaporating at the potential rate (Rutter et al., 1972).

The key issues in land surface parameterization are: i) the role of vegetation in controlling evapotranspiration and rainfall interception; ii) an adequate description of heat and water transfer in the soil; and iii) for high latitudes and over mountains a correct description of energy/water exchanges for the cryosphere. The direct influence of the land surface on time scales ranging from the diurnal cycle to the seasonal cycle is illustrated by the dual role of evaporation. Evaporation controls the amount of water kept in the soil during Spring, allowing its release during the Summer: in a climate model, systematic over-evaporation during the Spring means dry, warm Summer bias in the lower troposphere. However, for any particular Summer day, a wet surface will tend to evaporate more than a dry surface, and, for that reason, it will in general be cooler (see Sect. 2 on combination equations): for that reason, an over-evaporation for a typical Summer day, generally means a cold, wet bias for the boundary layer. The role of land-surface parameterization in NWP models reflects these two apparently conflicting roles.

Recent reviews of methods for representing LSP in NWP and climate models include Garratt (1993), Bougeault (1991), Blondin (1991), Rowntree (1991), Avisar and Verstraete (1990), Laval (1988), Verstraete and Dickinson (1986). To the above, more general papers, one should add at least Dickinson et al. (1991), and Sellers (1992), for the role of the biosphere in controlling the evapotranspiration, Dickinson (1992), Stricker et al. (1993), and Shuttleworth (1993a), describing the role of the surface in relation to the climate system, and Dooge (1992a, 1992b), for an hydrologist perspective on the subject. In this review, no attempt will be made to give comprehensive descriptions of land surface parameterization schemes: the reader has a large choice of review papers in the literature. Instead, we will focus on a few issues that we consider relevant: the choice reflects work at ECMWF in recent
years, in boundary layer aspects (Betts et al., 1993), and in the surface (Beljaars and Viterbo, 1994; Beljars et al., 1995; Betts et al., 1995; Viterbo and Beljaars, 1995).

Sect. 2 presents some general remarks on the role of the surface, while in Sect. 3 basic concepts on the soil energy and water budget are introduced: the role of vegetation vs. bare ground on defining the evapotranspiration rate is described, the concepts of interception and surface runoff as perceived by large scale atmospheric models are presented, and the Penman-Monteith equation is derived as an example of an useful interpretative tool combining the heat and water budget in the soil. Sect. 4 presents three examples of parametrization schemes, with different complexity level. Sect. 5 discusses ways of representing subgrid scale heterogeneity, and methods for defining initial conditions for the surface variables. Brief remarks on validation and intercomparison are made on Sect. 6, while Sect. 7 deals with processes affecting the cryosphere. The concluding section emphasizes areas of uncertainty, and current needs for data. Throughout the paper the emphasis will be on simplicity and correct physical representation of the processes: In NWP the sensitivity to initial conditions prohibits the use of too complex LSP models, with a large number of surface parameters, while for climate modelling, the complexity of the more advanced LSP (e.g. SSiB, Xue et al., 1993) can be an advantage to handle correctly the atmospheric/surface interactions.

2. General remarks
Fig. 2 shows the size of the moisture reservoirs of the terrestrial atmosphere and the marine atmosphere, the exchanges of moisture between them, and between the atmosphere and the surface below. Surface evaporation over sea is more than 6 times as large as the corresponding value over land. The sea surface evaporates at the potential rate, while over land there are additional mechanisms that reduce the evapotranspiration rate: dryness of the soil or, over vegetated areas, physiological mechanisms that can reduce or shut transpiration from the plant leaves and trunks making the water from the root zone effectively unavailable for the atmosphere above. Precipitation over land is about a quarter of that over sea. Note that precipitation exceeds evaporation over land, while over sea the reverse is true. In order to have a closed budget for the terrestrial atmosphere, advection of moisture across a vertical wall projecting over the continent boundaries has to match the difference precipitation minus evaporation. Advection is roughly half of the water evaporated over land, suggesting an annual recirculation ratio (ratio of the rainfall coming from local evaporation over total rainfall rate) of 67% (71/107). To close the hydrological cycle, the advection has to be matched by the river runoff: the globally averaged influx of fresh water into the ocean, is estimated in this way as $36 \times 10^{15} \text{kgyr}^{-1}$. For continental areas, annual runoff, evaporation and precipitation are approximately in the ratio 1:2:3. The rainfall rate and the size of the terrestrial atmosphere reservoir can be combined to give a time scale $4.5/107 \approx 0.042 \text{years} = 15 \text{days}$: the terrestrial atmosphere reservoir would be emptied by rainfall in 15 days. In a similar way the reservoir would be replenished by surface evapotranspiration in 23 days ($4.5/71 \text{years}$). The time scales associated to marine rainfall are only 7.5 days, and the corresponding value for evaporation is 6.8 days. This suggests: i) a more vigorous hydrologic cycle over the ocean; ii) a land surface control over large time scales (weeks to months), through the evapotranspiration flux of water at the surface. The implication of the above on the extended predictability of the atmosphere due to the exchanges of water with the land surface has been discussed in many papers (see e.g., Namias, 1958, Mintz, 1984, Dumenil and Bengtsson, 1993, Dirmeyer and Shukla, 1993), and were already mentioned in Sect. 1.
Before proceeding any further, in the interest of clarity we will now discuss a few conceptual notions of evaporation and define some basic quantities. Potential evaporation (PE) is the amount of water evaporated per unit area, per unit time from an idealized extensive free water surface under existing atmospheric conditions (Shuttleworth, 1993b). Several problems exist with this concept (see Brutsaert, 1982 for a thorough discussion). First, potential evaporation is often estimated based on measured near surface atmospheric conditions, which correspond to the idealized situation described above only shortly after an episode of precipitation or dew deposition. The amount of water in the soil conditions the actual evaporation, which, for a given incident radiation, determines the actual values of surface air temperature and humidity. If the soil was an idealized extensive free water surface the measured values of air temperature and humidity would change from its actual values. This sort of ambiguity has been the source of confusion in the literature about how to compute PE in atmospheric models: Pan (1990) present a method for its estimation, by use of a modified near surface air temperature, and Milly (1992) shows that this estimate matches early PE definitions of Manabe (1969) and Budyko (1974).
The other problem of potential evaporation is its failure to recognize the role of the surface cover. As detailed in following sections, the surface water vapour flux from a patchy natural covered surface is the sum of different contributions: the lakes and rivers, the bare ground, the wet fraction of the canopy, and plant transpiration from stems, leaves and trunks. When compared to the other contributions, plant transpiration has an additional resistance to the water vapour flux, dependent on physiological conditions, soil state and environmental factors. The maximum possible value of transpiration, so-called *unstressed evaporation* or *potential evapotranspiration* (Penman, 1948) corresponds to a minimum (non-zero) value of this resistance. If, in order to obtain the idealized *free water surface* defining PE, we do not change the roughness length, unstressed evaporation is always smaller than potential evaporation. Following common use in the literature, we will distinguish between *evaporation at the potential rate* (when the canopy is wet) and unstressed evaporation. Notwithstanding the problems referred above for its definition, PE is still a very useful concept as an upper "energy-limit" value of evaporation, for a given incident energy at the surface (Milly, 1993).

A synthetic illustration of the interactions between the land surface and the atmosphere is presented in Fig.3, adapted from Dooge (1992a). The diagram illustrates the behaviour of the soil and the atmosphere within a complete idealized cycle composed of a wet period followed by a dry period. Let us start just after a long episode of rainfall, point A in the picture. The soil water is available in abundance in the root layer (for a precise definition of the concept of field capacity referred in the picture see Sect. 2.3.2 later in the chapter), and its evolution (drying) is going to be determined by evaporation. While the soil has plenty of water, the rate of evaporation is controlled by the atmospheric moisture content in the near-surface: the regime is *controlled by the atmosphere* and the evaporation is at the potential rate. Below a certain level of soil moisture (point B in the picture), physiological mechanisms will limit the supply of water from the root layer into the atmosphere, and evaporation will drop below its maximum value (potential rate). The regime is under a *soil (vegetation) control*. When precipitation starts (point C), it will meet a soil dry enough during the initial stages, so that infiltration (the amount of water that falls as precipitation and is effectively collected by the soil for future use) will equal precipitation. The evolution of water in the soil is once more *atmospheric controlled*, via the rate of precipitation. Beyond a certain value of soil water (point D), the soil does not have the ability to infiltrate all precipitation, some of it goes into runoff. This last phase is again *soil controlled*: the state of the soil determines a rate of infiltration.

Land surface parametrizations have to represent correctly the surface fluxes and the evolution of soil moisture in all four phases of the cycle, and to switch from the atmosphere control into the soil control regime. The evolution of soil moisture will determine when point D will occur, and the evaporation formulation will determine point B. The crucial areas, from the point of view of the atmosphere, are B-C and C-D. During spring and summer (when the atmospheric evaporative demand is very large), the system stays much longer in the states B-D than in the opposite part of the cycle.

Before going into further detail, we will present an early example of awareness of the role of the different mechanisms controlling the surface heat and water budget: Richardson (1922), in his classical book on numerical weather prediction, identified practically all the
relevant subjects in land surface parameterization. First of all, he notes that lower boundary conditions for the atmospheric equations simplify greatly if one makes a forecast for the soil water content. The surface of the earth is separated in sea, bare ground and vegetated part, and he proposes to use the relative fractions of the above within a grid square. The soil heat budget and water budget are to be solved by discretizing the partial differential equations representing the soil heat transfer and soil water transfer. The depths of the 5 soil layers are in a logarithmic distribution, with a total soil depth of 1.5 m, common to soil temperature and soil moisture (in order to take into account easily the energy exchanges into phase changes in the soil water and the energy transported with the water in the soil).

He is probably the first author to write an equation for the water transfer in the unsaturated part of the soil. The idea of generalising Darcy’s law (established in 1856 for the flow of water in a saturated medium) to the flow of water in unsaturated soil, is normally attributed to Richards (1931). In fact, Richardson has done it nine years earlier. He proposes to integrate the resulting partial differential equation with precipitation and evaporation as top boundary conditions, and to specify soil water properties (matric pressure head and hydraulic
conductivity) depending on soil water content. Moreover he includes in the soil water balance an additional term representing the effect of transport of water in vapour phase. For soil heat transfer, the classical Fourier law is applied, with heat conductivity depending on soil moisture. Two additional terms are included: the first for the heat transported with the water, the second for the heat involved in phase changes (liquid-vapour) within the soil. The top boundary condition is the net heat flux at the surface, added to the energy content associated with water falling as precipitation.

For computation of the evaporation from a canopy, he recognizes the physiological control of plants in reducing transpiration below a certain threshold value of soil moisture, made dependent on soil type. The notion of canopy resistance, together with its "electrical" analogy, is properly introduced. An interception reservoir is used, representing the leaves, collecting precipitation and evaporating at the potential rate. The bare ground evaporation depends on the relative humidity in the air pores of the top soil layer.

All the principles currently used by most current surface parameterization schemes have been proposed by Richardson in 1922. The way he infers values for the constants involved is reminiscent of current practice, where one value for a particular site is generalized to the whole globe: a noticeable difference is that he performed field experiments in order to estimate parameters or study phenomena for which there was insufficient evidence, e.g., for canopy interception.

Before closing the section, we will present combination equations for estimating evaporation: a special family of equations obtained by simultaneous solution of the surface energy balance and turbulent transport of heat and water vapour, taking into account the internal plant resistance to transpiration. Penman (1948) was the first to obtain such an equation, for an open-water (or well-watered) surface. The introduction of the effect of vegetation (through the stomatal resistance of a single leaf or a canopy resistance for the effect of the entire canopy) gives rise to what is normally called the Penman-Monteith equation. For a review of combination equations, and the different derived forms of the Penman equation, see Brutsaert (1982), Milly (1991), and Monteith (1980, 1981). In the following, a sketchy derivation of the Penman-Monteith equation will be presented. Combination equations are interesting for a number of reasons. First, they are useful for micrometeorological estimates of evaporation, because they eliminate the surface temperature from the surface energy budget. Secondly, they can be used to estimate regional evaporation on a daily or monthly basis (Priestley and Taylor, 1972). Thirdly, they can be used in NWP models (Pan, 1990) as a basis for the surface parametrization. Finally, they are a powerful interpretative tool for analysing experimental or model results and understanding the physical mechanisms responsible for evaporation.

When storage terms in the vegetation are neglected, energy conservation at the interface soil/vegetation/atmosphere implies

$$R_n + G + LE + H = 0$$  \hspace{1cm} (1)

where $R_n$ is the net radiation at the surface (sum of net shortwave, the downward longwave, and the upward longwave emitted by the surface), $L$ the latent heat of vaporisation (sublimation if the water exists at the surface in the solid phase), and $G, E$ and $H$ are the
ground heat flux, rate of evaporation, and sensible heat flux, respectively. All fluxes are positive downwards and have units of Wm$^{-2}$. $G$ can also be interpreted as minus the rate of heat storage beneath the surface (vegetation plus soil), in which case (1) represents the heat budget of a finite vertical slab of soil+canopy. Eq. (1) states that all significant water in the soil available for evaporation exists in the liquid (solid) state, the water is transferred to the atmosphere in the vapour state at the expense, energetically, of the surface: the surface provides the latent heat, with a corresponding cooling in case of an upwards moisture flux.

Using Monin-Obukhov similarity, the sensible heat flux can be written as

$$H = \rho C_p C_H(z_{oh}, z_{om}, L) U_L (T_L - T_{sk})$$  \hspace{1cm} (2)$$

where $\rho$, $C_p$, $U$, $T$, $z_{oh}$, and $z_{om}$ are, respectively, the air density, specific heat at constant pressure, wind speed, temperature, roughness length for heat and momentum, the subscript $L$ represents an arbitrary level (at height $z_L$) within the surface layer, $T_{sk}$ is the skin or canopy temperature, the temperature at a point in the air immediately adjacent to the canopy (or the soil, in case of bare ground). $C_H$ is the exchange coefficient for heat and $L$ is the Obukhov length; a functional form of $C_H$ can be found, e.g., in Beljaars and Holstlag (1991). Eq. (2) can be rewritten in resistance form,

$$H = \rho C_p \frac{T_L - T_{sk}}{r_a}$$  \hspace{1cm} (3)$$

By comparing (2) and (3), we obtain

$$\frac{1}{r_a} = C_H U_L$$  \hspace{1cm} (4)$$

relating the aerodynamical resistance, $r_a$, to the product of the wind speed and the exchange coefficient for heat.

The evaporation flux, the flux of water between the air inside the stomata, at saturation, and the air in the surface layer can be modelled following (3). We have now conceptually two resistances, one from the air inside the stomata up to the surface of the leaves, $r_s$, and the second one from the surface of the leaves up to level $z_L$, $r_a$

$$E = \rho \frac{q_L - q_{sat}(T_{sk})}{r_s + r_a}$$  \hspace{1cm} (5)$$

where $q_L$ and $q_{sat}$ are the specific humidity at level $z_L$ and the saturation specific humidity, respectively. By assuming the same aerodynamical resistance for heat and water transfer, we are implicitly assuming identical roughness lengths for heat and moisture.

