4 Water Budgets in Ecosystems

John W. Pomery, Matthew K. MacDonald, Pablo F. Dornes, and Robert Armstrong

4.1 Introduction

Water is the foundation of all ecosystems, whether terrestrial or aquatic. In terrestrial ecosystems freshwater not only provides critical water supply for transpiration during plant photosynthesis and drinking water for animals, but also transports, redistributes and stores energy, nutrients and contaminants. In aquatic and snow ecosystems, water is the medium in which the ecosystem functions and so its state mediates all transactions in these systems. Ecosystems are not passive responders to water but through their structure and function can manage water and associated microclimate – forests, grasslands, organic terrain wetlands, and beaver ponds being just a few examples.

This chapter will examine the surface water budget in terms of the water continuity equation as a manifestation of the hydrological cycle. To solve the continuity equation for water, the chapter will review hydrological processes and how they interact with vegetation, animals, soils, geomorphology and climate in the context of the catchment. The coupling of the mass and energy continuity equations in controlling hydrological processes will be discussed. How hydrological processes and their ecosystem interactions are managed by humans will be introduced. Then the chapter will review calculation schemes for the surface water budget via one-dimensional land surface schemes and catchment-based hydrological models, noting the data requirements, uncertainty and limitations of these models and the balance required between model complexity and physical representation of hydrology. This will give the conceptual ideas and basic mathematics of conservation laws and transport processes that form the basis of many models in the forthcoming chapters.

4.2 Hydrological Processes as a Fundamental Component of Aquatic and Terrestrial Ecosystems

4.2.1 Hydrological Cycle

The hydrological cycle is the flow and storage of water, as liquid, solid or vapor, on and near the Earth’s surface. This cycling is a fundamental function of the Earth system and, through its associated latent energy transformations and other influences on land surface characteristics, ensures the habitability of the planet. A representation of the global

hydrological cycle is found in Figure 4.1 where it can be seen that there are substantial flows between ocean and land – evaporation and river discharge from land transfer water directly to the oceans or through precipitation and ocean water is evaporated and then forms precipitation over land. In general, precipitation, evaporation, infiltration, soil moisture and percolation are vertical (1D) processes whilst runoff, groundwater flow and river discharge are horizontal (2D) processes; therefore, land surface and soil-vegetation-atmosphere transfer schemes have tended to focus on the 1D processes in order to solve for evaporation and surface energy balance whilst hydrological models have tended to focus on 2D processes in order to estimate river discharge and lake storage.

Water budgets of terrestrial and aquatic ecosystems require a focus on the terrestrial hydrological cycle where key hydrological processes include:

– Precipitation of snowfall and rainfall: the fall of snow and rain to the surface, taking into account preferential deposition of snowfall to leeeward slopes in mountain terrain and of rainfall to windward slopes under strong winds in complex terrain. Precipitation can be derived from convective or frontal storms and can be driven by orography in mountains (Bastia et al., 1994), lake effects near water bodies (Norton and Bolsenga, 1993) and recycling where evaportranspiration rates are high (Eliahir and Bras, 1994).
Water Budgets in Ecosystems

Hydrological Processes of Aquatic and Terrestrial Ecosystems

- Redistribution of snow by wind: the transport of blowing snow by wind occurs as two-phase flow erodes snow from exposed low vegetation surfaces to sheltered, high vegetation surfaces. There is concomitant sublimation of blowing snow that reduces the amount that is deposited in drifts (Pomeroy et al., 1993).

- Interception of rainfall and snowfall by plant canopies: some rain drops and falling snowflakes are intercepted and stored in plant canopies where they are subject to evaporation and sublimation and also redistributed to the surface as drip, stemflow, and unloading. The efficiency of interception is controlled by both the intensity and phase of precipitation and the structure of the plant canopy, rigidity of branches and shape of leaves (Rutter et al., 1971; 1975; Pomeroy et al., 1998).

- Evaporation from water and soils: evaporation occurs directly from open water surfaces (lakes, ponds, wetlands, puddles) and from soils via exfiltration. Over open water, the evaporation rate is controlled by wind speed and humidity gradients with the atmosphere and is poorly indexed to net radiation due to substantial heat storage in water bodies and the influence of upwind stability (Granger and Hedstrom, 2010). Over soils, evaporation rates are controlled by energy and ventilation at the surface, tension forces in the soil, soil moisture content and permeability, including the effect of macropores and cracks (Milly, 1984). In the vadose zone, it is controlled by the rate of capillary rise from saturated porous media.

- Evapotranspiration from plant stomata: evapotranspiration is controlled by the plant's ability to withdraw water from the soil to the roots, transmit water upwards through plant stems and then release water from stomata as transpiration during photosynthesis. It therefore requires daylight and photosynthesis, adequate soil moisture, and is strongly associated with plant primary productivity (Monteith, 1965; Rosenzweig, 1968).

- Sublimation from snow and ice: sublimation is the vaporization of ice from snowpack, lake/river ice or glaciers and is strongly controlled by vapor pressure gradient between the ice and atmosphere and ventilation. It can also occur from blowing snow particles and snow intercepted in forest canopies and tends to operate most effectively when ice temperatures are below 0°C (Strasser et al., 2008).

- Melt of snow and ice: melt of ice occurs from snowpacks, lake/river ice or glaciers and involves phase change from solid to liquid. It occurs when the temperature of the ice has reached 0°C and further energy input goes into phase change (Male and Gray, 1981).

- Freezing of liquid water on the surface and in frozen soils: meltwater can refreeze at the base of a melting snowpack when the substrate is a frozen soil or cold glacier surface. This basal ice or superimposed ice can be an important water storage term in extremely cold regions. Meltwater can also refreeze after infiltrating a frozen soil, where its release of latent heat has an important role in warming the soil (Gray et al., 2001).

- Flow of glacier ice: glacial ice is a deformable, plastic material that flows downhill. Higher elevations of glaciers tend to have greater snow accumulation than melt, resulting in a positive local mass balance. Over time the snowpack metamorphoses into glacier ice and flows downhill to lower elevations where melt exceeds accumulation and the local mass balance is negative (Paterson, 1994).

- Infiltration of water into soils: infiltration occurs by direct entry of water into air-filled pore spaces in the soil and via movement into macropores and cracks in the soil. Downward percolation rates for flow from ponded water into an initially unsaturated soil are controlled by the hydraulic conductivity of the soil, which is a function of soil texture. Much more rapid infiltration occurs when macropores are present (Green and Ampt, 1911; Germann and Beven, 1985).

- Percolation of subsurface water: percolation involves water moving through the non-saturated zone. Infiltrated water replenishes the soil moisture deficiency and the excess moving downward by the force of gravity is called percolation. If it reaches the saturated zone and contributes to the ground water table it constitutes the process of recharge. Vertical unsaturated flow with the conservation of mass is described by the Richards equation, which evaluates the vertical rate of change of the hydraulic conductivity as a function of soil water content and the vertical rate of change of the pressure head or the matric suction gradient, also being a function of soil water content.