Assuming we can specify $r_a$ - standard turbulence estimates, using Eq. (4) - and $r_s$ - from the knowledge of canopy state and canopy type, and other environmental factors - and have an estimate for $q_L$ - standard SYNOP observation or lowest model level value - Eq. (5) has still one unknown, $T_{sk}$. By performing a Taylor expansion of $[q_{sat}(T_{sk}) - q_{sat}(T_L)]$, $(T_{sk} - T_L)$ can be eliminated by using Eqs. (1) and (3). If we retain the linear term in the expansion (see Milly,
1991, for higher order approximations), use the Clausius-Clapeyron equation for the slope of the saturation specific humidity function, $S = \frac{dq_{sat}}{dT} \bigg|_{T = T_L}$, the Penman-Monteith equation is finally obtained:

$$-LE = \frac{\rho \left[ q_{sat}(T_L) - q_L \right] + S(R_n + G)}{C_T \div \left( \frac{r_s}{r_a} + S \right)}$$

Evaporation is driven by a combination of two terms, the first proportional to the saturation deficit at screen level height (the advection term), and the second proportional to net radiation minus the soil heat storage (the energy term). During day time summer $G$ is small compared to $R_n$ for a vegetated area (typical instantaneous midday values are 80 Wm$^{-2}$ and 600 Wm$^{-2}$, respectively) and can either be neglected or made proportional to $R_n$. In winter, $(R_n + G)$ and $S$ are small and the saturation deficit term dominates. If we also have an unstressed canopy, $r_s$ takes its minimum value, and evaporation depends crucially on the aerodynamical resistance (the combined effect of stability and roughness length). Typical values of aerodynamical resistances are (see Shuttleworth, 1993b) 45 sm$^{-1}$, 18 sm$^{-1}$, and 6.5 sm$^{-1}$ for grassland, agricultural crop and forest, while the minimum (unstressed) stomatal resistance varies from a value of 60 sm$^{-1}$ over grassland to 120 sm$^{-1}$ over forest. The fact that smaller values of $r_s$ are associated with larger values of $r_a$ makes the first term in the denominator of (6) to change significantly between grassland and forest; with the numbers given above, it will vary by a factor of 10.

Using Eq. (6), we can now quantify some of the different idealized evaporation concepts presented earlier in this section. Potential evaporation is obtained when there in a well watered surface or a wet canopy ($r_s = 0$). It corresponds to the maximum possible value of evaporation; note also that $T_L$ has to be replaced by the value that it would assume when the surface is wet (see below). Unstressed evaporation (or potential evapotranspiration) is obtained by replacing $r_s$ by its minimum value: the effect of the canopy even in unstressed conditions can modify the first term in the denominator by a factor of 10, therefore making, e.g., forest unstressed transpiration substantially smaller than potential evaporation.

Pan (1990) used Eq. (6) as basis for the parametrization of evaporation in the NMC global model. Unstressed evaporation is computed, using a modified temperature $T'_L$, characteristic of a well wet surface; note that $T'_L$ changes the value of the advection term and the radiative term (through the thermal radiation emitted by the surface). A practical way of computing the effects of $T'_L$ is based on Taylor expansions of both terms. Actual evaporation is computed as the product of an evaporative fraction, $\beta$, times unstressed evaporation.

To summarize, the Penman-Monteith equation stresses that: i) evaporation is driven by two large environmental factors, an energy capping given by the net radiation, and an "advection" factor given by the saturation deficit term; ii) the rate of evaporation is controlled by the nature and state of the surface cover, given by the ratio $r_s/r_a$, the presence of $r_s$ indicating the physiological control by the plants.
3. Soil energy and water budget

3.1 SOIL ENERGY TRANSFER

Neglecting the coupling of water transfer in the soil with the heat transfer, we can assume the following Fourier law of diffusion to govern the soil heat transfer

\[
(rC)_s \frac{\partial T}{\partial t} = \frac{\partial}{\partial z} [\lambda_T \frac{\partial T}{\partial z}]
\]

where \((rC)_s\) is the volumetric soil heat capacity \((\text{Jm}^3\text{K}^{-1})\), \(T\) is the soil temperature \((\text{K})\), \(z\) is the vertical coordinate - distance from the surface \((\text{m})\), positive downwards, \(t\) is time \((\text{s})\), and \(\lambda_T\) is the thermal conductivity \((\text{Wm}^{-1}\text{K}^{-1})\). In Eq. (9), the heat transfer is assumed to occur only in the vertical direction.

The volumetric heat capacity and the thermal conductivity depend on the soil type and its water content, and therefore we will shortly review here general properties of soils. For more details on descriptive information on soil science see, e.g., Hillel (1982), Duchaufour (1984), Marshall and Holmes (1988). The soil is a three phase heterogeneous system, where the solid phase is called the soil matrix, the liquid phase is the soil water and the gaseous phase is the moist air trapped in its pores (Hillel, 1982). The soil phase includes the mineral matter and, in some cases, organic matter attached to the mineral grain and binding them together. The fraction of soil occupied by the soil pores is called porosity, or volume of air trapped over total volume, \(\theta_s\) \((\text{m}^3\text{m}^{-3})\). \(\theta\) is the symbol used here for soil wetness, and \(s\) the subscript for saturation. Indeed porosity coincides with the total amount of water that can be held by the soil in its pores (saturation soil wetness). Porosity of most soils is of the order of 0.5 \(\text{m}^3\text{m}^{-3}\), except for soils with high organic content, e.g., peat, where values as high as 0.8 \(\text{m}^3\text{m}^{-3}\) can be found.

The soil is characterized by its texture (the size distribution of the soil particles), structure (the spatial organization of the soil particles), composition (the types of minerals existent in the soil), and water content. For heat and moisture transfer, the most relevant variables are water content and texture, the former modulating the intensity of heat and moisture fluxes, while the latter determines essentially the amount of water that the soil can hold against the combined effects of gravity and pressure (see next section). In terms of texture, any given soil is normally characterized by its percentage of clay, silt and sand. For the purposes of classification, the United States Department of Agriculture (USDA) has defined a textural triangle, each side corresponding to a type of particles, sand, silt or clay. Any particular soil becomes a point inside this triangle, according to its texture. Areas within the triangle become soil types. USDA defines eleven soil types, ranging from the finer textures (clay), through the intermediate textures (loam), and coarser textures (sand).

In order to integrate Eq. (7) one needs to specify values for its coefficients. The soil heat capacity can be estimated as an weighted sum of the heat capacity of its phases (deVries, 1963, 1975). The air heat capacity being three orders of magnitude smaller than the other phases, we can write

\[
(rC)_s = (1-\theta_s)(rC)_m + \theta (rC)_w
\]

where the subscript \(m\) and \(w\) refer to the soil matrix and water respectively. \((rC)_m\) ranges
from $2 \times 10^6$ Jm$^{-3}$K$^{-1}$ for most minerals, up to $2.5 \times 10^6$ Jm$^{-3}$K$^{-1}$ for organic matter. Because $(\rho C)_w$ is $4.2 \times 10^6$ Jm$^{-3}$K$^{-1}$, typical values for $(\rho C)_i$ are around $3 \times 10^6$ Jm$^{-3}$K$^{-1}$, and vary by a factor of two depending on its water content. For a frozen soil, Eq. (8) can be modified to

$$
(\rho C)_s = (1 - \theta_s)(\rho C)_m + \theta_w(\rho C)_w + \theta_i(\rho C)_i
$$

(9)

where $\theta_s$ and $\theta_i$ stand for the liquid and ice water content of the soil, respectively and $(\rho C)_i$ = $1.9 \times 10^6$ Jm$^{-3}$K$^{-1}$. Thermal conductivity depends not only on soil texture and water content, but also soil structure. de Vries (1975) developed a general theory, that can be used for a given soil in the field, but inadequate (too detailed) for application to large scale modelling. A simple expression in terms of water content and soil type can be found in McCumber and Pielke (1981). For a medium texture soil (loam, type 5 on the USDA classification), values range from 0.428 Wm$^{-1}$K$^{-1}$ for a dry soil up to 2.24 Wm$^{-1}$K$^{-1}$. Values for coarser (sandy soils) can be twice as large.

A complete theory of heat and moisture transfer must describe the moisture transfer under the combined influence of gradients of temperature and moisture content (see next section), and the heat transfer under the influence of temperature gradients and mass flow of moisture. The theory has been developed by Philip and deVries (1957) and deVries (1958), and further generalized by Milly (1982), in the presence of hysteresis, but it is not in use by current GCM parametrizations. In practice Eq. (7) can be used for all cases, provided (9) is used for the heat capacity, and some other relation is used to define heat conductivity in terms of soil wetness. A notable exception is the treatment of frozen soils. In the absence of snow cover, a soil that is cooling from, say 5°C, up to the freezing temperature will stay at a constant temperature, around 0°C, during a couple of weeks (see Williams and Smith, 1992, and Verseghy, 1991), and afterwards will continue its cooling. This corresponds to heat released by the phase change of the water in the soil into ice: until the whole water in the soil column is frozen, any further radiatively driven cooling will not produce lower soil temperatures. This is a common phenomenon in high latitudes during the autumn and early winter; the reverse mechanism, when the soil is thawing in spring, holds the temperature close to 0°C for much longer than Eq. (7) would predict. If a model does not consider phase changes of water, and their impact on the heat budget, the near soil temperature will reveal a cold bias during the freezing in the autumn, and a warm bias in spring thawing - the warm bias will be less common, because the soil will typically be covered with snow, acting as an insulator layer.

The top boundary condition for the integration of Eq. (7) is of the flux form, the net heat flux at the soil/atmosphere interface. At the bottom, the boundary condition can be given (1) as a no-flux boundary condition (if the soil is deep enough), (2) by specifying a seasonal heat flux at the bottom, or (3) specifying a seasonally varying temperature at the bottom. In practice, for integration of Eq. (7) one needs to discretize in space and time. Discretization in space means choosing soil layers of a given depth, each layer will be characterized by its thermal inertia (Warrilow et al., 1986), the upper layers changing more rapidly than the lower layers (see also Dickinson, 1988).

3.2 SOIL-WATER TRANSFER
The movement of water in the unsaturated zone of the soil obeys the following equation
\[ \rho_w \frac{\partial \theta}{\partial t} = -\frac{\partial F}{\partial z} + \rho_w S_\theta \]  

(10)

where \( \rho_w \) is the density of water (kg m\(^{-3}\)), \( F \) is the water flux in the soil (positive downwards, kg m\(^{-2}\)s\(^{-1}\)), and \( S_\theta \) is a volumetric source term (m\(^3\)m\(^{-3}\)s\(^{-1}\)), corresponding to root extraction (the amount of water transported from the root system up to the stomata - due to the difference in osmotic pressure - and then available for transpiration), and phase changes of ice to liquid water. Eq. (10) assumes that horizontal transfers are negligible, which holds for a grid-box of 50x50 km: for much smaller scales, local terrain slope can induce large horizontal water fluxes.

As seen in Sect. 2., Richardson (1922) and Richards (1931) extended Darcy’s law to the flow of water in the unsaturated case, expressing the water flux in terms of a gradient of the hydraulic head, \( \psi \) (units m of water), and gravitational potential,

\[ F = -\rho_w [\gamma(\psi) \frac{\partial \psi}{\partial z} - \gamma(\psi)] \]  

(11)

Once more we are neglecting lateral gradients of \( (\psi-z) \) in the mechanisms responsible for water transfer. The matric head is homogeneous (has the same units) to the symmetric of pressure \( p \), and \( g(\psi-z) \) can be interpreted as an energy per unit mass, or the work necessary to extract from the soil a unit mass of water against capillarity and gravity. The hydraulic conductivity \( \gamma \) (m s\(^{-1}\)) is a function of the pressure head. In order to integrate (10) with the flux definition (11), an expression is needed for \( \psi = \psi(z) \), or, alternatively, \( \psi = \psi(\theta) \). Such an expression would allow us to write

\[ F = -\rho_w [\gamma(\theta) \frac{\partial \psi}{\partial z} - \gamma(\theta)] \]  

(12)

Hysteresis effects mean that \( \gamma \) is neither a unique function of \( \psi \) or \( \theta \). Defining the hydraulic diffusivity, \( \lambda \) (m\(^2\)s\(^{-1}\)), as

\[ \lambda = \gamma \frac{\partial \psi}{\partial \theta} \]  

(13)

(12) can be rewritten as

\[ F = -\rho_w [\lambda(\theta) \frac{\partial \theta}{\partial z} - \gamma(\theta)] \]  

(14)

Combining (14) with (10), we obtain

\[ \frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z} [\lambda(\theta) \frac{\partial \theta}{\partial z}] - \frac{\partial \gamma}{\partial z} + S_\theta \]  

(15)

231
Eq. (15) with a zero source term is normally called Richards equation. Neglecting the second and third term in the r.h.s., Eq. (15) looks like a Fickian equation for diffusion. Hillel (1982) cautions against pushing too far this analogy, because the process modelled by Eq. (15) is not molecular (or turbulent) diffusion but water transfer through a porous media. The other difference between (15) and (7), of a more practical nature, is in the forcing (Savijärvi, 1992): the heat flow equation in the soil is forced in spring and summer mainly by net radiation at the surface, a sum of harmonics corresponding to the diurnal and the seasonal forcing modulated by the presence of cloudiness. In contrast, the forcing of (15) at the surface is the quasi-random signal of precipitation.

Integration of Eq. (15) needs functional relationships for $\lambda = \lambda(\theta)$ and $\gamma = \gamma(\theta)$. Hillel (1982) reviews several empirical equations proposed, and Mahrt and Pan (1984) compared an extensive set of measured relationships for different textures. They all obey the functional form (Clapp and Hornberger, 1978, Cosby et al., 1984)

$$\gamma = \gamma_{sat} \left( \frac{\theta}{\theta_{sat}} \right)^{2b-3}$$

$$\lambda = \lambda_{sat} \left( \frac{\theta}{\theta_{sat}} \right)^{b+2}$$

(16)

where $\gamma_{sat}$ and $\lambda_{sat}$ are the values at saturation of hydraulic conductivity and diffusivity, respectively, and $b$ is a non-dimensional coefficient, called the Clapp and Hornberger exponent. Note that Eq. (16) indicates a stronger variation in terms of soil moisture for conductivity than diffusivity, with twice as many orders of magnitude covered. Clapp and Hornberger and Cosby et al. measured values for $\gamma_{sat}$ and $\lambda_{sat}$ for a very large sample of soils classified in the eleven soil classes of the USDA classification. The three coefficients vary widely over the textural classes: i) $\gamma_{sat}$ varies between $1 \times 10^{-4}$ and $1.76 \times 10^{-2}$ ms$^{-1}$ between coarser textures (sand) and finer textures (clay), respectively; ii) $\lambda_{sat}$ varies between $8.22 \times 10^{-5}$ m$^{2}$s$^{-1}$ and $2.18 \times 10^{-2}$ m$^{2}$s$^{-1}$, between coarser and finer textures; iii) $b$ values range from 2.8 to 11.5 for coarser and finer textures, respectively. For a given soil type, variations in terms of soil moisture are even more dramatic. For a medium texture soil (loam), $\gamma$ varies between $1.07 \times 10^{10}$ ms$^{-1}$ and $1.48 \times 10^{10}$ ms$^{-1}$ for a range of acceptable values of soil moisture from a dry soil to a wet soil, respectively; $\lambda$ for the same soil moisture values is $5.77 \times 10^{7}$ m$^{2}$s$^{-1}$ and $9.29 \times 10^{5}$ m$^{2}$s$^{-1}$, respectively.

The non-linear variations of $\lambda$ and $\gamma$ in terms of soil moisture makes Eq. (15) (or its counterpart expressed in $\psi$ terms) difficult to integrate, even numerically. Wang (1992) reviews analytical solutions of the Richards equation for some simple representative cases covering each of the four quadrants of the hydrological rosette (Fig. 3), with very idealized formulations for precipitation, evaporation and the recharge flow as lower boundary condition (see also Eagleson, 1978). Boundary conditions for Eq. (15) are, at the top, infiltration (that part of precipitation that is available to wet the soil, i.e., precipitation minus surface runoff minus interception) and, at the bottom, the flow of water from underneath. Two lower boundary conditions can be envisaged (Abramopoulos et al., 1988), corresponding to the two limit cases of bedrock or no bedrock underlying the soil domain, the former specifying $F$ as its drainage ($\gamma$) part, while the latter corresponds to a zero water flux. Some authors (see e.g., Wang, 1992, Dümenil and Todini, 1992) include in the lower boundary condition a base flow,
mimicking the effect of lateral subterranean flow, as well as vertical water loss.

From all generalizations to equation (15) (see Philip and deVries, 1957; Philip, 1957; Milly, 1982), there are two of items of relevance for large scale atmospheric models. The first one relates to the inclusion of water vapour. For arid regions, the soil vertical transfer in the upper few cm is driven by variations of the water vapour in the soil pores (for a more detailed discussion see Sect. 3.3 below). However, because this transfer occurs in a very shallow top layer, a vertical resolution typical of current parameterization schemes prevents its explicit consideration. Mahrt and Pan (1984) proposed a parameterized way of dealing with it. The second deals with the inclusion of the solid phase of water, essential for modelling the soil water transfer in high latitudes. Its is possible to write additional equations for the conservation of frozen water at different soil layers (Verseghy, 1991, Pitman et al., 1992). An alternative way to deal with it is to specify the percent of soil water in the frozen state, in terms of temperature, and to specify its effect on soil water transfer in terms of a frozen water matric potential (Peter Cox, personal comm., Black and Tice, 1988, Williams and Smith, 1992, Miller, 1980). The latter approach avoids the use of additional prognostic variables, but requires the same vertical discretization for soil moisture and temperature. The term $S_0$ in that case will include melting/freezing effects, with an equivalent term for the energy equation. Proper consideration has to be given to the root extraction term, in order to make it ineffective in the presence of frozen soils: in boreal forests this retards the beginning of significant evaporation until late May or June when the whole column of the soil is thawed, long after the solar radiation starts to increase (Sellers et al., 1995).

3.3 BOUNDARY CONDITIONS

3.3.1 Surface sensible heat flux

A way of defining the sensible heat flux was already presented in Eq. (2), including the dependency on atmospheric stability and the consideration of different roughness lengths for heat and momentum. Many versions of Eq. (2) have been presented recently in the literature and the reader is referred to Dyer (1974), Garratt (1992), Högström (1988), and Stull (1988) for details. We will mention here only a couple of relevant points. Beljaars (1995) include in the definition of velocity a convective velocity scale, or free convection velocity, $w_*$, characterizing the intensity of turbulence in well mixed unstable boundary layers. Its inclusion ensures a proper asymptotic limit of the stability dependent functions in case of free convection. The skin layer temperature, $T_{sk}$, can also be used a temperature representative of the surface. The skin layer temperature is representative of a infinitesimally thin layer with zero heat capacity, isolating the underlying soil from the radiative heating above. The use of the skin layer, instead of the average soil temperature characteristic of the top few cm in the soil, reduces the ground heat flux amplitude and phase errors, and the phase error in the diurnal cycle of sensible and latent heat flux (Betts et al., 1993). If an average soil temperature is used, the depth of the soil represented by that temperature has to be warmed, for the temperature to rise in response to the radiative forcing, introducing a phase lag of 1-2 hours between the peak of radiative forcing and the peak of the top soil temperature (Beljaars and Betts, 1993).

Most GCMs do not distinguish the roughness length for heat and momentum. However, the mechanisms responsible for drag are related to bluff body effects, and are more efficient than the conduction at the interface through which the heat transfer is performed. Brutsaert (1982) presents abundant experimental evidence for the difference between the two roughness
lengths, while Beljaars and Viterbo (1994) show the impact of properly separating the two values onto spring systematic errors on near surface humidity over well wet mid-latitude areas (Cabauw, the Netherlands). The problem is more serious in GCMs because roughness lengths are order of magnitudes larger than their micrometeorological values (see Mason, 1988, 1992), where the necessary compensating decrease of the roughness for heat is hardly ever used. A stability formulation in terms of $z/L$ allows a direct use of measured stability functions, and an easy way of separating heat and momentum roughness length. The disadvantage is the need to solve a non-linear equation for every time step expressing the Obukhov length in terms of the bulk Richardson number. A stability formulation in terms of a Richardson gradient number, for $z_{an}^{a} / z_{ob}$ is presented in Mascart et al. (1995).

3.3.2 Evaporation

Formally, an equation analogous to (2) can be written for the evaporation,

$$E = \rho C_{H} \|U\| [a_{L} q_{L} - a_{s} q_{sat}(T_{sk} p_{s})]$$

(17)

where $a_{L}$ and $a_{s}$ are factors dependent on the soil cover and state, vegetation type and phenological state, environmental factors like radiative forcing and humidity saturation deficit at the surface: $a_{L} a_{s} < 1$. The first approach for specifying evaporation in a GCM was given by Manabe (1969) based on the concept of "evaporative factor" $\beta$ (Budyko, 1974). In our notation, that corresponds to

$$a_{L} = a_{s} = \beta = \min(1, \frac{\theta}{\theta_{max}})$$

(18)

where $\theta$ and $\theta_{max}$ are the actual and maximum contents of a single reservoir of water (bucket) in the soil. A separate prognostic equation for $\theta$ is needed for the definition of (18) at every time step. Milly (1992) discusses the proper use of (17) and (18); as mentioned earlier, in some cases a different equilibrium temperature, characteristic of the conceptual irrigated surface, is needed, defined as $a_{s} q_{sat}(T_{sk}) = \beta q_{sat}(T_{eq})$, and compatible with Budyko’s notion of evaporative fraction and potential evaporation.

The bucket model is inadequate for a number of reasons. First, the use of only one layer in the soil (a single soil reservoir) does not allow a quick response to precipitation events. For bare soil, it is an observational fact that, after a very brief period at the potential rate, evaporation is greatly reduced after the loss of water in the first few cm of the soil (representing at most 1 cm of water). The bucket model overestimates bare soil bare soil evaporation in almost all regimes. For vegetated areas, the bucket model fails to recognized the impact of the canopy resistance on evaporation. For unstressed vegetation, the bucket model gives potential evaporation ($r_{t}=0$), instead of unstressed evaporation (minimum, non-zero, value of $r_{t}$), as an estimate for real evaporation. As discussed before, this will produce a significant over-estimation of transpiration: this leads to overevaporation following a precipitation event, producing a rapid drying of the soil and, afterwards, too small evaporation because there is not enough water available in the soil for evaporation. For that reason, a number of alternative methods exist to estimate evaporation in GCMs, and in the following sections we will present their physical/physiological basis.
3.3.2.1 Canopy evaporation

The canopy evaporates due to two different mechanisms. The wet part of the canopy, covered by a thin film of water, shortly (a few hours) after a precipitation event or deposition of dew, evaporates at the potential rate (Deardorff, 1978); note that this can be substantially larger than the unstressed transpiration, hence the interest in the distinction. In our notation \( a_L = a_s = 1 \). The role of surface parametrization in this case is to establish the fraction of the canopy that is wet, and to have a predictive equation for the intercepted water, in order to establish how long the canopy can evaporate at the potential rate.

For dry canopy, most of the current parametrization schemes use the so-called "big leaf-big stomata" (Monteith, 1965) approach. These models recognize that plants evaporate due essentially to a two-path mechanism: i) the first, physiologically controlled, transports water from the root zone up to the stomatal cavities, where water vapour is at saturation; ii) the second, aerodynamically (environmentally) controlled, transports water from the stomata up to the air in the surface layer. Plants regulate evaporation by controlling the aperture of the stomata. In a steady state, the aperture of the stomata is the mechanism that controls the root water upflow, and is enough to characterize the whole physiological control of evaporation. In most species, stomata regulate the outflow of water vapour and the intake of carbon dioxide for photosynthesis; the energy required for opening and closing the stomata, and for the uptake of the root zone soil water is provided by radiation (photosynthetically active radiation, PAR, mainly the visible part of shortwave radiation, roughly 55% of the total shortwave). Under most environments, the system appears to operate in such a way as to maximize the carbon dioxide intake for a minimum water vapour loss. When soil moisture is scarce, the stomata close to prevent desiccation of the plant.

The integrated behaviour of the whole canopy can be conceptually represented by the schematic resistance diagram of Fig. 4 (from Dickinson et al., 1991). The right hand side represents the sensible heat transfer, while the left hand side represents evaporation. The equation corresponding to such a resistance network is

\[
E = \frac{\rho q_L^e - q_c}{r_a + r_c} = \rho \frac{q_L - q_{sat}(T_{sk})}{r_a + r_c}
\]

(19)

where \( q_c \) is the specific humidity in the stomata cells. The canopy resistance is the integral mean of the resistance of the individual leaves, assumed to act in parallel (Dickinson et al., 1991):

\[
r_c = \frac{<r_s>}{L}
\]

(20)

where the brackets represent a leaf-area inverse average (conductances, \( 1/r_s \), are linearly averaged)

\[
<(\ )> = \frac{\int_L L \frac{dL}{L} (\ )}{\int_L L (\ )}
\]

(21)
The quantity measured for a single plant is the stomatal resistance, $r_s$. Most of the atmospheric models use the multiplicative approach of Jarvis (1976), to express the surface resistance as a product of a minimum factor times a number of limiting factors

$$r_s = r_{\text{min}}f_1(PAR)f_2(\theta)f_3(T)f_4(D)$$

(22)

where $f_1$, $f_2$, $f_3$, and $f_4$ are functions of PAR, temperature, relative humidity, and soil wetness, respectively. $f_1$ is a bell-like shape, peaking at some (species dependent) optimal temperatures, $T_{opt}$, and close to 0 for too low or too high temperatures. The saturation deficit conductance function, $f_2$, is a dominant mechanism for forests, but unimportant for grassland.

The dependence on soil wetness is implemented differently for almost every surface parametrization scheme. $\theta$ represents a measure of the water in the root zone, which is typically the top 1 m of the soil, but varies with ecological type: i) It is shallower (around 0.5 m) for crops, being "man-nurtured" have a less adaptive nature; ii) Around 1 m for mid-latitude and boreal forests; iii) several metres (up to 10 m) deep for savannah-like vegetation; and iv) only 0.5 m for the trees of equatorial forests. A map of the depth of the root zone can be found in Patterson (1990).

All models shut evaporation ($f_2 = 0$) below a critical point, $\theta_{pwp}$, the permanent wilting point, or the value at which plants can no longer recover turgidity even when placed in a saturated...
environment (Hillel, 1982). Being related to the energy involved in the transport of water up to the root zone, the permanent wilting point corresponds to a very large value of the matric potential, $\psi$. A value of $\psi_{pwp} = 153$ m (15 bar) is widely accepted as characteristic value for most soils. Based on Cosby et al. (1984) estimates, it is possible to obtain $\theta_{pwp}$ for each soil type of the USDA classification. Values range from 0.033 m$^3$m$^{-3}$ for coarser sandy soils, up to through 0.145 for loamy intermediate soils, and 0.25 for finer soils. Patterson (1990) presents a critical review of the experimental estimates of $\theta_{pwp}$.

All models have a stress-free evaporation ($f_{st}' = 1$) beyond a critical point which is a fraction of the field capacity; the fraction should lie in the range 0.5 to 0.8 (Shuttleworth, 1993b), but in parametrization schemes ranges between 0.7 and 1. The field capacity is the maximum amount of water an entire column of soil can hold against gravity, or the equilibrium mean value of a column soil water 24-48 hours after wetting the soil (Hillel, 1982). Having such a "loose definition", it is not surprising that estimates of field capacity for one particular soil type vary so widely. Hillel (1982) discusses the problems associated with the concept and use of field capacity, and Patterson (1990) reviews experimental estimates for different types of soil. It is common to associate field capacity to a certain value of hydraulic conductivity, with values ranging from 0.05 mm d$^{-1}$ to 0.1 mm d$^{-1}$ (Warrilow et al., 1986, Eagleson, 1970, Jacquemin and Noilhan, 1990); the value of $\theta_{cap}$ can be estimated in this way from Eq. (19).

These estimates give generally too small a value of field capacity, and an alternative quantitative way of defining it is to assign a certain value of $\psi$, ranging from 1 m (0.1 bar) to 3.4 m (0.33 bar). Patterson (1990) concludes that no single value of $\psi_{cap}$ is appropriate for every soil texture. A value of 3.4 m overestimates $\psi_{cap}$ for fine textures and underestimate it for coarse textures. Estimated values of $\theta_{cap}$ range from 0.093 through 0.288 up to 0.384 m$^3$m$^{-3}$, for coarse, medium, and fine texture soils, respectively. Given that evaporation is zero below $\theta_{pwp}$, and at the unstressed rate above $\theta_{cap}$, the availability, or available water capacity, $\theta_{ava}$, defined as

$$\theta_{ava} = \theta_{cap} - \theta_{pwp}$$

is a very important quantity (Patterson, 1990, Henderson-Sellers et al., 1993). When multiplied by the root depth, it gives the total amount of water available for evaporation or total water holding capacity (units m of water): this quantity determines the size of the reservoir and, as such, is crucial in determining time constants in long drying periods. A geographical distribution of total water holding capacity is given by Patterson (1990).