- Capillary rise of water in porous media: capillary rise transports water into the unsaturated zone above the water table as a result of the upward movement of capillary water. The thickness of the capillary fringe varies inversely with pore size of a soil or rock.

- Saturated flow over porous media (groundwater flow, interflow): infiltrated water may percolate deeper than the unsaturated zone to recharge groundwater and later become hillslope flow, which may seep into streams. Thus, the transfer of surface water to streamflow is the result of saturated flow in porous media (i.e., aquifer discharge) and is governed by the Darcy equation.

- Macropore subsurface flow: this preferential flow is a result of a variety of factors such as flow through non-capillary cracks or channels within a soil profile, or by water flowing downhill as interflow through highly permeable or fractured materials above the water table.

- Overland flow: this is water flowing over the ground surface toward a stream. It is usually an episodic process closely related to precipitation rate and soil maximum infiltration rates. According to the generation process, it can be related to precipitation characteristics (Hortonian flow) or to flow over existing saturated soils.

- Channel flow: this is water flowing in defined river channels as either open channel flow or flow under river ice. Hydraulic relationships between channel wall friction, slope and channel width control flow rates and discharge. Ice cover complicates these relationships and can result in restricted flow.

These processes operate via flow dynamics with water stores, such as depressional storage (lakes, ponds, wetlands), interception storage (canopy rain or snow load), detention storage (rivers, overland flow), snowpacks, glaciers, soil moisture and groundwater. A more detailed representation of terrestrial hydrological processes
and stores is shown in Figure 4.2. Some of the processes are directly linked to and controlled by the ecosystem, for instance plant canopies and interception and evapotranspiration.

Hydrological processes can be used to calculate the terms of the continuity equation to estimate the water budget of any particular store of water of interest on the Earth's surface. For terrestrial ecosystems, the term of interest is usually soil moisture due to the control of soil moisture on plant transpiration and photosynthesis. For aquatic ecosystems, the term is usually surface depressional storage in lakes, ponds or wetlands and river discharge.

In the simplest terms the continuity equation applied to hydrology can be found as:

\[ P - E = \Delta S + R \]  

where \( P \) is the precipitation rate, \( E \) is the evaporation rate (including sublimation and evapotranspiration), \( S \) is water storage (surface and subsurface), \( \Delta \) denotes the change over a unit time, and \( R \) is river discharge rate (runoff) – all terms are mass of water per unit area of the surface. Equation 4.1 can be applied to a control volume as shown in Figure 4.3, but it is very scale dependent. For long time intervals of years and large spatial scales, \( \Delta S \) tends toward 0. However, at smaller temporal or spatial scales, the change in storage term can be very large. It is strongly mediated by interactions with vegetation and animals and has been managed by humanity for many thousands of years.

Whilst attractively and perhaps deceptively simple, Equation 4.1 cannot be directly applied in practice because the terms are aggregations of various hydrological processes that have varying degrees of importance in different climates and ecosystems and for different spatial scales. There is also strong seasonality, episodic behavior and temporal, and spatial coherence in various terms. For instance, when and where both precipitation and evaporation are large together, there may be little response in storage and runoff. This can occur when light rainfall events occur relatively frequently and evenly over level topography such as plains. But with large precipitation rates occurring when or where evaporation is small, such as in winter when snowfall occurs or during heavy rainstorms or on mountain tops, the storage and runoff terms respond dramatically. Similarly, ecosystems that have the potential to store water as snowpacks, glaciers, interception, soil moisture, groundwater or surface water will have attenuated or delayed runoff responses to precipitation events. Large storage terms tend to evaporate leading to attenuated runoff or may be delayed as storage in snow and ice, which must melt to form runoff, or in slowly draining sub-surface water that may form runoff days or even months later. Not all storage need form runoff – there are internally drained lakes, ponds and wetlands that will store water until it evaporates or contribute to deeper infiltration and never form river discharge. These are common in arid, semi-arid and previously glaciated environments where the connectivity to the stream channel from the drainage basin is very poor or non-existent.
To actually apply Equation 4.1 it must be broken down into hydrological process terms and the temporal and spatial scale of the control volume for application of continuity must be clearly defined. Breaking down the term for precipitation shows that:

$$ P = P_n^* + P_R^* $$ \hspace{1cm} 4.2

where the net snowfall rate, $P_n^*$ is composed of the snowfall, $P_n$, plus the flux of snow redistributed horizontally by wind or vegetation, $Q$, in and out of the control volume:

$$ P_n^* = P_n + (Q_{in} - Q_{out}) $$ \hspace{1cm} 4.3

Snow redistribution by wind must be calculated as blowing snow using two-phase flow physics and is important in the arctic, alpine, glacierized and cold regions steppe environments. Snow can be redistributed long distances from sparsely vegetated surfaces to lee slopes, topographic depressions, treelines, shrub-tundra and tall steppe or prairie vegetation such as tall grasses, shrubs or trees. There is also snow redistribution from snow intercepted in forest canopies into nearby gaps or clearings, but this term is usually considered small. A conceptual diagram of snow redistribution processes and their interaction with vegetation is shown in Figure 4.4.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{diagram.png}
\caption{Case for snow redistribution in cold regions: (a) forest with clearings. (b) alpine and treetline (Harder and Pomeroy, 2014). Reproduced with permission of the publisher.}
\end{figure}

The net rainfall rate, $P_R^*$ is the rainfall rate, $P_R$ (including direct fog deposition) plus drip in from adjacent plant canopies, $D_m$, and stemflow in from adjacent plant canopies, $SF_m$, where adjacent canopies are outside of the control volume of interest:

$$ P_R^* = P_R + D_m + SF_m $$ \hspace{1cm} 4.4

The difficulty in this conceptualization of the continuity equation is that it considers interception in the plant canopy, $I$, as part of the change in storage term $\Delta S$, and interception losses to evaporation and sublimation as part of the evaporation term. Only the stemflow and drip terms are included in net rainfall. As such, the net rainfall rate defined here is not that which would be measured under a plant canopy as it does not have the evaporative losses taken from it yet. However, it is the above canopy rainfall rate plus any small-scale redistribution of rainfall by the plant canopy via drip and stemflow.

Breaking down the term for evaporation shows surprising complexity, more of the interception terms and substantial involvement with the terrestrial ecosystem via transpiration, direct evaporation and interception:

$$ E = T + E_o + E_{ex} + I_e + I_{so} + Q_{so} + S_b $$ \hspace{1cm} 4.5

where, $T$ is the transpiration rate from plant stomata, $E_o$ is the rate of evaporation from open water at the surface (lakes, ponds, puddles, wetland surfaces, rivers), $E_{ex}$ is the evaporation of infiltrated water from the soil (direct soil evaporation), $I_e$ is the evaporation of intercepted rainfall directly from the plant canopy, $I_{so}$ is the sublimation of intercepted snowfall directly from the plant canopy, $Q_{so}$ is sublimation of blowing snow and $S_b$ is direct sublimation from the surface snowpack or glacier ice surface. Note that evaporation and sublimation rates can be negative when condensation is occurring as a consequence of phase change from water vapor to either liquid or solid – this is not direct fog droplet deposition, which is part of the precipitation term. A conceptual diagram of these processes and how they interact with the surface water budget is shown in Figure 4.5; this is an example from a mountain basin in Patagonia with the local annual importance of various fluxes. Transpiration from the plant stomata depends on photosynthesis, which depends on available energy, temperature, soil moisture, nutrients, plant species, growth state and plant health. Transpiration tends to only occur during daylight and occurs most rapidly from healthy plants in a growth phase with adequate soil moisture, warm temperatures, exposure to direct solar radiation, good nutrient supply for photosynthesis and well-developed root systems. Transpiration can be enhanced by water vapor deficits in the atmosphere and strong winds if species are not adapted to arid conditions. Lower rates of transpiration occur during cold and/or damp conditions, from arid-zone species, non-vascular plants, unhealthy plants (infected by disease, parasites or impacted by insects or grazing), plants with poorly developed root systems or inadequate soil moisture availability for the roots. Transpiration in some ecosystems can be modulated by all of rooting zone soil water supply, solar radiation, air temperature, soil temperature, season and nutrient supply, making estimation of transpiration rates very uncertain. Open water, soil water and snowpack evaporation, and sublimation rates tend to be controlled by atmospheric
accessed by plant roots and the groundwater store, \(GW\), in normally saturated, typically deep porous media layers defined as confined or unconfined aquifers:

\[
S = G + SWE + I + D + D_T + O + S_M + GW
\]

The glacier and snowpack storage terms are controlled by the accumulation of snowfall and redistributed snow, the sublimation of snow and the melt of snow. These storage terms can be important interannually as glaciers or seasonally as snowpack and can be the largest storage terms in cold regions. There is increasing interest in these multiphase environments as habitats for microbial life, and in the seasonal snowpack a wide range of organisms can be found from microbial to fungal to sub-nival plants, herbivores and carnivores including large mammals and a food web (Jones et al., 2001). The interception of rainfall and snowfall in plant canopies is a very short-term store (minutes for rainfall, up to months for snowfall) that provides water for some lichens and mosses and stores water in a high energy well-ventilated environment where evaporation or sublimation is likely. Any interception that does not evaporate or sublimates forms drip and streamflow or unloads as redistributed snow.

Depressional storage is a tremendously important term as it includes most surface water storage, and some stores are very large such as inland great lakes. There is a continuum of sizes from lakes that are considered inland seas to small ephemeral puddles. These are incredibly important habitats for aquatic ecosystems – from the full range of microbial, algal, plant, invertebrate, fish, insects, birds and mammals that inhabit freshwater lakes, permanent ponds and wetlands to sparser ecosystems in muskegs, ephemeral ponds, saline lakes and puddles. But they are also important for terrestrial ecosystems as a source of drinking water for non-aquatic insects, birds and mammals or temporary habitat for some organisms that live both on land and water. Not only are the depth, spatial extent, connectivity and fluctuation of these water bodies important to the ecosystem, but also their temperature, sediment load, dissolved oxygen concentration, organic and inorganic chemistry and optical transparency are critically important to life. The formation of seasonal ice covers is also very important for the ecosystem.

Surface water and shallow sub-surface storage have a strong control on hydrology as well-defined storage, \(D\), and river discharge, \(R\), relationships are established for water bodies that are connected to the stream channel network:

\[
R = a \cdot D^b
\]

where \(a\) and \(b\) are coefficients. This linear reservoir relationship presuming a stream channel flowing downhill from the water body, in which the discharge rate is governed by the hydraulic head gradient and therefore the storage level of water in the lake, pond or wetland. Well-defined storage-discharge relationships occur where fluvial erosion has shaped geomorphology such that a dense dendritic network of stream channels connects water bodies. This is common in most temperate rivers basins. However, some surface water is stored in internally drained depressions or in depressions not connected to any stream except by fill and spill runoff mechanisms in which, when the storage exceeds some threshold, surface runoff can connect to a downstream channel. This is
more common in regions where geomorphology is defined by exposed bedrock, aeolian erosion or by glacial processes from recent (<20,000 years) glaciation. Detention storage is defined as moving surface water and so is found from the depth of overland flow and river discharge on the land surface at any one time. Overland flow is very episodic and nearly non-existent in many environments except during heavy rainfalls or rapid snowmelt events, but the storage of detained water in large rivers can be significant in some environments. Organic layer and soil moisture storage can be difficult to distinguish as there is a continuum between poorly decomposed organic deposits such as the leaf litter and duff layer and mineral soils with a high organic content. Importantly, organic layers do not and soils do contain roots or vascular plants where uptake of water from the pore space to the plant for transpiration occurs. However, they are both porous media with varying degrees of structure and a variety of pore and capillary sizes, from micropores where capillary tension forces dominate to macropores where gravity drainage dominates and open channel flow velocities are possible. Uptake of water into root systems occurs due to strong tension gradients from tremendous osmotic pressure deficits in the root-trunk system. It is facilitated by fungi in many soils. Water is taken up into the plant and transmitted through intercellular flow paths in its stem or trunk to the leaf. During photosynthesis, water is made available to the stomata, which open and permit transpiration of water to water vapor as part of the complex photosynthetic process. Soil moisture and groundwater are also sometimes hard to distinguish. Soils are considered to be in the rooting zone and normally unsaturated. The soil layer is part of the vadose zone (Figure 4.6), which is normally unsaturated and may include porous media that are not soils. However, perched saturated layers can occur in soils after large infiltration events and lay above unsaturated layers. In the traditional groundwater layer

the porous medium is saturated. Water movement in porous media determines changes in $O + S_m + GW$. Richard’s equation can be used to determine the mass flux, $q$, between horizontal porous media layers of volumetric soil moisture $\theta$. It is expressed here in two dimensions with consideration for phase change due to soil freezing, $F$:

$$\frac{\partial \theta}{\partial t} = -\nabla q + F.$$  \hspace{1cm} 4.8

The mass flux, $q$, can be found from Darcy’s Law if flow conditions are non-turbulent:

$$q = K(\theta) \nabla (\psi(\theta) + z)$$  \hspace{1cm} 4.9

where $K$ is the hydraulic conductivity, which is a function of $\theta$ when unsaturated and otherwise a property of the permeability of the porous medium, $\psi$ is the matric suction, also a function of $\theta$ (zero when saturated) and $z$ is height above a datum. Matric suction can cause capillary rise from the saturated groundwater into the vadose zone, thereby providing water for root uptake. It should be noted that most soils and subsurface materials have substantial macropores for which Darcy’s Law does not apply and where flow is governed by pipeflow considerations and occurs at high velocities. Many of these macro pores are created by organisms, either former roots or tunnels, as well as due to cracking of heavy textured soils and subsurface materials.