The different ways of applying $f_{st}'$ relate to the shape of the curve linking the 0 and 1 branches. Some authors have a linear dependency in $\psi$ (Abramopoulos et al., 1988; Federer, 1979), while others have linearity in $\theta$ (Warrilow et al., 1986; Noilhan and Planton, 1989; Blondin, 1991). The other difference lies in the specification of $\theta$ in multilevel models: i) It can be taken as a weighted average value over the root zone, with the weights proportional to the root profile (Blondin, 1991); ii) It can be taken as the minimum value of the soil moisture in the root zone (Abramopoulos et al., 1988), the water is borrowed from where it is "energetically cheaper"; and iii) $f_{st}'$ is a product of corresponding values for each root layer (Federer, 1979). A large uncertainty exists to what method is a better representation of nature.
3.3.2.2 Soil (bare ground) evaporation

As illustrated schematically by Fig. 5 (from Mahfouf and Noilhan, 1991), soil (bare ground) evaporation is due to a combination of two physical processes (Kondo et al., 1990, Mahfouf and Noilhan, 1991): i) Molecular diffusion from the water trapped in the pores of the soil matrix up to the surface/air interface, defined by the humidity roughness length, $z_{eq}$ (Brutsaert, 1982); ii) Laminar and turbulent exchange in the air between $z_{eq}$ and screen level height (e.g. 2 m above $z_{eq}$). Process ii) can be characterized by the atmospheric resistance, $r_{a}$, defined by Eq. (4). Process i) involves a soil resistance, $r_{soil}$, and is dependent on relative humidity adjacent to the free-water surface in the soil matrix. The humidity equilibrium value can be expressed by an exact thermodynamic relationship in terms of soil temperature close to the pores (Philip, 1957). In dry situations the relative humidity in the pores has a strong vertical gradient in the top few mm of the soil. $r_{soil}$ is inversely proportional to the diffusivity of water vapour, and strongly dependent on soil texture and structure (Kondo et al., 1990).

An accurate description of bare soil evaporation can be obtained with a soil model with circa 10 layers in the top 5 cm of the soil, in order to model explicitly the diffusivity of water vapour (McCumber and Pielke, 1981, Sasamori, 1970, Camillo et al., 1983). The small time step required in such a model makes its cost prohibitive for large scale problems. The depth of the first soil layer in GCMs is typically a few cm, too coarse to define explicitly the water vapour transfer.

Mahfouf and Noilhan (1991) made a comparative study of several formulations of evaporation over bare ground. Given the above description it is not surprising that most methods reviewed are very sensitive to the top soil discretization of NWP models. The methods presently available were classified in bulk parametrization approaches ($\alpha$- type and $\beta$- type, following Kondo et al., 1990) and threshold methods.

$\alpha$- methods model the evaporation as a bulk transfer of water vapour between $z_{eq}$ (assumed at relative humidity $\alpha$), and a reference height, typically the height of the lowest model layer. They correspond, in Eq. (22), to $a_i=I$, $a_i=\alpha$. $\alpha$ can assume several parametrized forms (see e.g., Noilhan and Plamton, 1989 and Barton, 1979), depending on the soil water typically in the top 5 cm of soil. In $\beta$-methods the evaporation is a fraction $\beta$ times the bulk transfer of water vapour between the air trapped in the soil pores close to the water in the soil and the reference height. $\beta$ can be related to a resistance, $r_{soil}$, for the transfer of water between the soil pore and $z_{eq}$, and is dependent on the availability of water in the top soil layer (see e.g. Deardorff, 1978, Dorman and Sellers, 1989). The $\beta$-method requires the specification of the

Fig. 5 Schematic illustration of the interaction between the atmosphere and a bare soil.
relative humidity, $h$, of the air trapped in the soil pores according to Philip (1957) formula (and different from $\alpha$, relative humidity at $z_{eq}$). It corresponds, formally, to a rewriting of Eq. (5), with the stomatal resistance replaced by $r_{soil}$, and $q_{sat}(T_{sk})$ replaced by the specific humidity in the soil pores. Using the notation of Eq. (17), $a_{f}=1/(r_{a}+r_{soil})$, $a_{z}=h/(r_{a}+r_{soil})$.

In threshold formulations evaporation proceeds at potential rate above a certain amount of soil water. In drying conditions, when the soil water is depleted, evaporation is determined by the water flux from below (see Mahrt and Pan, 1984, Wetzel and Chang, 1987, Abramopoulos et al., 1988). The evaporation is given by

$$E = \min\{E_{l}, \frac{Q_{a}}{r_{a}}[q_{sat}(T_{sk})-q_{a}]\}$$

(24)

where $E_{l}$ specifies a maximum water flux from below. In Eq. (24) the first term in the r.h.s is the soil limited evaporation (regime of soil supply in Fig. 3), while the second term is the atmospheric demand term. $E_{l}$ can be specified in several ways, e.g., the hypothetical water flux between the centre of the top soil layer and the surface assumed at $\theta_{pp}$ (Mahrt and Pan, 1984), who noted that the threshold method depends on the estimation of the soil water flux in a top layer of depth $d$, and is highly sensitive to the value of $d$.

The main conclusions of Mahfouf and Noilhan (1991) can be summarized as follows. Threshold methods are very sensitive to the specification of the depth of the top soil layer, where the diurnal cycle is felt, and provide neither a correct diurnal cycle of evaporation, nor the correct transition (on weekly timescales) between the regime of atmospheric demand and the regime of soil supply. $\alpha$- and $\beta$- methods give results of similar quality at day time. During night time, special care has to be taken in order to handle correctly the dew deposition: the two methods give differences in cumulative evaporation of up to 20%.

3.3.3 Interception

The role of vegetation in intercepting precipitation and collecting dew is a relatively old subject in the meteorological literature. Horton (1919) established statistical relationships, valid for monthly periods, between measured net rainfall reaching the ground and over the canopy. Parallel to the statistic estimates of interception (the difference between the rainfall over the canopy and the net rainfall reaching the ground), physically based models of the interception reservoir have been developed, starting with the work of Rutter et al. (1972, 1975). Those models evolved from a bulk representation of the canopy into multi-layer plant models (Sellers and Lockwood, 1981). Multi-layer canopy models are too detailed to be coupled with GCMs, but the basic mechanisms of the bulk representation of vegetation have been incorporated into many land surface parameterization schemes (Deardorf, 1978, Dickinson et al., 1986, Sellers et al., 1986, Xue et al., 1991, Noilhan and Planton, 1989, Blondin, 1991). For a review of suitable interception models for coupling with atmospheric models see Mahfouf and Jacquemin (1991). They do not include a proper treatment of interception of snowfall: see Miller (1967) for a discussion of snowfall interception.

All interception models obey the following equation for the evolution of the grid-box estimate of the canopy water content, $W_{i}$ (m)
where all quantities are positive downwards. \( P^* \) (kg m\(^{-2}\)s\(^{-1}\)) is a modified precipitation rate at the top of canopy level, taking into account the subgrid scale distribution of precipitation, \( a_E \) is the evaporation coming from the interception reservoir, or wet evaporation, \( D \) (kg m\(^{-2}\)s\(^{-1}\)) is a rate of drainage from the canopy (it is a sink term in the equation, \( D \leq 0 \)), and \( E_i \) (kg m\(^{-2}\)s\(^{-1}\)) is the evaporation rate of the wet fraction of the grid-box, defined as the potential evaporation rate. \( a_p \) and \( a_i \) are non-dimensional factors expressing respectively the efficiency of interception and the percent of the grid-box covered by the interception reservoir (wet fraction). \( a_p P^* + D \) is the interception of rain by the canopy, \( I \) (kg m\(^{-2}\)s\(^{-1}\)). \( P - I = P - a_p P^* - D \) is called throughfall, all "rainfall" that is available at the ground level. Some authors separate drainage into dripping from the canopy and stemflow, but we will ignore the distinction here. Because the total size of the interception reservoir is very small (0.1-1 mm), interception has to balance wet evaporation on a daily time scale. Note that when \( P = P^* \), or \( a_p = 1 \), the absolute value of throughfall and drainage are identical. What distinguishes any individual model is the specification of \( a_p, a_i, D \), and the subgrid scale assumptions involved in \( P^* \).

\( a_p \), expressing the probability of interception of rainfall by the canopy, is estimated to be 0.75 over a Corsican pine plantation, by Rutter et al. (1972), while Shuttleworth (1988a) gives 0.92 for the Amazon forest. Warrilow et al. (1986) and Dolman and Gregory (1992) use \( a_p = 1 \). The sensitivity of daily canopy evaporation to the value of \( a_p \) is small when \( a_p \) lies between 0.85 an 1 (Shuttleworth, 1988b). Noilhan and Planton (1989) set \( a_p = C_v \), where \( C_v \) is the fraction of the grid-box covered by vegetation, while Blondin (1991), somewhat empirically uses \( a_p = 0.25 C_v \), probably as an attempt to include the sub-grid scale variability of interception.

When an interception model is used to validate point measurements, the observed precipitation rate can be used directly as a forcing term in Eq. (25), \( P^* = P \). However, global model grid-boxes vary between 60x60 km\(^2\) and 400x400 km\(^2\) and, for this scales, it is unrealistic to assume an uniform coverage of precipitation over the grid-box, particularly in the case of convective precipitation. The atmospheric parametrization schemes provide an average rainfall rate representing the average value over the grid-box, separated in large scale precipitation (occurring due to supersaturation of the mean values of humidity) and convective precipitation. The sub-grid distribution of precipitation is described in many hydrological and meteorological papers. A review of formulations that have been applied to GCMs can be found in Thomas and Henderson-Sellers (1991) and Pitman et al. (1993); see also Chapter 4 for an example of the sensitivity of the surface branch of the hydrological cycle of GCMs to the sub-grid scale distribution of precipitation. A proper way of taking into account the heterogeneity should be dependent on the grid-box size and the time step used. Note that although the instantaneous cover of precipitating convective elements within a grid-box is a small value (of the order of 0.1), the effective cover can be several times larger, in order to take implicitly into account the movement of convective cells inside the grid-box during one time step. Two distributions for precipitation intensity inside a grid-box have been used: an exponential distribution (Shuttleworth, 1988b), and a box distribution (Viterbo and Illari,
The idea is to assume that an equation like (25) holds for each grid point, with \( P^* = P_t \), where \( P_t \) represents the local precipitation rate; the value of \( P^* \) is obtained by multiplying \( P_t \) by its probability density function (pdf) value and integrating over the grid-box. Note that \( P_t \), the grid-box precipitation rate given by the atmospheric forcing, is the expected value of \( P_t \) over the grid-box. \( P^* \) obtained in this way depends on a shape parameter of the pdf which, for the box distribution of Viterbo and Illari (1994), is the fraction \( k \) of the grid-box covered by precipitation, taken as 1 for large-scale precipitation and 0.5 for convective precipitation. For the functional form of \( P^* \) using the exponential distribution see Dolman and Gregory (1992); a slightly different distribution was used by Eltahir and Bras (1993).

Let us now turn our attention to the evaporation part of Eq. (25). The fraction \( a_t \) multiplying the potential evaporation rate was first defined by Rutter et al. (1972) as \( W_t/W_{\text{max}} \), where \( W_{\text{max}} \) is a constant, the maximum capacity of the interception reservoir. Assuming a constant potential evaporation, Eq. (31) will give an exponential decrease of the water in the interception reservoir. Under the sole effect of evaporation the interception reservoir would never go to zero, which led Deardorff (1978) to assume a power law for \( a_t = \min\{1, (W_t/W_{\text{max}})^{23}\} \). The previous expression can be multiplied by the vegetation cover, in which case \( (W_t/W_{\text{max}})^{23} \) can be interpreted as the fraction of the vegetation cover that is wet (Warrilow et al., 1986, Noilhan and Planton, 1989). The maximum size of the interception reservoir, \( W_{\text{max}} \), is proportional to the leaf area index, \( L_f \) times the amount of water that can be stored on a single leaf \( (W_{\max}) \) or, in case the reservoir includes ponding of bare soil (Blondin, 1991),

\[
W_{\text{max}} = [C_v L_f(1 - C_v)] W_{\max}
\]

The remaining term to be specified in Eq. (25) is drainage. This is the term where most existing formulations differ. We will disregard stem flow in our discussion, because its consideration require an additional reservoir of water related to the storage in the trunks (Rutter et al., 1975). Formulations in the literature are divided in two broad classes: i) Allowing drainage before the interception reservoir is saturated (Rutter et al., 1972, Warrilow et al., 1986, or Dolman and Gregory, 1992, for the case of a sub-grid scale distribution of precipitation); ii) Threshold formulations, where drainage occurs only if the reservoir is saturated (Dickinson, 1984, Noilhan and Planton, 1989, or Blondin, 1991, for the case of sub-grid scale distribution of precipitation).

The total storage of the canopy reservoir, \( W_{\text{max}} \), is the main parameter controlling the behaviour of a particular interception model (Shuttleworth, 1988b). Model results are also sensitive to details of the sub-grid distribution of precipitation, as shown by Shuttleworth (1988b), Dolman and Gregory (1992), Thomas and Henderson-Sellers (1992) and Viterbo and Illari (1994). If the canopy storage is also allowed to vary within a grid-box, the drainage and interception terms assume a different form, depending on the assumed pdf distribution for \( W_t \); a complete generalization of Eq. (25) can be found in Eltahir and Bras (1993).

Since the processes described in this section are very fast (small time scales), the integration of Eq. (25) poses some numerical problems. The first is stability: the method of Kalnay and Kanamitsu (1988) is commonly used, where the fastest non-linear term (evaporation) is treated implicitly. The second problem relates to time truncation errors (see Dolman and Gregory, 1992). The third problem is mass-conservation: it is not trivial to formulate a numerical implementation of threshold methods for the drainage term that conserves mass for the water
substance. In the ECMWF model, a rate of evaporation from the interception reservoir is needed for the specification of the total evaporation, for use as a boundary condition for the turbulent vertical diffusion of water vapour in the atmosphere. Because of the semi-implicit scheme used in the time integration of the surface equations, it is not possible to ensure a perfect matching between wet evaporation, as seen by the atmosphere, and the amount of water lost by the interception reservoir. Viterbo and Beljaars (1995) discuss a scheme that attempts to minimize the different time truncation errors described above, while at the same conserving the mass of the water substance and having the necessary requirements of stability.

3.3.4 Runoff and infiltration

A proper account of runoff is not common in most GCMs, specially those used in NWP. In connection to ongoing international programs, such as the Global Water Budget Experiment (GEWEX) and GCIP (GEWEX Continental-scale International Project), it is likely that this situation will change rapidly in the near future. Note that a correct treatment of runoff will be specially important for coupled ocean-atmospheric models (to define the fresh water inflow to the oceans) and long time scales. In this section, some basic terminology and concepts from hydrology will be introduced and the relation with similarly named quantities in GCMs will be explored. A qualitative description of runoff will be presented, emphasizing its role in the soil water controlled part of the hydrological rosette (phase A-D, Fig. 3). A rainfall-runoff scheme, making use of sub-grid scale distribution of soil parameters, and currently used in GCMs (Dümenil and Todini 1992; Liang et al. 1994) is presented; all other discussions related with sub-grid heterogeneity will be postponed to Sect. 5. Two routing schemes, for the lateral transport of soil water in GCMs, will be referred to. They allow the computation of fresh water discharge of the major rivers into the ocean, and can be used directly during a coupled ocean-atmosphere GCM run, or for post-processing GCM output. The model runoff averaged over a large basin is often compared to river discharge to the oceans (see e.g., Dümenil and Todini 1992).