The mass budget described previously is linked to the Earth surface energy budget in many ways. The evaporation, sublimation and melt terms all involve phase change and so the rates are controlled by available energy and the latent heat transfer coefficient. An example for evaporation rate, $E$ is:

$$E = \frac{Q_E}{\lambda} = \frac{Q^* - G - H}{\lambda}$$  \hspace{1cm} 4.10

where $Q_E$ is the energy available for evaporation and $\lambda$ is the latent heat of vaporization of water. As $Q_E$ is often unknown, it can be estimated as the net radiation flux, $Q^*$, less the ground heat flux, $G$, less the sensible heat flux (turbulent heat exchange) with the atmosphere, $H$. These terms can be estimated from the net solar radiation and net thermal radiation to the surface, ground and atmospheric temperatures and wind speed. Similar equations are developed for sublimation and melt using the latent heat of sublimation and of fusion. The energy budget is considered in more detail in Chapter 3 and evapotranspiration processes in Chapter 9.

4.3 SVATS – 1D Land Surface Schemes and Other Representations of Water Movement

4.3.1 Overview and Approach

The difficulty in solving the preceding hydrological process terms for a near surface control volume has led to the development of sophisticated numerical land surface schemes (LSS) with their implicit soil-vegetation-atmosphere transfer scheme (SVATS)
solutions. These physical calculations are coupled mass and heat budgets where the mass and heat terms are linked via phase change, radiative transfer and conductive heat transfer terms. Vegetation is parameterized where necessary to impact the calculations of water and energy movement and updating of state variables of storage terms. Animal and human manipulation of the land surface is generally ignored. There are a wide variety of LSS, some of which are listed in Table 4.1.

### 4.3.2 Canadian Land Surface Scheme and Variants

To describe all of these LSS would be exhaustive and unproductive as they have very similar approaches. To illustrate the primary functions of a LSS, one, the Canadian Land Surface Scheme (CLASS; Verseghy, 1991; Verseghy et al., 1993) is described in detail here and noted where it differs from other LSS. It should be noted that CLASS is frequently updated and so new versions may deviate from the following description. CLASS calculates the energy and water balances of the land surface from an initial starting point, making use of atmospheric forcing data to drive the calculations. When CLASS is run in coupled mode with an atmospheric model, the forcing data are passed to it at each time step from the parallel atmospheric model simulation. CLASS then produces surface parameters such as albedo and surface radiative and turbulent fluxes, which are in turn passed back to the atmospheric model. CLASS can also be run in uncoupled or offline mode, with forcing data derived from a separate atmospheric model run or from field measurements. CLASS calculates separate vertical energy and water balances for four subareas: canopy over snow, canopy over bare ground, bare ground and snow-covered ground (Figure 4.7). Physically based algorithms are used to calculate: evaporation and evapotranspiration; evapotranspiration and sublimation from vegetation canopy; interception, throughfall and drip of rainfall and snowfall; freezing and thawing of liquid and frozen water on the canopy and in soil layers; surface ponding and freezing of ponded water; sublimation from the snowpack; snowmelt; infiltration of rain into the snowpack; infiltration into soil; soil water movement between soil layers in response to gravity and suction forces; and temporal variation of snow albedo and density. Four
vegetation types are included in CLASS: needleleaf trees, broadleaf trees, crops, and grass. Each vegetation type is assigned a background value for physiological parameters such as albedo, roughness length, maximum and minimum leaf area index, etc. Certain physiological parameters vary throughout a simulation using annual or diurnal functions.

4.3.3 Mass and Energy Budget Calculation in CLASS

The surface energy balance equation for a non-vegetated surface is:

\[ Q^* + H + kE = G_0 \]  \hspace{1cm} 4.11

with the net radiation, \( Q^* \):

\[ Q^* = K^* + L^* \] \hspace{1cm} 4.12

where \( K^* \) and \( L^* \) are the net shortwave and longwave radiative fluxes, respectively, absorbed at the surface. \( H \) is the sensible heat flux, \( kE \) is the latent heat flux and \( G_0 \) is the surface heat flux into the ground or snowpack. \( K^* \) depends on incoming shortwave radiation, \( K^1 \), and surface albedo, \( a \), as:

\[ K^* = (1 - a)K^1. \] \hspace{1cm} 4.13

\( L^* \) is calculated as the difference between incoming longwave radiation, \( L^j \), and the radiation emitted by the surface, which is assumed to radiate as a black body:

\[ L^* = L^j - \sigma T_0^4 \] \hspace{1cm} 4.14

where \( \sigma \) is the Stefan-Boltzmann constant and \( T_0 \) is the surface temperature. The surface heat flux is calculated from the surface layer temperatures.

CLASS solves the non-linear surface energy balance equation iteratively using the Newton-Raphson method, with a maximum of five iterations. Iterative solutions are also found in BATS, CLM, MAPS, Noah-MP, SIIB2, VIC, and VISA. The surface temperature, \( T_0 \), at each iteration step is updated if the residual of the energy balance, \( RESID \), is greater than 5.0 W m\(^{-2}\), using:

\[ T_0 = T_{0-} - \frac{RESID}{\frac{d(T_0^2)}{dT} + \frac{d(T_0)\, d(T)}{dT}} \] \hspace{1cm} 4.15

where the subscript \(-\) denotes values calculated prior to incrementing \( T_0 \).

4.3.4 Turbulent Transfer Calculation

First-order closure is most commonly used for estimating turbulent fluxes of heat and moisture between the atmosphere and land surface. The widely applied bulk aerodynamic formulae are given by:

\[ H = \rho c_p C_M u(z)[T_0 - T(z)] \] \hspace{1cm} 4.16

for sensible heat flux (\( H \)), and \( \gamma \):

\[ \gamma E = \rho L_v C_H u(z)[q_0 - q(z)] \] \hspace{1cm} 4.17

for latent heat flux (\( \gamma E \)). In these equations, \( C_M \) is the scalar transfer coefficient assumed to be the same for both sensible and latent heat at reference height \( z \), \( u \) is the wind speed, \( T \) is the temperature, and \( q \) is the specific humidity with the subscript 0 indicating that it is the state at the surface.

Similarly, the momentum flux (\( \tau \)) is given by:

\[ \tau = -\rho C_D u(z)^2 = -\rho u^2 \] \hspace{1cm} 4.18

where \( C_D \) is the drag coefficient and \( u \) is the friction velocity. The drag and transfer coefficients depend on atmospheric stratification, which is commonly parameterized using Monin-Obukhov similarity theory or a Richardson number approach.

Variations of Obukhov length parameterizations are also used in the ECMWF land surface model, CLM, JULES, Noah-MP, SWAP, and VISA. The Obukhov length, \( L \), is the height above which buoyant production of turbulence dominates over shear production. \( L \) is used to characterize atmospheric stratification and is given by:

\[ L = -\frac{u^2 T(z)}{\kappa \varepsilon \theta_{v0}} \] \hspace{1cm} 4.19

where \( \varepsilon \) is the von Kármán constant, \( T \) is the air virtual temperature, \( g \) is acceleration due to gravity, and \( \theta_{v0} \) is the virtual temperature heat flux at the surface. The drag and scalar exchange coefficients are given by:

\[ C_D = \frac{\kappa^2}{\left[ \ln \left( \frac{z}{z_0} \right) - \psi_{m}(\zeta) \right]^2} \] \hspace{1cm} 4.20

\[ C_H = \frac{\kappa^2}{\left[ \ln \left( \frac{z}{z_0} \right) - \psi_{h}(\zeta) \right] \left[ \ln \left( \frac{z}{z_0} \right) - \psi_{s}(\zeta) \right]} \] \hspace{1cm} 4.21

where \( z_0 \) and \( z_{0s} \) are the surface roughness lengths for momentum and scalar transfer, respectively, and \( \psi_{m} \) and \( \psi_{s} \) are stability functions for momentum and scalar exchange.

The stability functions are given by the integrals:

\[ \psi_{m,k} = \frac{z/L}{\zeta} \int_0^{z/L} \left( 1 - \frac{\phi_{m,k}(\zeta)}{\zeta} \right) d\zeta \] \hspace{1cm} 4.22

where \( \phi_{m} \) and \( \phi_{s} \) are the universal functions for momentum and scalar exchange, respectively. Examples of those used in JULES (Dyer, 1974) for unstable conditions, and in Beljaars and Holtslag (1991) for stable conditions are:

\[ \phi_{m}(\zeta) = \begin{cases} (1 - a\zeta)^{-1.4} & -1 < \zeta < 0, \\ 1 + \zeta + b\zeta(c - d\zeta)e^{-\zeta} & \zeta \geq 0 \end{cases} \] \hspace{1cm} 4.23
where $a = 16$, $b = 2/3$, $c = 6$, and $d = 0.35$ are coefficients determined experimentally.

Values for roughness lengths are flow dependent; however, the most common approach used in land surface models is to use a constant value. The roughness length for heat and moisture transfer ($z_{0h}$) is smaller than that for momentum ($z_0$); $z_{0h}$ is commonly parameterized as a fraction of $z_0$. CLASS uses $z_{0h}/z_0 = 3.0$ (also in IAP94). JULES and ISBA use $z_{0h}/z_0 = 10.0$. The roughness length for momentum of snow is set to 0.001 m, while $z_{0h}$ vegetation are specified parameters.

CLASS does not employ the combination approach to evapotranspiration such as developed by Pennman (1948) and enhanced by Monteith (1965). Rather, it uses a Dalton-type bulk transfer approach with adjustments for unsaturated surfaces using resistance formulations to link vegetation and soil to the atmosphere. The surface evaporation efficiency coefficient, $\beta$, is used to calculate the soil surface specific humidity and, therefore, affects the magnitude of latent heat fluxes. $\beta$ characterizes surface specific humidity. If there is snow cover or water ponded on the surface, then $\beta$ is set to 1.0 and the surface specific humidity is set to the saturation specific humidity.

CLASS uses the relationship from Lee and Pielke (1992) to calculate $\beta$ as a function of volumetric soil moisture content $\theta$:

\[
\beta = \begin{cases} 
1.0, & \theta_{fc} > \theta_{fc,1} \\
0.25 \left[ 1 - \cos \left( \pi \frac{\theta}{\theta_{fc,1}} \right) \right]^{2}, & 0.04 \leq \theta_{fc,1} < \theta_{fc,1} \\
0.0, & \theta_{fc,1} < 0.04 
\end{cases} 
\]

where $\theta_{fc,1}$ is the field capacity of the first soil layer and is calculated from the soil saturated hydraulic conductivity, which is calculated using the widely applied Clapp and Hornberger (1978) relationships. This cosine parameterization is also used in CLM. Alternative parameterizations that focus on critical and wilting points are used in Noah-MP, JULES and MOSES.

### 4.3.5 Ground Heat Flux

The ground heat flux, $G_0$, is calculated by deriving a linear equation as a function of $T_0$ by assuming that the variation of temperature within a soil or snow layer with depth can be expressed using a quadratic equation. For bare ground, the linear equation for $G_0$ has slope and intercept as functions of the average temperatures, thicknesses, and top and bottom thermal conductivities of the top three soil layers:

\[
G_0 = a_1 T_1 + a_2 T_2 + a_3 T_3 + a_4 T_0 + a_5 
\]

where $T_1$, $T_2$ and $T_3$ are the average temperatures of the first, second and third soil layers, respectively, and the $a$ terms are the coefficients. For snow-covered ground, the linear equation for $G_0$ has slope and intercept as functions of the average temperature, thickness and thermal conductivity of the snowpack. Snowmelt occurs in two ways: if the solution of the surface energy balance results in $T_0 > 0°C$ or if the energy balance calculations for the snowpack results in a snowpack temperature $T_0 > 0°C$.

CLASS uses six soil layers, typically with depths of 0.10, 0.35, 1.10, 2.10, 3.10, and 4.10 m below ground surface. The finite-difference scheme of the one-dimensional heat conservation equation is applied to each soil layer, giving the change in average soil layer temperature, $T_i$, over time step $\Delta t = 1800$ s as:

\[
T_i(t + \Delta t) = T_i(t) + \frac{[G(T_{z_i-1},t) - G(T_{zi},t)]}{C_i/\Delta t_i} \pm S_i
\]

where $G(T_{zi}, t)$ and $G(T_{zi}, t)$ are the conductive heat fluxes at the top and bottom of soil layer $i$. $C_i$ is the soil volumetric heat capacity, $\Delta t_i$ is the layer thickness and $S_i$ is included for cases of freezing or thawing, or groundwater percolation. The conductive heat fluxes between soil layers are calculated from average layer temperatures by assuming that temperatures within each layer vary according to a quadratic function of depth.

### 4.3.6 Soil

The soil layer moisture contents are calculated using a conservation equation analogous to that for heat. For average layer volumetric liquid water content, $\bar{\theta}_{li}$ is:

\[
\bar{\theta}_{li}(t + \Delta t) = \bar{\theta}_{li}(t) + \frac{F(T_{zi}, t) - F(T_{zi}, t)}{C_i/\Delta t_i}
\]

where $F(T_{zi}, t)$ and $F(T_{zi}, t)$ are the liquid water flow rates at the top and bottom of soil layer $i$. Changes in frozen water content, $\bar{\theta}_{li}$, occur if $T_i(t + \Delta t) > 0°C$ while ice is present, or if $T_i(t + \Delta t) < 0°C$ while $\bar{\theta}_{li}$ is greater than a limiting value of 0.04. Below the surface, $F(T_{zi})$ are calculated as one-dimensional Darcian fluid flow as used in most land surface models. Soil water vapor movement and liquid water movement according to temperature gradients are ignored. Soil water suction and hydraulic conductivities are calculated based on soil texture from the widely applied Clapp and Hornberger (1978) relationships. $F(0)$ is the infiltration rate at the surface.