Infiltration is that part of the precipitation flux that contributes to wet the soil. At the interface soil-atmosphere, continuity of the water flux can be written as

$$I_f = T - Y_s$$

(27)

where $I_f$ and $Y_s$ are, respectively, the infiltration and the surface runoff (units kgm$^{-2}$s$^{-1}$). Eq. (27) is part of the boundary condition to integrate Eq. (15) for the soil water transfer, the remaining part being evaporation. For most GCMs, and certainly for all current NWP models, Eq. (27) defines $Y_s$ as a quantity that is lost to the model water cycle: there is no horizontal transfer of water in the soil. $Y_s$ represents in these GCMs an instantaneous, local response to precipitation, with no time-lag and no notion of a drainage network or a drainage basin attached to it. $Y_s$ is a scalar because no direction of water transfer needs to be specified.

In hydrology studies, $Y_s$ in Eq. (27) represents a divergence of an horizontal soil water flux at the surface. The possible different paths taken by rainfall and snow melt water until it reaches a stream channel are described schematically in Fig. 6 (from Dunne, 1978). Water impinging on a hill side of limited infiltration capacity becomes overland flow, but if the precipitation (snow melt) is absorbed by the soil and encounters some shallow depth of impervious material, it creates a subsurface horizontal flow, named subsurface streamflow.
Subsurface streamflow that returns to the ground surface and leaves the hill side is named return flow; together with the precipitation that falls into saturated areas, they form the saturation overland flow. Finally, if precipitation is absorbed by the soil it will slowly percolate until it reaches the saturated zone in the soil, and will eventually be diverted horizontally to the nearest stream channel. The water collected by the different paths detailed above, arrives to the stream channels and is collected in catchments of different sizes until eventually arriving at the ocean as fresh water inflow.

The role of parametrization in GCMs, with regard to the term $Y_r$, would be to represent the transport by the above 4 mechanisms, plus the water transport by the stream channels. The time scales of the response to a precipitation event by each of the above mechanism are very different. Hydrologists separate flow into *streamflow* responding directly (but with a time and space lag) to a precipitation event, and *baseflow*, that part of the flow that is to a large extent independent of precipitation events (but its intensity will reflect long term anomalies of precipitation). The first 3 mechanisms described in Fig. 6 are essentially contributors to storm runoff, while path 4 will be partly a contributor to baseflow and partly to the storm flow (depending on the position of the water table).
As described in Fig. 3, when precipitation starts following a dry spell, it will first meet a dry soil, with a large infiltration capacity (route B-C in the figure): the rate of change of soil wetness is determined by the precipitation rate, the regime is atmospheric controlled. If the precipitation event is long enough, the infiltration capacity - maximum rate of top infiltration for a given soil condition - will reduce while the soil gradually fills up with water, until it eventually becomes smaller than the precipitation rate (point D in the picture). Beyond that point, part of the precipitating water goes into runoff; the fraction of water that goes into runoff is determined by the soil state and its morphology, area C-D in Fig. 3 is under soil control.

Two possible mechanisms, namely the Horton runoff and the Dunne runoff (see, e.g., Bras, 1990), have been identified to describe the different water paths in Fig. 6. Any real runoff event will be due to either one of the two mechanisms or a combination of the two, dependent on the precipitation intensity, type and state of soil, morphology of the terrain, and the soil cover. The Horton mechanism occurs mainly in up slope areas, in those parts of the watershed where conductivities are lowest; runoff occurs in areas where the infiltration capacity is lower then the precipitation rate. The near surface soil water content increases until the hydraulic conductivity at the surface (or infiltration capacity) is identical to the precipitation rate; beyond that point runoff occurs. The soil wetting front comes from above in this mechanism. In the Dunne mechanism, in near-channel wetlands, overland flow or surface runoff will occur when the water table rises up to the surface: it happens preferentially in those parts of the watershed where the water table is shallowest. Runoff starts when the water table rises and eventually meets the surface; the soil wetting front comes from below in this mechanism. The Horton mechanism was the first to be identified, and is responsible for runoff in arid and semi-arid areas, where the total annual precipitation will tend to come in only a few large events, or in humid areas where the original vegetation and soil structure has been destroyed: the reduced vegetation cover or soil erosion create a land surface with low infiltration capacity. In humid temperate areas with a dense vegetation, the soil infiltration capacities are larger and the precipitation is well distributed during the year: the Dunne mechanism will be predominant.

It is clear from the above description that the runoff fraction \( Y/P \) will depend on soil type, terrain slope, moisture conditions, vegetation cover, precipitation intensity, and the sub-grid scale variability of the above.

The Arno scheme, designed to represent in a GCM some of the basic mechanisms detailed above is presented in the following. Based on a standard hydrological "lumped" model (Becker 1992), it was adapted for use in GCMs by Dümenil and Todini (1992) and Wood et al. (1992), in connection with a single soil moisture reservoir ("bucket"), and later with multi-level reservoirs (Liang et al. 1994, Katia Laval, person. communication). The version of Dümenil and Todini (DT) is described. The original bucket model, without the "hydrology" parametrization gives the time variation of water in a soil of depth \( D \)

\[
\rho_w D \frac{\partial \theta}{\partial t} = P - Y_s + E
\]  

(28)
where \( P - Y_s = I_f \). \( I_f \) is computed as a threshold quantity

\[
I_f = \begin{cases} 
P & D\theta < W_c \\
0 & D\theta \geq W_c 
\end{cases}
\]  

(29)

\( W_c \) is the capacity of the reservoir, originally defined as 15 cm worldwide by Manabe (1969). The term \( Y_s \) has been generalized by DT to represent storm runoff as overland flow and drainage through the reservoir. To compute the storm runoff, a local balance equation of precipitation, runoff and infiltration (similar to Eq. 29), is coupled to a sub-grid scale distribution of local storage capacity. For a given water content, \( w \), the storage capacity distribution is the fraction of the basin area (or grid-cell in a GCM) for which the infiltration is less or equal \( w \). Using that information for the spatial distribution of water, the local balance equation can be integrated for the entire basin. The shape of the storage capacity distribution curve can be parametrically controlled, dependent on the standard deviation of orography in a grid-cell: steeper sub-grid scale slopes imply larger values for runoff. A grid value for storm runoff, taking into account the sub-grid scale distribution of soil moisture, is therefore obtained.

Runoff, as produced by Eq. (29), together with the model sub-grid scale distribution assumptions, represents the amount of water that is lost locally by the system soil-atmosphere: we will refer to \( Y_s \) in this section as source runoff (Miller et al. 1994). A horizontal advection scheme in the soil can be applied to that field of local imbalance, or source runoff, to produce an horizontal flow of water across land grid-boxes, eventually arriving into the mouth of the rivers. If \( M \) (units kgm\(^{-2}\)) represents the water in the soil that is free to move (representing the river flow and groundwater flow),

\[
\frac{\partial M}{\partial t} = \nabla F + Y_s
\]

(30)

where \( F \) (units kgm\(^{-1}\)s\(^{-1}\)) is a vector field representing the horizontal flux of water in the soil/surface of the model. It is proportional to the quantity of water upstream, with the direction and the proportionality being a specified set of advection rates \( u \). Sausen et al. (1992) and Miller et al. (1994) present alternative ways of defining the advection rates.

4. **Examples of parametrization schemes**

As emphasised in the previous sections, land-surface parametrization schemes for use in GCMs define the surface boundary conditions for the atmospheric momentum, heat and moisture equation: surface stress, surface sensible heat flux and surface evaporation flux, respectively. Because of the long timescales involved with soil moisture, and the diurnal cycle of the forcing, net radiation at the surface, land-surface parametrization schemes need to deal accurately with a timescale ranging from 1 hour to several months.

The Project of Intercomparison of Land Surface Parametrization Schemes (PILPS, Henderson-Sellers et al, 1993) is at the centre of a very large effort of coordinating the modellers' efforts. Its goals include a review of existing parametrization schemes (part of Phase I, already completed), followed by comparison of the model results when forced by
<table>
<thead>
<tr>
<th>Key Model</th>
<th>Number Canopy Layers</th>
<th>Interception Treated</th>
<th>Number of Layers Included for T</th>
<th>Number of Layers Included for Θ</th>
<th>Canopy</th>
<th>Philosophy for Temperature</th>
<th>Soil moisture</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>B BEST</td>
<td>1</td>
<td>yes</td>
<td>3</td>
<td>2</td>
<td>Penman/Monteith</td>
<td>force-restore</td>
<td>Philip-de Vries</td>
<td>Pitman et al (1991)</td>
</tr>
<tr>
<td>C BUCKET</td>
<td>0</td>
<td>no</td>
<td>0</td>
<td>1</td>
<td>instantaneous surface heat balance</td>
<td>bucket + variation</td>
<td></td>
<td>Robock et al (1993)</td>
</tr>
<tr>
<td>D CLASS</td>
<td>1</td>
<td>yes</td>
<td>3</td>
<td>3</td>
<td>Penman/Monteith</td>
<td>heat diffusion</td>
<td>Darcy's Law</td>
<td>Verseghey (1991)</td>
</tr>
<tr>
<td>G ISBA</td>
<td>1</td>
<td>yes</td>
<td>2</td>
<td>3</td>
<td>Penman/Monteith</td>
<td>heat diffusion</td>
<td>force-restore</td>
<td>Nolhan and Planton (1989)</td>
</tr>
<tr>
<td>H topaT</td>
<td>1</td>
<td>yes</td>
<td>1</td>
<td>2</td>
<td>Penman/Monteith</td>
<td>heat diffusion</td>
<td>Philip-de Vries</td>
<td>Famiglietti and Wood (1994)</td>
</tr>
<tr>
<td>I LEAF</td>
<td>1</td>
<td>yes</td>
<td>7</td>
<td>7</td>
<td>Penman/Monteith</td>
<td>heat diffusion</td>
<td>Darcy's Law</td>
<td>Avisar and Pielke (1989)</td>
</tr>
<tr>
<td>J LSX</td>
<td>2</td>
<td>yes</td>
<td>6</td>
<td>6</td>
<td>Penman/Monteith</td>
<td>heat diffusion</td>
<td>bucket</td>
<td>Manabe (1969)</td>
</tr>
<tr>
<td>K MAN69</td>
<td>0</td>
<td>no</td>
<td>1</td>
<td>1</td>
<td></td>
<td>bucket</td>
<td></td>
<td>Manabe (1969)</td>
</tr>
<tr>
<td>L MILLY</td>
<td>0</td>
<td>no</td>
<td>1</td>
<td>1</td>
<td></td>
<td>heat diffusion</td>
<td></td>
<td>Abramopoulos et al (1988)</td>
</tr>
<tr>
<td>M MIT</td>
<td>0</td>
<td>no</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>Entekhabi and Eagleson (1989)</td>
</tr>
<tr>
<td>N MOSAIC</td>
<td>1</td>
<td>yes</td>
<td>2</td>
<td>3</td>
<td>Penman/Monteith</td>
<td></td>
<td>Darcy's Law</td>
<td>Koster and Suarez (1992)</td>
</tr>
<tr>
<td>O NMC-MRF</td>
<td>1</td>
<td>yes</td>
<td>1</td>
<td>1</td>
<td>lumped with soil</td>
<td></td>
<td></td>
<td>Pan (1990)</td>
</tr>
<tr>
<td>P CAPS</td>
<td>1</td>
<td>yes</td>
<td>2</td>
<td>2</td>
<td>Penman/Monteith</td>
<td>heat diffusion</td>
<td>diffusion</td>
<td>Mahrt and Pan (1984)</td>
</tr>
<tr>
<td>Q PLACE</td>
<td>1</td>
<td>yes</td>
<td>30</td>
<td>30</td>
<td>Ohm's law analogy</td>
<td>force-restore</td>
<td></td>
<td>Wetzell and Chang (1988)</td>
</tr>
<tr>
<td>R RSTOM</td>
<td>-</td>
<td>no</td>
<td>1</td>
<td>1</td>
<td></td>
<td>bucket + variation</td>
<td></td>
<td>Milly (1992)</td>
</tr>
<tr>
<td>S SECHIBA</td>
<td>1</td>
<td>yes</td>
<td>2</td>
<td>2</td>
<td>Penman/Monteith</td>
<td>force-restore</td>
<td>Choisnend</td>
<td>Duccoud et al (1993)</td>
</tr>
<tr>
<td>T SSIB</td>
<td>1</td>
<td>yes</td>
<td>2</td>
<td>3</td>
<td>Penman/Monteith</td>
<td>force-restore</td>
<td></td>
<td>Xue et al (1991)</td>
</tr>
<tr>
<td>U UKMO</td>
<td>1</td>
<td>yes</td>
<td>4</td>
<td>1</td>
<td>Penman/Monteith</td>
<td>heat diffusion</td>
<td>diffusion</td>
<td>Warrillow et al (1986)</td>
</tr>
<tr>
<td>V VIC</td>
<td>1</td>
<td>yes</td>
<td>1</td>
<td>2</td>
<td>Penman/Monteith or full energy balance</td>
<td>heat diffusion</td>
<td></td>
<td>Liang et al (1994)</td>
</tr>
<tr>
<td>W BIOME</td>
<td>1</td>
<td>yes</td>
<td>1</td>
<td>1</td>
<td>Penman/Monteith</td>
<td>force-restore</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1  Characteristics of several land surface parametrization schemes
synthetic data sets and field experiment data. Table 1, from Pitman et al. (1993) summarises the characteristics of a large number of schemes. Many schemes include a fast-responsive reservoir for evaporation (the interception reservoir), and separate descriptions of evaporation from the bare ground and transpiration. The representation of longer timescales (such as a complete cycle of the hydrology "rosette" in Fig 3) is guaranteed if the soil water transfer is capable of quick infiltration of soil water at the beginning of a precipitation event, and retention of that water in the soil until the evaporative demand of the atmosphere reclaims it. Heat stored in the soil during summer can serve as a source of energy for the winter months. Finally, in high latitudes or over mountains, the representation of the insulation effects of the snow mantle, with its higher albedo and large heat capacity, is essential for a bias-free near surface temperature.

Land Surface Parametrization schemes (LSP) can be classified in two classes in terms of their potential use:

1) For short, medium and extended range (1-2 days, 3-10 days, 10 days up to a season, respectively) forecasts with NWPs. Operational requirements and the sensitivity to critical conditions prohibits the use of two complex LSPs, with many prognostic variable parameters.

ii) On the other hand, longer term (climate) integrations are more dependent on initial conditions, and the complexity of advanced LSPs can be an advantage for correctly handling the atmosphere/surface interaction.

Because of the lack of observations to initialise the surface state variables, land-surface models in NWP context work very often in "climate mode", with the initial conditions taken from a very short-range forecast (see section 7). Since this requires a correct handling of the longer timescales by the land-surface scheme, the trend should be to try to unify the schemes and abolish the above distinction (see section 4.2 below, ISBA has been applied to both mesoscale and climate modelling).

In the following we will describe a few schemes: the bucket model, chosen for its simplicity and historical importance, the ISBA model, a force-restore scheme, and the ECMWF model, a multilevel model of heat and water soil transfer. The emphasis is on highlighting a few relevant properties; for a complete description of the features of each model, the reader is referred to the documentation of these schemes.

4.1 THE BUCKET MODEL
The bucket model has been introduced by Manabe (1969) as the first attempt to parametrize land-surface processes in GCMs. It is based on Budyko's concept of evaporative fraction, $\beta$. The evaporation is the product of $\beta$ times the potential evaporation. There is a single reservoir of water in the soil, with its contents changing with the combined action of precipitation and evaporation. $\beta$ is given by the ratio of the current contents of the reservoir over its maximum capacity. Runoff exists when the reservoir exceeds its maximum capacity.

As mentioned in the previous section, the bucket model overestimates evaporation over bare ground in all regions, and for dry conditions over vegetated areas. Milly (1992) describes
some improvements to the bucket model that greatly reduce the above bias. Manabe and Delworth (1988, 1989) demonstrated that, when coupled to a GCM, the bucket model correctly works as a low-pass filter to the precipitation forcing; it is capable, therefore, of correctly representing the soil as a source of low frequency variability to the atmosphere, via the evaporation. Pan (1990) has applied the single reservoir concept, but uses a Penman-Monteith evaporation, estimating skin temperature under saturated soil conditions.

4.2 ISBA
ISBA (Noilhan and Planton, 1989) represents a refinement of Deardorff (1978) model. The soil is composed of a thin reservoir at the top and the root layer underneath. The diurnal and seasonal timescale force the thermal equations, while for the water budget the forcing for each layer is parametrized based on the soil texture and Clapp and Noinberger (1979) relationships. Evapotranspiration is a sum of the bare ground component, interception and transpiration.

There are, however, several characteristics that distinguish ISBA. It is a model that has been applied successfully at single-point (Noilhan and Planton, 1989), mesoscale and to the global GCMs (Mahfouf et al, 1991). There has been a constant validation effort (see summarizing table from Mahfouf and Noilhan, 1994) comparing model results to observations in field experiments in a variety of atmospheric conditions and for a wide range of ecosystems. Moreover the author’s demonstrate here that it is possible to derive the model parameters from soil/vegetation characteristics in a consistent way (Noilhan and Lacarrère, 1995): parameter definition is based on a thematic mapping of the terrain, and the concept of "effective" parameters (see next section), representing the heterogeneity of the terrain in a GCM grid box.

4.3 ECMWF SURFACE MODEL
The motivation to develop the current ECMWF surface model (Viterbo and Beljaars, 1995), in operations since August 1993, was to correctly model heat and water air-soil transfer with timescales ranging from the day cycle to the season. Diagnostics of errors of the previous model (Blondin, 1991), when compared against field experiments data (Betts et al., 1993, Beljaars and Betts, 1993) formed the basis to develop a new scheme. The new scheme has 4 predicted soil layers for water and temperature, plus a predicted skin temperature, and a thin surface water layer representing interception of precipitation and collection of snow. The bottom boundary conditions are zero heat flux and free drainage. The 4 soil layers have depths of 7.21, 72, and 189 cm, with a root zone in the first three layers. The deepest layer acts as a reservoir and a memory for the longer time scales (of order a year). The formulation of the soil hydraulic properties is based on the Richards equation with the Clapp and Honberger (1978) relationships for the definition of the soil hydraulic diffusivity and conductivity.

The introduction of a skin temperature, which is calculated from flux equilibrium at an infinitesimally thin layer at the surface, reduces the errors in the ground heat flux, and the phase errors in the sensible and latent heat fluxes at the surface. A smaller roughness length for heat (and moisture) than momentum was introduced with the following benefits: i) it increases the difference between the surface and air temperatures, verifying better with FIFE data (Betts and Beljaars, 1993); ii) it improves the accuracy of the surface longwave emission for the FIFE data (Beljaars and Betts, 1993); iii) it improves the simulation of winter evaporation for the Cabauw data set (Beljaars and Viterbo, 1994).
The evaporation is based on a vegetation fraction, derived from the data set of Wilson and Hendersson-Sellers (1985), treating separately on a grid-box the evaporation of the wet canopy fraction, transpiration from the dry canopy, and bare ground. The scheme has been extensively validated, with emphasis on the long time scales, based on single-column integrations for the FIFE, ARME and Cabauw data sets (Viterbo and Beljaars, 1995), and the SEBEX data set (Beljaars and Viterbo, 1995). Some examples of the one-column validation effort will be presented in Sect. 6. Parallel to the one-column integrations, a 5-year 3D run was performed (Viterbo and Beljaars, 1995), and the performance of the scheme in data assimilation/forecast mode was analysed for the US floods of July 1993 (Beljaars et al., 1995).

5. Subgrid-scale heterogeneity

The size of current GCM grid boxes ranges from 60 x 60 km (ECMWF T213 model used for NWP) up to 4° x 5°, while the time steps used range from 5 minutes up to 1 hour. It is clear that none of the variables or parameters described in the previous chapter are homogeneous in a GCM grid box. Specially for the larger time steps, the forcing to a surface scheme cannot realistically be assured to be uniform in time either: e.g., convective precipitation hardly ever has a constant intensity during periods longer than 10 minutes. This section will describe ways of dealing with the heterogeneity in a grid box or subgrid scale heterogeneity.

Land surface is heterogeneous in all spatial scales, both in its physiography and its physical state. For example, Wetzel and Chang (1987) report on an analysis of variance of soil moisture in the top 10 km of soil, for spatial scales ranging from $10^1$ m$^2$ to $10^3$ km$^2$, and they found significant variance in the whole spectrum. Another example of variability can be found in any satellite image of a snow covered surface, revealing heterogeneity in the snow cover, linked mainly (but not only) to terrain elevation and orientation of the slope.

The problems involved in subgrid scale heterogeneity can be illustrated with the water balance equation (15), together with the parametric relations for the conductivity and diffusivity (16), and the boundary conditions at the top, defining the surface water fluxes, and at the bottom, the drainage flux.

- The first non-homogeneous quantity is the evaporation flux. Parts of a grid-box size square over land will be covered with vegetation, which in turn can have moisture deposited on its leaves. In winter, there will be snow covered and snow free parts of the terrain. As detailed in the previous chapter, the rate of evaporation of all the above parcels will be substantially different. Heterogeneity in the fluxes (which are part of the output of a land-surface scheme) is discussed in Sect. 5.1.

- The main forcing term of Eq (15) is precipitation. Precipitation is far from uniform in a grid-box size portion of land: i) the size of the rain generating convective cells is typically 1/10 of a grid-box (Emanuel et al, 1994); and ii) frontal rain shows very often organisation in the mesoscale, with variations in intensity in scales only partly resolved even in the finer resolution GCMs. Heterogeneity of the forcing will be discussed in Sect. 5.2.

- Most of the components of the land-surface parametrization depend on parameters which cannot be considered homogeneous in a grid-box, even in a uniform fraction of it: e.g. the dry vegetation part of a grid-box will have non-uniform evaporation
rates arising from a varying canopy resistance, corresponding to different stomatal apertures within the canopy, due to a different phenological state. Other parameters (e.g. conductivities, diffusivities, roughness length) will vary in the same way. Heterogeneity of the parameters will be detailed in Sects. 5.2-5.3.

- Finally, the variable soil moisture itself cannot be considered homogeneous in a grid-box. Reasons include the variability in the forcings, parameter and fluxes as described above, but also non-accounted variability arising from, e.g., differences in soil type, terrain shape, vicinity of local features like an aquifer or a shallow-water table, etc. "Intrinsic" heterogeneity in the prognostic variable, e.g. that part of the variance that has not been taken into account by the above three items, will be detailed in Sect. 5.4.

### 5.1 HETEROGENEITY IN THE SURFACE COVER

Let us consider a grid-box with a fraction \((I-C_{\text{LSM}})\) covered by lakes. The land cover part of the grid-box will have a fraction covered by snow, \(C_{\text{Sl}}\). The snow free part will be covered by bare ground \((I-C_s)\), while a fraction \(C_i\) of the vegetation will be covered by intercepted water. The grid-box evaporation can therefore be written as a weighted sum of the expression of each different section (Abramopoulos et al., 1988; Noilhan and Planton, 1989; Pitman, 1991; Blondin, 1991; Duondré et al., 1993):

\[
E = \frac{C_{\text{LSM}} (C_{Sn} E_{Sn} + (1 - C_{Sn}) (C_i E_i + (1 - C_i) E_v) + (1 - C_v) E_g))}{(1 - C_{\text{LSM}}) E_{Sm}}
\]  

(31)

In the above expression, \(C_{\text{LSM}}\) and \(C_v\) are constant, geographically prescribed fractions (although the vegetation cover \(C_v\) can be allowed to vary seasonally), while \(C_{Sn}\) and \(C_i\) depend on the nature of the surface but also on the state of the soil, and therefore will vary in a forecast. Typically \(C_{Sn}\) will depend on the amount of snow and \(C_i\) on the water content of the interception reservoir. Lake evaporation, \(E_{Sm}\), evaporation of intercepted water, \(E_i\), and snow sublimation, \(E_{Sn}\), can be computed at the potential rate. Equation (19) can be used for dry canopy evaporation, \(E_v\), while \(E_g\) can be computed by any of the methods discussed in Sect. 3.3.2. Note that all the different evaporation rates will be computed in this approach by using a single value for the temperature and soil water, representative of the entire grid-box.

If the different heterogeneity elements corresponding to its fraction are distributed randomly in a grid-box, equation (1) gives a good approximation of the grid-box evaporation (type A heterogeneity, as detailed in Shuttleworth, 1988b). However, if there is organisation in the heterogeneity (e.g. a single rectangular lake occupying half a grid-box) mesoscale calculations can develop and transport a significant part of the heat and water involved (Avisar and Chen, 1993, type B heterogeneity in Shuttleworth, 1988b; see also Section 5.4).

The approach of equation (1) can be downscaled one step further. If there is more than one species present in the grid-box, the transpiration can be written as (Duondré et al, 1993; Avisar and Pielke, 1989; Li and Avisar, 1994)

where the summation is done across the \(N_v\) vegetation species, with cover \(C_{vr}\). For the computation of \(E_{vr}\), a different canopy resistance function can be applied to each species. In this way, each grid-box is schematically distributed in tiles (up to 7 different vegetation ecotypes in Duondré et al, 1993). As for Eq. (31), Eq. (32) is only valid if the different
\[ E_v = \sum_{i=1}^{N_v} C_{vi} E_{vi} \]  

(32)

micro-elements which conceptually aggregated constitute a tile are randomly distributed in the grid-box. The approach detailed in equations (31) and (32) is inexpensive, both in computational terms and in terms of memory. The only computational burden is to calculate the evaporation formula \( N \) times while carrying, in the case of equation (32), an additional file detailing the different ecosystems existing in a grid-box.

One step further in the line of complexity is to have different energy and water budgets for each of the separate tiles (or fractions, as represented in the previous two equations). The grid-box will be characterised not only by one temperature and soil water content, but \( N \) temperatures and \( N \) soil water contents, corresponding to the \( N \) different tiles (Avissar and Pielke, 1989; Li and Avissar, 1994). The cost in memory can be prohibitive for GCMs because it multiplies the number of surface prognostic fields by \( N \). Note that all the different surface tiles will have the same atmospheric forcing, namely, radiation, precipitation and near surface atmospheric variables.

5.2 PRESCRIBED SUBGRID SCALE DISTRIBUTION OF VARIABLES

Over a scale of 100 x 100 km, precipitation rate is far from uniform: the characteristic size of an individual convective element is of the order of 1-10 km, whereas its lifetime is of the order of the timestep used in GCM models (0.5-1 hour). During that time the individual convective element will be advected, hence the fraction of the 100 x 100 km square wetted by the cloud will be larger than the value in any snapshot at any given time. The actual precipitation intensity will follow some distribution law, with slope and amplitude dependent on the grid rise and the time interval considered (Entekhabi and Eagleson, 1989; Warrilow et al., 1986). The output of a convection scheme in a GCM (e.g. Tiedtke, 1989) will be a precipitation rate, corresponding to a mean rate in that timestep and for the grid box considered. The partitioning of that rain into interception, and subsequently into infiltration and run-off, will be markedly different if the rain is assumed to fall uniformly or if, say, it falls in a fraction \( k \) of the grid box with amplitude \( Pl/k \) (see Viterbo and Illari, 1994, for more details). The more concentrated in space the rainfall is, the higher the chances of reaching soil infiltration limits and therefore producing runoff.

Three distributions (probability density functions, pdf) of precipitation have been used in GCM modelling: a \( \gamma \)-distribution (Eagleson, 1978b; Entekhabi and Eagleson, 1989), an exponential distribution (Warrilow et al., 1986; Dolman and Gregory, 1993; Shuttleworth, 1988b; Pitman et al, 1990) and a box-distribution (Viterbo and Illari, 1994). All these distributions depend on parameters that define their sharpness; reliable estimation of these parameters based on observation events exists only for a handful of cases (e.g. Eagleson and Wang, 1985), with a larger uncertainty in the tropics because of the dearth of observational evidence. The sensitivity of the output of land-surface schemes (evaporation and runoff) to the value of the sharpness parameter was analysed: i) in stand-alone mode (i.e. running the land surface model uncoupled from the GCM) by Pitman et al (1990) and Pitman et al (1993) (see also review by Thomas and Henderson-Sellers, 1991) and ii) when fully coupled to a GCM (Viterbo and Illari, 1994). The sharpness parameter should be made dependent on
resolution (Eltahir and Bras, 1993): it depends on grid rise $\Delta x$, timestep $\Delta t$ and precipitation intensity $P$.

When a (pdf) $g(P)$ for precipitation exists, the fraction of the grid box with rainfall intensity between $P_i$ and $P_i + dP_i$ is $g(P_i)dP_i$, with $\int_0^\infty g(P)\ dP = P$. Applying Eq. (15) for the soil water balance over that fraction of the grid-box, the forcing is written as $g(P_i)dP_i$ and the result can be integrated across the range of precipitation intensities. The scheme integration can be made numerically (with techniques similar to the ones detailed in Avissar and Pielke, 1989) during the GCM integration, or analytically beforehand (Dolman and Gregory, 1993). The latter case will produce a modified equation (15); for simple distributions the modification made to the equations is trivial (Viterbo and Ilari, 1994).