Most land surface models have analytical infiltration schemes due to the computationally expensive requirements of numerical schemes. CLASS uses the two-stage Mein and Larson (1973) analytical infiltration parameterization for uniform soils and constant rainfall intensity. Two-stage refers to separate calculations for pre-pended and ponded infiltration rates, relaxing the Green and Ampt (1911) assumption of constant head at the surface. The infiltration rate is given by:

\[
F(t) = \begin{cases} 
K_w(p_w + Z_t)/Z_t, & t < t_p \hspace{1cm} \text{(unsaturated)} \\
K_s(p_s + Z_t)/Z_t, & t \geq t_p \hspace{1cm} \text{(saturated)} 
\end{cases}
\]

where $K_w$ is the hydraulic conductivity at the wetting front, $p_w$ is the soil water potential at the wetting front, $Z_t$ is the infiltration depth and $Z_s$ is the ponding depth. $t$ is the infiltration time and $t_p$ is the ponding start time. Green–Ampt-type infiltration schemes are used in SWAP and as an option in CRHM.
Zhao and Gray (1997; 1999) used results from a physically based numerical model to develop a general parametric expression for estimating infiltration into frozen soils in prairie and boreal forest environments. The relationship related infiltration to total soil saturation (liquid + frozen water) and temperature at the beginning of snowmelt, the soil surface saturation during melt and the infiltration opportunity time. Infiltration to frozen soil calculations are grouped into three categories:

**Restricted** - infiltration is completely restricted due to impermeable surface conditions such as ice lens formation or saturated conditions;

**Limited** - unsaturated conditions where capillary flow predominates and infiltration is primarily controlled by soil physical properties;

**Unlimited** - gravity flow predominates and water infiltrates rapidly into macropores and other large air-filled voids.

A parametric equation is used for the limited infiltration case:

\[
F(0) = C S_0^{0.92} (1 - S_0)^{1.64} \left( \frac{273.15 - T_f}{273.15} \right)^{-0.45} T_f \leq 273.15
\]

where \( C \) is an empirical constant equal to 2.10 and 1.14 for prairie and forest soils, respectively, \( S_0 \) is the soil surface saturation, \( S_0 = \theta / \theta_b \) is the pre-melt pore saturation of the upper soil layer with \( \theta_b \) being the volumetric soil moisture (liquid + frozen water) at the start of infiltration, \( T_f \) is the pre-melt temperature of the upper soil layer and \( t_0 \) is the infiltration opportunity time. \( t_0 \) is estimated from SWE as:

\[
t_0 = 0.65 \text{SWE} - 5.
\]

The maximum amount of water that can infiltrate in the limited case, the water storage potential \( (W_{sp}) \), is constrained as:

\[
W_{sp} = 0.6 \theta_b (1 - S_0) z_p
\]

where \( z_p \) is depth of a highly permeable surface layer (e.g., thickness of organic layer or depth of surface-connected cracks).

The thermal and hydraulic properties of each of the modeled soil layers are determined differently for different ground types. Soil thermal conductivities are used to calculate heat fluxes between soil layers, and at the soil-atmosphere and soil-snow interfaces, thus affecting the magnitudes of non-radiative fluxes. CLASS uses the parameterization of Côté and Konrad (2005) for soil thermal conductivity. Soil thermal conductivity, \( \lambda_{soil} \), is calculated using a relative thermal conductivity, \( \lambda_r \), which has a value of 0.0 for dry soils, \( \lambda_{dry} \), and 1.0 at saturation, \( \lambda_{sat} \):

\[
\lambda_{soil} = [\lambda_{sat} - \lambda_{dry}] \lambda_r + \lambda_{dry}
\]

\( \lambda_r \) is calculated from the degree of saturation, \( S_r \), as follows

\[
\lambda_r = \frac{x S_r}{1 + (x - 1) S_r}
\]

where \( x \) is an empirical coefficient.

\( \lambda_{dry} \) is calculated using an empirical relationship with different coefficients for mineral and organic soils:

\[
\lambda_{dry} = 0.73 \exp(-2.76\theta_b) \quad \text{for mineral soils}
\]

\[
\lambda_{dry} = 0.30 \exp(-2.00\theta_b) \quad \text{for organic soils}
\]

where \( \theta_b \) is the soil porosity. \( \lambda_{sat} \) is calculated using the linear averaging approach of de Vries (1963), as suggested by Zhang et al. (2008), rather than geometric averaging used in Côté and Konrad.

\[
\lambda_{sat} = \lambda_{w} \theta_{wp} + \lambda_{sat}(1 - \theta_{wp})
\]

\[
\lambda_{sat} = \lambda_{i} \theta_{ip} + \lambda_{sat}(1 - \theta_{ip})
\]

where \( \lambda_{w} \) and \( \lambda_{i} \) are the thermal conductivities of water and ice, respectively.

### 4.3.7 Snowpack

The snowpack is modeled as a single layer of variable depth using the same equations for the surface energy balance and heat fluxes as presented previously. Incoming shortwave radiation, \( K^a \), is allowed to penetrate the snow surface, decreases exponentially with depth following Beer’s law and can be absorbed by the underlying soil.

Blowing snow involves the horizontal redistribution and sublimation of snow. Despite its importance to mass budgets in high altitude and high latitude cold regions (e.g., Pomeroy and Li, 2000), these processes have yet to receive widespread parameterization in hydrological models and land surface schemes. Blowing snow calculations are included in a few land surface and hydrological models as options: CRHM and VIC. PBSM calculates blowing snow transport and sublimation rates for steady-state conditions using mass and energy balances. PBSM was initially developed for application over the Canadian Prairies, characterized by relatively flat terrain and homogeneous crop cover. Refer to Pomeroy and Gray (1990), Pomeroy and Male (1992), Pomeroy et al. (1993) and Pomeroy and Li (2000) for details on algorithm development.

PBSM is for fully developed blowing snow conditions and is therefore restricted to minimum fetch distances of 300 m following measurements by Takeuchi (1980). Blowing snow transport fluxes are the sum of snow transport in the salination and suspension layers, \( F_{sab} \) and \( F_{susp} \), respectively. Salination of snow must be initiated before snow transport can occur in the suspension layer and blowing snow sublimation can occur. \( F_{well} \) is calculated by partitioning the atmospheric snow stress into that required to free particles from the snow surface, to that applied to nonerosible roughness elements (vegetation stalks or shrubs) and to that applied to transport snow particles (Pomeroy and Gray, 1990):
fractional snow covered area, $f_s$ is a function of average snow mass or depth. The following linear function is used in CLASS, ECMWF model and SIB2:

$$f_s = \min\left(\frac{d_s}{d_0}, 1\right)$$

where $d_s$ is snow depth and $d_0$ is a threshold parameter set to 0.10 m.

The density of fresh fallen snow affects heat transfer within snowpacks and the atmosphere. There are no physically based parameterizations of fresh snow density in use as they require detailed simulations of crystal size, shape and packing; rather there are a number of empirical functions based on combinations of air temperature, humidity and wind speed. CLASS calculates fresh snow density as a function of air temperature using an equation presented by Hedstrom and Pomeroy (1998):

$$\rho_{s, f} = 67.92 + 51.25 \exp\left(\frac{T(z) - T_{ref}}{2.59}\right)$$

for air temperatures below 0°C, and using an equation from Pomeroy and Gray (1995):

$$\rho_{s, f} = 119.17 + 20.0(T(z) - T_{ref})$$

for air temperatures at or above 0°C.

Snow density generally increases over time due to grain metamorphism, compaction from the weight of overlying snow and the refreezing of meltwater. Snow density is commonly used to parameterize thermal conductivity, liquid water content and, indirectly, snow cover fraction (Pomeroy and Brum, 2001). CLASS uses an empirical equation in which the density of snow, $\rho_s$, increases exponentially from the fresh snow value, $\rho_{s, f}$, to a maximum possible snow density, $\rho_{s, max}$:

$$\frac{d\rho_s}{dt} = 0.01(\rho_s - \rho_{s, max})$$

where the value 0.01/3600 is an empirically determined time scale. The maximum snow density is calculated from snow depth following Tabler et al. (1990):

$$\rho_{s, max} = A_S - \frac{204.70}{d_s} \left[1 - \exp\left(-\frac{d_s}{0.673}\right)\right]$$

where $A_S$ is set to 700.0 kg m$^{-3}$ for snowpacks near 0°C, and to 450.0 kg m$^{-3}$ for colder snowpacks following Brown et al. (2006). Similar empirical parameterizations are used in the ECMWF land surface model and ISBA.

Snow albedo exerts a strong control on the timing of snowmelt and land surface-climate feedbacks. Albedo depends on physical characteristics of snowpacks (i.e., grain structure, depth, contaminants) and also on the solar angle and spectral distribution of radiation. In CLASS, snow albedo is modeled using empirical exponential decay
functions. Snow albedo, $\alpha_s$, decreases exponentially from a fresh snow value of 0.84 using the function

$$\alpha_s(t + \Delta t) = (\alpha_s(t) - \alpha_{s,old}) \exp \left( -\frac{0.01 \Delta t}{3600} \right)$$

where $\Delta t = 1800$ s. The background old snow albedo, $\alpha_{s,old}$, is set to 0.50 if the melt rate is non-negligible or the snowpack temperature is greater than $-0.01 \, ^\circ C$, otherwise $\alpha_{s,old} = 0.70$. The snow albedo is reset to 0.84 if a snowfall greater than or equal to 0.1 mm occurs. Similar empirical parameterizations are used in ISBA, the ECMWF land surface model and Noah-MP.

The thermal conductivity of snow, $\lambda_s$, is used along with the vertical temperature gradient to calculate the heat flux through the snowpack. Most models parameterize an effective thermal conductivity as a quadratic or power function of snow density. CLASS uses an empirical equation from Sturm et al. (1997):

$$\lambda_s = \begin{cases} 
3.233 \times 10^{-6} \rho_s^2 - 1.01 \times 10^{-3} \rho_s + 0.138, & \rho_s \geq 156.0 \\
0.234 \times 10^{-3} \rho_s + 0.023, & \rho_s < 156.0
\end{cases}$$

The retention of liquid water in snowpacks controls the timing of runoff. Gravitational drainage of liquid water from snowpacks can be rapid due to high porosity and preferential flow pathways, and capillary forces maintain an irreducible water content. Some highly detailed snow models calculate vertical water velocities; however, this can make a model much more computationally expensive, potentially unstable and differences are only realized on short time scales. Rather, most land models consider runoff from snowpacks once a holding capacity is exceeded. CLASS uses a constant snowpack maximum liquid water retention capacity $\gamma_{s,max} = 4\%$ by weight. Noah-MP also uses a constant maximum liquid water retention capacity.

4.3.8 Vegetation and Transpiration

In 1802, Dalton showed the rate of evaporation from a water surface is directly proportional to the differences between the saturation vapor pressures at the surface temperature of the water and the dew point of the air (Penman, 1947). In CLASS, surface temperature is used to estimate the saturated specific humidity at the surface of the canopy. The implicit assumption in this method is that leaf sub-stomatal cavities are saturated at the temperature of the leaf surface (Verseghy et al., 1993). The humidity gradient can then be determined between the surface and that measured at some reference height above the surface from the air temperature and relative humidity. The flux of water vapor along this gradient also takes into consideration the aerodynamic resistance of the canopy via turbulent transfer and the logarithmic wind profile and the canopy resistance.

This Dalton-type approach is widely used for estimating surface fluxes and is commonly applied in land surface parameterization schemes (Mahrt, 1996; Sellers et al., 1997). This may be attributed in part because the method: 1) can be relatively simple to apply, 2) is driven by surface temperature, which is commonly diagnosed by iterative solutions to the surface energy balance in land surface schemes, and 3) provides a direct estimate of the flux-gradient between the surface and atmosphere. The BT method may also be applied to both land surfaces and open water surfaces and has the potential for directly integrating remotely sensed surface temperature data, obtained via field measurements or derived from airborne or satellite imagery.

Based on the model diagnosis of surface temperature from an iterative solution to closing the surface energy balance, the parameterization requires measurements or estimates of air density, surface temperature, vapor pressure, wind speed, vegetation height, and soil moisture used in the calculation of $r_c$:

$$E = \frac{\lambda_p (q_v(T_v) - q)}{r_a + r_c}$$

where $q_v$ is the saturated specific humidity (kg kg$^{-1}$) at the surface temperature ($T_v$) and $q$ is the specific humidity of the air (kg kg$^{-1}$). Application of Equation 4.50 to non-saturated surfaces requires consideration of the resistances of water vapor transfer to the atmosphere. Estimates of the aerodynamic resistance are obtained assuming a logarithmic wind profile formulation:

$$r_a = \frac{\ln \left( \frac{z}{z_0} \right)^2}{k^2 u}$$

where $u$ is the wind speed at the reference height, $z$, $d = 0.67$ h is the displacement height of the vegetation (m) and $k$ is the von Kármán constant (0.41). Estimates of canopy resistance are obtained using the general model proposed by Jarvis (1976) and the experimental relationships developed by Verseghy et al. (1993) for the multiplicative factors describing environmental stress effects on stomatal control:

$$r_c = r_{c,\text{max}} f_1 f_2 f_3 f_4$$

where $r_{c,\text{max}}$ represents the minimum unstressed canopy resistance (s m$^{-1}$). The multiplicative factors describe stomatal control as a representative value of $f$ under what may be considered optimal conditions for plant growth and a value $>1$ under stressed conditions. $f_1$ increases under conditions when light is limiting and is a function of the incoming solar radiation, $K_L$ (W m$^{-2}$), required for photosynthesis:

$$f_1 (K_L) = \max(1.0, (500/K_L - 1.5))$$

$f_2$ is a function of the vapor pressure, $e$, deficit (mb) required to maintain water and nutrient uptake to the plant, which increases as the plants ability to transmit water from the soil rooting zone is exceeded:

$$f_2 (\Delta e) = \max(1.0, (\Delta e/5.0))$$

$f_3$ is a function of soil moisture supply, specifically the soil moisture tension, $\psi$ (m), which increases with decreasing soil moisture:
\[ f_s(\psi) = \max(1.0, \frac{\psi}{40.0}) \]

where \( \psi \) is derived using the Campbell power law function for specific soil texture classes based on the air entry tension, \( \psi_{ae} \), porosity, \( \varphi \), a pore size distribution index, \( b \), and soil moisture, \( \theta \) (Campbell, 1974):

\[ \psi = \psi_{ae} \left( \frac{\varphi}{\theta} \right)^b \]

\( f_s \) is a function of temperature with an operating range between 0°C and 40°C:

\[ f_s(T) = 1.0 \text{ if } T < 40^\circ C \text{ or } > 0^\circ C \]

or

\[ f_s(T) = \frac{5000}{r_{min}} \text{ if } T > 40^\circ C \text{ or } > 0^\circ C \]

and indexes the range of temperatures at which transpiration may be considered to occur.