The technique detailed above can be applied to any variable: A statistical-dynamic approach for considering heterogeneity is any approach where the variations in the input and the model state are quantified in terms of different probability density functions (pdf), and the output computed as the contribution of the forcing with the different model states, properly weighted by their own pdf functions. The runoff scheme of Dümenil and Todini (1992) presented in section 3.5 is another example where the infiltration capacity of the soil in a grid box is assumed to follow a distribution law. Further examples can be found for variations one parameter (Avissar, 1992) or in the state variables (Wetzel and Chang, 1988; Johnson et al, 1993). Avissar (1992) considers the subgrid scale variation of stomatal conductance. The surface energy budget equation is written for a given value of stomatal conductance. The appropriate soil water and heat equations are also computed for the given value of stomatal conductance. The system is valued for the ground surface temperature. The result is computed numerically for all values in the range of stomatal conductances and the solutions are averaged according to the pdf. Wetzel and Chang (1988) consider their statistical dynamical variations of both the soil water (with a pdf taken from observations) and of stomatal resistance. Averaged results are computed numerically by double integration across the soil moisture and the stomatal resistance range, and results are compared to observations. The traditional approach, considering only mean grid-square values, is shown to underestimate evaporation in cases of dry soil and underestimates it when the soil is close to field capacity. Johnson et al (1993) report on the testing within a GCM of a scheme where both runoff and evaporation used pdfs to characterise their dependence on precipitation and soil wetness. Their method generalises Dümenil and Todini's (1992) work. Finally, Mahrt (1987) computed the grid-averaged surface fluxes when the the static stability is not uniform in one grid-box.

5.3 EFFECTIVE PARAMETERS AND BLENDING HEIGHT

As mentioned before, the effect of subgrid scale heterogeneities depends on their length scale. For smaller length scales (less than 10 km) no organization is seen in the PBL and turbulence is responsible for the vertical transport of heat, moisture and momentum (Shuttleworth, 1988b). For the above conditions it is possible to define the blending height, the minimum height in the PBL where mean atmospheric conditions are approximately in equilibrium with the underlying surface (Wieringa, 1986; Mason, 1988; Claussen, 1991). As an example, let us consider the dry vegetated part of the canopy, fraction $C_a(1-C_d)$ in equation (31). Previously, we have already seen how to compute the average surface flux as a mean across the heterogeneity elements of terms proportional to the difference between an atmospheric term and the appropriate surface term. The atmospheric term has to be computed at (or
above) the so-called blending height.

A simpler alternative way of considering the heterogeneity in, e.g., the dry vegetated part of the canopy, is to use values of a mean surface state in the computation of the fluxes, but with modified parameters in the transfer coefficients (or surface resistances). The modified parameters that yield the correct value for the mean fluxes are called *effective parameters* (Fiedler and Panofsky, 1972; Mason, 1988; Warrilow and Buckley, 1989; Noilhan and Lacarrère, 1995). It is useful at this stage to distinguish between *primary* parameters and *secondary* parameters (Noilhan and Planton, 1989, Noilhan and Lacarrère, 1995). In the following, we will assume that a data set of the primary parameters exists at a finer resolution than the GCM resolution. The values of a few primary parameters (e.g. depth of soil, dominant type of soil texture, dominant type of vegetation) in a detailed data set determine the values of secondary parameters (roughness length, LAI, main stomatal resistance, fractional vegetation cover) at the resolution of the original data set. The problem consists of finding an effective value of each of the secondary parameters at the GCM resolution by stable averaging procedures. The averaging operator for each parameter is chosen for consistency with an arithmetic averaging of the fluxes themselves; in other words, for a given

![Fig. 7](image_url)

*Fig. 7* Comparison between the area-averaged surface fluxes computed from the 3D model (solid points) and the 1D prediction with dominant (dashed line) or effective (solid line) properties. The vertical bars represent the standard deviation of the fluxes in the 3D model. Latent heat flux (a), sensible heat flux (b), net radiation (c), soil heat flux (d), plant transpiration (e), and bare soil evaporation (f) (from Noilhan and Lacarrère, 1995).
parameter, the average should be linear in the related quantity that scales linearly with the fluxes. For instance the effective canopy conductance of a GCM grid box is defined as a weighted sum of the conductances of the elementary areas. For other examples see Warrilow and Buckley (1989) and Noilhan and Lacarrère (1995).

Fig. 7 (from Noilhan and Lacarrère, 1995) presents a comparison between the fluxes computed with a detailed mesoscale model and the fluxes computed using the effective parameter concept for a situation in the HAPEX-MOBILHY data base (André et al, 1986). Despite the non-linearity dependence of surface fluxes on vegetation parameters and soil water content, it is clear that the effective surface fluxes computed with a 1D model match the 3D mesoscale estimates with a relative error of ±10%. For the cases examined, the effects of non-linearity were found to be smaller for the vegetation parameters (e.g. stomatal conductance) than for the soil water transfer parameters (e.g. hydraulic diffusivity).

As pointed out by Noilhan and Lacarrère (1995) and Blyth et al (1993), the effective parameter approach does not work when i) mesoscale fluxes are of the same order of magnitude as turbulent fluxes (see next section); and ii) water, snow and ice surfaces are included in the average, because heat and evaporation fluxes are governed by quite different physical mechanisms. When one of those conditions exist, the fluxes should be aggregated using either Eq. (31), with a single surface temperature or different surface temperatures (Sect. 5.1).

5.4 ADDITIONAL REMARKS
In the previous sections we have seen how variations in the input of the model and its parameters imply spatial variations in the output; the above variations would, in principle, be present at the smaller spatial scales (e.g. below 1 km) and hamper even the mesoscale simulations that are taken as "truth" in most heterogeneity studies. Avissar (1991, 1992) calls variations in these spatial scales "heterogeneity" and argues that some of the variability in these scales might be "self-regulating" (Avissar, 1993). Studying the variations in stomatal resistance in an otherwise homogeneous potato field, he argues that a decrease in the stomatal resistance in a single leaf causes transpiration to increase and this, in turn, will humidify the surrounding air. As a result, the specific humidity difference for a neighbouring leaf increases and causes the stomatal resistance of those neighbouring leaves to increase, thus compensating for the transpiration increase of the first leaf. If "self-regulation" dominates, the sensitivity of quantities like stomatal resistance to environmental variables at the GCM scale is much less than its corresponding value at the plot scale.

For length scales of land surface variations larger than 10 km, mesoscale circulations are generated affecting the whole depth of the PBL (Avissar and Pielke, 1989; Bougeault et al, 1991). The concept of blending height is no longer valid (Clausen, 1993). It has been argued that these circulations are responsible for the vertical transport of sensible and latent heat of the same order to magnitude of the turbulent fluxes (Li and Avissar, 1994; Chen and Avissar, 1994; Segal et al, 1988). Avissar and Chen (1993) present the equations for the so-called "mesoscale kinetic energy", defined in a similar way to the turbulent kinetic energy (Stull, 1988), but where the averaging operator is limited to mesoscale lengths. The results obtained suggest mesoscale fluxes larger than turbulent fluxes are normally obtained in calm situations, whereas when there is some mean wind, turbulence takes over as the main mechanism of vertical transport. For the much more difficult problem of heat and moisture
vertical transport in stable cases and in the presence of topography see the review by Kaimal and Finnigan (1994).

6. Validation and intercomparison

6.1 POINT VALIDATION

Results of field experiments measuring surface fluxes and variables as local point values or on a regional scale are regularly used to test surface parametrization schemes. To give just a few examples, ARME (Shuttleworth et al, 1984) data over the Amazon basin have been used in validating SiB (Sellers et al, 1989), ISBA (Noilhan et al, 1993) and the ECMWF new surface model (Viterbo and Beljaars, 1995). FIFE (Sellers et al, 1988, 1992) data over the Konza Prairie in Kansas, US and Cabauw data in the Netherlands have been used to validate the ECMWF model (Bets et al, 1993; Beljaars and Viterbo, 1994) and, in the case of FIFE, to validate ISBA. HAPEX/MOBILHY data in Southern France have been used to test the ISBA model and the ECMWF model. SEBEX data, over the Sahelian region, have been used to validate the ECMWF model (Beljaars and Viterbo, 1995). For a review of recent experiments useful for validation of surface parametrization data, the reader is referred to Shuttleworth (1991). Noilhan et al. (1993) describes all the validation efforts for what is probably the most thoroughly validated surface parametrization model, ISBA. A summary of field experiments is presented in Table 2. It is clear from that table that many valuable experimental data sets have never been used to validate any large-scale surface model. Snow is very much underrepresented in the above table. Campaigns that have just finished for the Canadian boreal forest (BOREAS, Sellers et al, 1995) or are currently under way for the Scandinavian forest (NOPEX, 1994) will hopefully fill that gap in validation data.

In pointwise validation the surface model runs coupled to a PBL-model (1-column) or uncoupled (surface 1-column, or 0-column), forced by time series of observed values. The quality of the model can be assessed by comparing its output (e.g. latent and sensible heat, or soil moisture time series) with observed values. 0-column validation requires a time series of observed values of surface wind and near surface atmospheric temperature and humidity, solar radiation and downward longwave reduction (Henderson-Sellers et al., 1993). The advantages are: i) current automatic measurement stations allow the measurement of the above quantities during several months, therefore the model behaviour can be validated in its longer timescales; ii) since radiation is part of the forcing, mismatch between the model output and observations can be attributed to the surface model/surface flux formulations. On the other hand, because of the absence of negative feedback from the BL, model drifts are perhaps exaggerated in this validation model (Jacobs and de Bruin, 1992). Nevertheless, 0-column validation remains the only form of validation giving a clear message about the quality of the land-surface scheme. A major intercomparison exercise, PILPS (Henderson-Sellers et al, 1993) is currently under way, whereby most of the existent surface models is compared in 0-column mode, forced first by synthetic data sets and then by a time series of observed values. The first observed data sets chosen were Cabauw and HAPEX/MOBILHY. It is planned, in the next phase of PILPS, to couple each of the land-surface schemes with the same host GCM and compare their behaviour in true interactive model. A similar European intercomparison exercise, SLAPS, was designed to assess the quality of the land surface schemes when compared to catchment-scale models developed by hydrologists, and concentrates in the hydrological aspects of the land-surface schemes.
<table>
<thead>
<tr>
<th>Experiment</th>
<th>Reference</th>
<th>Location</th>
<th>Data</th>
<th>Model validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARME, 1983-85</td>
<td>Shuttleworth et al., 1984</td>
<td>Amazon basin</td>
<td>Surf. energy and water balance, soil</td>
<td>ISBA, SiB, ECMWF</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>wetness</td>
<td></td>
</tr>
<tr>
<td>SEBEX, 1989-90</td>
<td>Wallace et al., 1991, Dolman</td>
<td>Sahel</td>
<td>as above</td>
<td>ECMWF</td>
</tr>
<tr>
<td></td>
<td>et al., 1993</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAPEX-MOBILHY, 1986</td>
<td>André et al., 1986</td>
<td>Southwest France</td>
<td>as above, info on mesoscale variability</td>
<td>ISBA, ECMWF, PILPS</td>
</tr>
<tr>
<td>Cabauw, 1987</td>
<td>Beljaars and Viterbo, 1994</td>
<td>Netherlands</td>
<td>surf. energy and water balance</td>
<td>ECMWF, PILPS</td>
</tr>
<tr>
<td>La Crau, 1987</td>
<td></td>
<td>Marseille, France</td>
<td>as above</td>
<td></td>
</tr>
<tr>
<td>LOTREX, HIBE88, 1988</td>
<td></td>
<td>Germany</td>
<td>as above</td>
<td></td>
</tr>
<tr>
<td>Niger, 1988</td>
<td></td>
<td>northern Sahel</td>
<td>as above</td>
<td></td>
</tr>
<tr>
<td>HEIFE, 1990</td>
<td></td>
<td>Tibetan plateau, Gobi</td>
<td>as above</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>desert</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kurex-88, 1987-88</td>
<td></td>
<td>basin of river Seym</td>
<td>hydrological data, including runoff</td>
<td>ISBA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>and snow</td>
<td></td>
</tr>
<tr>
<td>EFEDA, 1991-95</td>
<td></td>
<td>Central Spain</td>
<td>surf. energy and water balance</td>
<td></td>
</tr>
<tr>
<td>HAPEX-SAHEL, 1992</td>
<td></td>
<td>Sahel</td>
<td>as above</td>
<td></td>
</tr>
<tr>
<td>BOREAS, 1991-95</td>
<td></td>
<td>Canada</td>
<td>as above, snow measurements</td>
<td></td>
</tr>
<tr>
<td>ABRACOS, 1991-95</td>
<td></td>
<td>Amazon basin</td>
<td>surf. energy and water balance</td>
<td></td>
</tr>
</tbody>
</table>

Table 2  Recent field experiments and their use for validation of land surface models.

When an estimate of the regional distribution of the fluxes is available (as in HAPEX/MOBILHY), another possible way of validation is to run a mesoscale model over the field experiment area. This type of modelling has been essential in validating the equivalent parameter concept (see previous section and Noilhan and Lacarrère, 1995) and is crucial to assess the importance of the organisation of fluxes in the mesoscale (Avissar and
Pielke, 1989).

It needs to be stressed that the validation effort has be be made in close cooperation between the experimentalists and the modellers, the former bringing their expertise in doing error analysis and suitable averaging of the data, while the latter bring the interpretation of model results and suggestions for improvement. Field experiments collect a very large volume of data (of the order of a few Gbytes, see Sellers et al, 1992) that has to be aggregated into a much smaller set of values that will be the forcing data set for the surface model, and another set that will serve to validate the results (each set being no more than a few Kbytes). This is by no means a trivial exercise (Betts et al, 1993; Beljaars and Viterbo, 1995) but, once it has been done for use with one model, it is readily available for validating the other model.

One final remark. Field experiments are normally organised with concentrated measuring efforts on a few days to few weeks, the so-called Intensive Observation Periods (IOP), with gaps for some quantities, such as evaporation between two consecutive IOPs. Due to the different timescales involved in soil moisture (ranging from the diurnal to the seasonal scale), it would be desirable to have, in future experimental efforts, continuous monitoring of evaporation, radiation and sensible heat fluxes, together with soil moisture in the root zone, at least for an entire season. In that way, a closed soil water budget can be performed for the observations.

6.2 OTHER FORMS OF VALIDATION
Forecast/assimilation systems have the infrastructure to monitor the forecast results against observations (Strauss and Lanzinger, 1993; Lanzinger, this volume). Near-surface atmospheric variables routinely compared with the plentiful SYNOP observations include 2 metre temperature and humidity, low-level cloudiness and precipitation. As detailed in Lanzinger (this volume), the interpretation of the data displayed gives important clues on model problems, and the signature of model/analysis changes is often found in a long time series of data.

A set of global, continental and regional scale data sets that can be used to validate results from GCM climate runs or to monitor the performance of NWP assimilation/forecast systems is presented in Table 3. In this type of validation, it is not always obvious to link errors in the variables to deficiencies in one specific parametrization scheme. For instance, feedbacks between model processes can be responsible for model deficiencies in a variable like precipitation (Arpe, 1991). Some of the data sets in table 3 are estimated: for instance latent heat fluxes and the other components of the surface energy balance are estimated based on the extent of empirical formulas and energy conservation principles. These methods, although making the data set self-consistent in energetic terms, cast some doubt on their validity for verifying model results.

To give another example, runoff data has been used to validate its model counterpart for major rivers (Russell and Miller, 1990; Dümenil and Todini, 1992). Most of the time problems in model runoff are related to deficiencies in model precipitation, rather than the details of model surface drainage or infiltration. Separation of errors in forcing (the precipitation field) from the errors in the surface model is a difficult task (Miller et al, 1994).
<table>
<thead>
<tr>
<th>Dataset</th>
<th>Parameter</th>
<th>(Measured/Estimated)/Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jaeger, 1983</td>
<td>Rainfall</td>
<td>M / global</td>
</tr>
<tr>
<td>Legates and Willmott, 1990</td>
<td>Rainfall</td>
<td>M / global</td>
</tr>
<tr>
<td>UNESCO, 1974</td>
<td>Runoff</td>
<td>M / global</td>
</tr>
<tr>
<td>Henning, 1989</td>
<td>Energy balance components</td>
<td>E / global</td>
</tr>
<tr>
<td>Wallis et al., 1991</td>
<td>Precipitation, runoff, temp.</td>
<td>M / US</td>
</tr>
<tr>
<td>Hollinger and Isard, 1994</td>
<td>Soil moisture</td>
<td>M/Illinois, US</td>
</tr>
<tr>
<td>Vinnikov and Yenserkepova, 1991</td>
<td>Soil moisture</td>
<td>M / former USSR</td>
</tr>
</tbody>
</table>

Table 3  Global data sets that can be used to validate surface-related GCM results.

7. Initial values
Despite the sensitivity of medium-range weather forecast to initial conditions of soil water (see reviews by Garratt, 1993 and Rowntree, this volume), there are at present very few methods to define the soil water content in data assimilation systems. Current techniques for ground-based measurement of soil water content (reviewed in Hillel, 1982; Cuenca and Noilhan, 1991; Schulin et al., 1992) include the gravimetric method, neutron scattering, electromagnetic techniques, and tensiometer. None of the above methods is adequate for routine measurements, and therefore there has been an effort, in recent years, to provide satellite based estimates of soil water content (see reviews in Choudury, 1991; Schmugge and Becker, 1992). The techniques can use information from the thermal infra-red channels, or from active or passive microwave systems, but they have several calibration problems and provide only estimates for the water contents in the top few cms of the soil. We can safely state that none of the current methods of measurement of soil water can provide a weekly global estimate of the water soil contents in the root zone.

Almost all schemes for initialisation of soil water in NWP are based on finding the equilibrium value of soil moisture, given climatological estimates of sensible and latent heat fluxes, and radiative fluxes at the surface (see e.g. Mintz and Serafini, 1992). They are therefore inappropriate for use with data assimilation schemes, where the goal is to find a soil water field representing an adequate balance of real-time estimates of the above fluxes. Many current NWP prediction systems circumvent the problem by assigning short-term forecast values to the initial conditions of soil moisture. Any deficiencies in the land-surface model or, more seriously, in the forcing terms (precipitation and net radiation) will cause the model to drift in time, and these short-term forecast values will eventually be affected by biases, corresponding to the climatic bias of the land-surface model or the near surface atmospheric forcing.
The only methodology that can be applied in a forecast/data assimilation system is based on the ideas of Mahfouf (1991). As mentioned in the previous section, the error of short-range forecasts of summer-time near-surface temperature and dewpoint, when compared to the plentiful SYNOP observations (of the order of 10000 for any synoptic time), is normally a good indicator of the quality of the soil water field; in broad terms, too warm and too dry near surface atmospheric model states during daytime are associated with a wrong partitioning of the available surface radiative energy in latent and sensible heat flux, and too dry values for the soil wetness. Mahfouf (1991) developed an optimal interpolation scheme for initialisation of soil water, relating the analysis increments of soil moisture to short-range forecast errors of near-surface temperature and dew point.

As with any data assimilation scheme (Daley, 1991) used for medium-range forecasts, it is more important to initialise correctly the variables in the system associated with a longer memory, larger timescales. Variables with a sub-diurnal timescale will adjust their initial values to some values compatible with the model physics; the adjustment process will take just a few hours and, beyond this adjustment period, the "memory" of the initial conditions is lost. For soil moisture, this means having a special concern about the deeper layers, other than the shallow top layer, because the former have a timescale of weeks, as compared to the diurnal timescale of the latter. Notice that the deeper layers interact still with the atmosphere via transpiration, transporting the water from the root layer to the atmosphere.

Mahfouf (1991) developed an optimal interpolation scheme (Daley, 1991) for initialising the soil water, relating the analysis increments of soil moisture to short-range forecast errors of near-surface temperature and dewpoint. The scheme has been developed further and applied to initialisation of soil water in a mesoscale model (Bouttier et al, 1993b). As shown by onedimensional sensitivity and numerical simulation studies, in bare ground areas the error in two-metre temperature and dewpoint is associated to error in the top soil water layer, while in vegetated areas it is related to the root layer soil moisture errors. The most critical aspect of the algorithm is the definition of the optimum coefficients in the matrix relating the errors in two-metre temperature and dewpoint to the analysis increment errors in the top and root layer soil moisture. A continuous parametric formulation described in Bouttier et al (1993a) allows for the computation of these coefficients at each model grid point. They will depend upon solar zenith angle and surface characteristics (vegetation coverage, roughness length and soil texture).

A 48-hour clear-sky period from the HAPEX-MOBILHY experiment (André et al, 1986) was studied by Bouttier et al (1993b) and is presented in Fig. 8. The results of the assimilation demonstrate the rapid convergence of the method when starting from a wet or dry soil moisture guess.

Recently, a simplified version of the Mahfouf (1991) algorithm was introduced in operations at ECMWF (Viterbo and Courtier, 1995). It corrects for a dry and warm bias in the surface and boundary layer forecasts in late spring and early summer. The bias is associated to an underprediction of cloud cover causing too much solar radiation at the surface, driving too large an evaporation rate, and drying the soil too quickly and too early. The soil water initialisation scheme prevents the soil water values from drying too quickly. Fig. 9, comparing the root mean square (rms) error for 20 forecasts run with (moist) and without (ops) the soil
Evolution of soil water in 48-hour data assimilation with a mesoscale model. Top row: Starting condition dry. Bottom row: Starting condition wet. Values are shown for top soil water (left) and root soil water (right). Solid curves (A): mean values over the area; Dashed curves: maximum and minimum values; Solid curve, unlabelled: reference simulation, representing the truth (from Bouttier et al, 1993b).

Fig. 8

Z 200 N. AMERICA 1/6/94 - 20/6/94 (20 cases)
Difference moist - ops (95% conf. level estimation)

Mean root mean square of geopotential height for control (ops) and test (moist), and mean difference and confidence interval at 95% level, 1-20 June 1994, 20 kPa North America.

Fig. 9
water initialisation scheme, shows the large impact of the initial values scheme in the forecast quality. The impact on the rms comes mainly from a reduction in the lower tropospheric bias of temperature: a wetter surface causes a colder lower troposphere and, due to the hydrostatic relationship, a shallower 20kPa model surface. The large sensitivity of summertime continental forecasts to the definition of soil water initial conditions confirms the results reviewed in Garratt (1993).

8. Snow modelling
The effects of the snow mantle have to be taken into account for appropriate consideration of the surface thermal balance on high latitudes and mountainous regions in GCMs. The presence of snow reduces the energy available at the surface: the albedo of fresh snow is 0.85, while the albedo of a natural surface is in the range 0.12 to 0.25 (Dickinson, 1988). Snow melting is the most important source of soil moisture in spring in high latitudes. In general, due to its thermal properties, snow acts as an insulator between the air above and the soil underneath (Peixoto and Oort, 1991; Walsh et al., 1985). The thermal and mass budget of a layer of snow lying on the ground is relatively easy to establish on a local scale (e.g. Anderson, 1976), but more complex at scales of a typical GCM grid box, due to the heterogeneity of snow cover. In spite of its importance, there is very little observational evidence of relevance to GCMs on typical melting rates, albedo, snow cover and metamorphic changes followed by the snow mantle.

Most, if not all, current GCMs carry a prognostic equation for snow mass (Manabe, 1969)

\[ \rho \frac{dS}{dt} = F - M \]  

(33)

where \( S \) is the snow depth (m of water equivalent), \( F \) is the snow fall (Kg m\(^{-2}\)s\(^{-1}\)) and \( M \) is the rate of melting. The snowfall is either given as an independent amount from other physical parametrizations in the model, or is the total amount of precipitation if the surface air temperature is below a certain threshold. The role of snow parametrization schemes is to specify the melting rate, \( M \), and the snow albedo entering the thermal budget equation.

Melting conditions are met when the equilibrium ground temperature is above 0°C. In that case, an adjustment to 0°C is made by melting the necessary amount of snow, and the upper soil reservoir collects the melted water. The melted water exceeding the maximum capacity of that reservoir is lost into runoff. Albedo in snow covered areas is modelled as a background value plus a correction dependent on the snow amount. Some models take into account snow "masking" by the vegetation in the computation of the albedo, to reduce the snow covered area in the presence of tall vegetation, when compared to bare ground terrain (Blondin, 1991). Snow contribution to the albedo can also be made dependent on temperature: snow under melting conditions is made darker to simulate the effect of surface ponding (Dickinson et al., 1986).

Snow cover fraction (the fraction of the grid box covered by snow) is important for its effects on the albedo and melting. Inspection of any satellite image reveals that snow cover fraction is essentially dependent on vegetation, and typographic details such as slope and aspect. In GCMs snow cover fractions normally depend linearly on the amount of snow up to a
threshold value, beyond which it is taken as 1. The threshold value can depend on roughness length, in a crude way of parametrizing orography and vegetation effects.

More complex snow models are normally introduced by carrying extra predictive variables: snow density (Pitman et al., 1991; Verseghy, 1991; Douville et al., 1995), snow temperature (Verseghy, 1991; Dümenil, private communication), snow albedo (Douville et al., 1995). Details of the snow pack metamorphism, e.g., distinguishing between coarse and finer grain, or between old dark snow and finer fresh snow, can be considered by introducing a snow age (time elapsed since last snowfall, Verseghy, 1991) dependency on the density and on the albedo. For its independent thermal budget, the snow pack is considered as an additional variable-depth layer, with thermal conductivity and heat capacity dependent on snow density.

Melting of the snow pack can occur in these more complex models in two different ways: surface melting and deep melting. If the surface energy balance equation gives a temperature above 0°C, melting of snow occurs at the expense of the excess of energy obtained by cooling the surface to 0°C. The resulting amount of water percolates into the snow layer and might refreeze within the snow pack at some unspecified depth, in a process called ripening of the snow pack. The remaining water wets the upper layer of soil. Deep melting occurs by heat conduction from underneath the snow pack, if the soil is above 0°C. There is a similar adjustment of temperature as in the surface melting, but no ripening is allowed, the water being immediately available to the soil layer. Note that a separate thermal budget of the snow layer is necessary for proper separation of the two melting mechanisms and the ripening of the snow pack.

As referred earlier in Sect. 3, phase changes of the water in the soil is another important mechanism in high-latitudes (Black and Tice, 1988, Williams and Smith, 1992, Miller, 1980). A parametric inclusion of the effects of the solid phase of water, although essential for modelling the soil water and energy transfer in high latitudes, is not considered in most GCM models. Its is possible to write additional equations for the conservation of frozen water at different soil layers (Verseghy, 1991, Pitman et al., 1991). Modifications to the traditional treatment include, in order of importance: i) The thermal effects related to the latent heat of fusion/freezing; ii) Substantial reduction in transpiration in the presence of a frozen ground; iii) Soil water transfer dependent on a soil water potential including the effect of frozen water. There are indications that these effects are very important for characterising the role of boreal forests in the climate system (Sellers et al., 1995).

**9. Conclusions**

Current atmospheric models play a major role in estimating the surface branch of the hydrological cycle. The review presented above tried to emphasize the special role of NWP forecast/data assimilation systems as a continuously operating numerical laboratory for the study of the interactions of the atmosphere with the underlying surface. The only practical way of estimating globally the geographical distribution of the different terms of the surface hydrological budget on a day to day basis is to use data assimilation within a global forecasting system. In this systems, conventional synoptic data plus satellite observations are combined with a very short range forecast to produce an analyzed state of the atmosphere. Short range integrations starting from these initial states can provide an internally consistent (although model dependent) picture of precipitation, evaporation and runoff. Since the hydrologic formulation of current atmospheric models is a key component of the process
described above, it is important to review the current state-of-the-art land surface models, and to identify the deficiencies and areas where current research is likely to bring substantial improvements.

However, operational data assimilation systems do not provide a time-homogeneous sequence of values, because of the regular changes in either the data assimilation methods or in the forecast model used. Alternative estimates, based on multi-year climate integrations of the model, are hampered by the systematic errors in the models: they do not benefit from the corrective influence of the observations, regularly fed in the data assimilation cycles. In order to combine the benefit of a long period series of simulation given by the climate integrations with the controlling effect of the observations given by the data assimilation, several centres started recently re-analysing the atmosphere with a frozen system (Bengtsson and Shukla, 1988). The ECMWF Re-Analysis Project is currently re-analysing the atmosphere at T106 31 levels, for the years 1979-1993. The project will be completed in 1996, and the examination of results related to the surface-atmosphere interaction will provide invaluable insight on mechanisms involved in different timescales, ranging from the diurnal cycle to the seasonal cycle.

The last 20 years were characterised by a wide acceptance in the GCM community of the role of vegetation in controlling evaporation. The PILPS project is currently the catalyst in the only way ahead to develop and improve parametrizations: validation and comparison with observations. There is much to learn on the longer time scales of the atmosphere related to soil water contents, the complicated interaction between soil water and precipitation, and the role of snow in the climate system.

References


Black, P.B., and A.R. Tice, 1988: Comparison of soil freezing curve and soil water curve data for
Viterbo, P.: A REVIEW OF PARAMETRIZATION SCHEMES...


VITERBO, P.: A REVIEW OF PARAMETRIZATION SCHEMES...


Manabe, S., 1969: Climate and the ocean circulation, I. The atmospheric circulation and the hydrology...


Pitman, A.J., Z.-L. Yang, and A. Henderson-Sellers, 1993: Sub-grid scale precipitation in AGCMs:


Sellers, P.J., 1985: Canopy reflectance, photosynthesis and transpiration. Int. J. Remote Sensing, 6,
1335-1372.


Viterbo, P., and A.C.M. Beljaars, 1995: An improved land surface parametrization scheme in the ECMWF model and its validation. Accepted for publication in *J. Climate*.