### 4.3.9 Limitations of LSS

For all of their complexity, LSS struggle to calculate realistic water budgets at scales important for ecosystems. This is partly because of the large number of unconstrained parameters that must be set from sparse or non-existent observations or ecosystem-type lookup tables. It is also because LSS are essentially 1D representations of the water budget that attempt to homogenize vast swaths of the Earth’s surface, whereas in nature, 3D interactions and ecosystem variety are important to the hydrological cycle. The next section will examine this variability and how hydrological models attempt to include it without becoming overwhelmed with physical equations and uncertain and poorly constrained parameter values.

### 4.4 Hydrological Models and Simulations

#### 4.4.1 Classification of Hydrological Models

Although there are several ways of classifying hydrological models (Chow et al., 1988; Singh, 1995; Abbot and Ressgaard, 1996), modeling approaches may be distinguished by three main characteristics (Grayson and Bloschl, 2000):

1. the nature of the basic algorithm (empirical, conceptual and physical or process-based),
2. the approach to input or parameter specification (stochastic or deterministic), and
3. the spatial representation (lumped or distributed).

**Empirical models** are derived from data; therefore they are not based on scientific laws describing physical processes. Since the model structure relies on a given range of data, the applicability and validity of the resulting model is limited to this range of data. **Conceptual models** are based on a theoretical understanding of the hydrological processes. They generally use physical laws but in a highly simplified form. Conceptual models contain parameters that may have physical significance; however, most of the parameters are conceptual and hence the definition of their values relies on calibration. A typical example of a conceptual rainfall runoff model is the SLURP model which makes use of the linear reservoir and channel concepts (Kite and Kounen, 1992). In contrast, **physically based models** are deterministic calculations based on physics, chemistry and biology to describe hydrological processes and generally have physically identifiable parameters.

Typically, the process representation and inputs of hydrological models are deterministic in that they are based on physical laws such that the same inputs and model parameterizations generate the same model output. Stochastic models include some random component that limits the exact model prediction and are often associated with given probabilities and confidence intervals.

Lumped models deal with a catchment as a single computational unit. They relate precipitation inputs to discharge outputs without any consideration of the spatial patterns of the hydrological processes and basin characteristics. Therefore, they cannot capture the lateral or horizontal redistribution of moisture in soils and in the drainage network. Conversely, distributed models explicitly account for the spatial patterns of process response. Well-known examples of distributed hydrological models are the TOPMODEL (Beven and Kirkby, 1979; Beven et al., 1995; Beven, 1997), the SHE model (Abbot et al., 1986; Ressgaard and Storm, 1996), the SWAT model (Arnold et al., 1998), and the WATFLOOD model (Kounen, 1988). However, these models use different approaches for process representation. For example, while TOPMODEL uses a conceptual approach based on a detailed topographic description, the SHE model integrates a 3D groundwater model, a 2D diffusive wave approximation for the overland flow and a 1D full dynamic component of the river flow.

#### 4.4.2 Physically Realistic Hydrological Modeling

Many older hydrological concepts often persist in hydrological models despite being dismissed by more recent scientific investigations. This situation is not new (e.g., Klemes, 1986). Predictive problems caused by these misconceptions are particularly evident for cold, semi-arid, and arid regions that are outside of the temperate zones that have been the primary regions of hydrological model conceptualization and development, but are found in all regions (Pomeroy et al., 2013). The following are a few examples of components that should be in physically based hydrological models – their inclusion means physically identifiable parameters can be more easily found, that scaling problems can be minimized and the need for parameter calibration and regionalization is vastly reduced.

(1) There are several relationships to estimate solar and longwave radiation from latitude, time of year, air temperature range and humidity that can be used to drive
energy balance snowmelt and combination-type evapotranspiration algorithms (Walter et al., 2005; Sicart et al., 2006; Shook and Pomeroy, 2011).

(2) There are many dynamic vegetation growth and rooting algorithms available to provide resistance to potential evaporation, and complementary feedback relationships are available for when vegetation, roots and soils are relatively unknown (Granger and Gray, 1989; Armstrong et al., 2010; Brimelow et al., 2010).

(3) Precipitation phase is controlled by the psychrometric equation, snowfall gauge undercatch can be very substantial but is correctable and sublimation can consume a substantial proportion of snowfall in dry environments. Sublimation can be estimated using energy balance and aerodynamic approaches (Pomeroy and Gray, 1995; Harder and Pomeroy, 2013).

(4) Macropores caused by plants, animals and humans can provide a primary soil flowpath and infiltration equations can be modified for macropore flow. The variance of soil properties over a hillslope is a critical influence on variable contributing areas, fill and spill mechanisms, which control saturated flow at the soil-bedrock interface (Beven and Kirkby, 1979; Beven and Germann, 1982; Tromp-van Meerveld and McDonnell, 2006).

(5) Drainage basin contributing areas expand and contract as depressional storage and saturated flow pathways fill and empty; the relationship between contributing area and storage is non-linear hysteretic but can be modeled using network connectivity concepts (Spence and Woo, 2003; Phillips et al., 2011; Shook et al., 2013).

(6) Sub-surface flow abounds and its velocity is controlled by soil and topographic parameters, not overland open channel flow velocities (Henderson and Wooding, 1964; Sabsevari et al., 2010).

(7) Infiltration and soil hydraulics are controlled by the interaction of soil ice content and porosity over time, which is controlled by coupled energy and mass balance equations and influenced by the depth of freezing and the presence of permafrost (Zhao and Gray, 1999; Gray et al., 2001; Quinton and Gray, 2001).

4.4.3 Scaling Issues

Blöschl and Sivapalan (1995) defined the term “scale” as a characteristic length or time and the term “scaling” as a change in scale. Moreover, upsizing means transferring information from smaller to larger scales (i.e., aggregating) whereas downsampling refers to the opposite transference of information, where the information is disaggregated from large to small scales.

In general, the scale at which the data are collected is different from the scale at which predictions are needed. Measurements are made to obtain information about natural processes; however, measured data will not exactly reproduce the natural variability of the processes because of observational error and the limited spatial dimensions of measurements. Hence, patterns of measured data will differ from true natural patterns.

For example, precipitation is measured at widely spaced points. Even in an instrumented catchment, spacing between gauges may be on the order of 5–10 km, which is longer than the length scales of convective storms. Spatial interpolation of monthly precipitation can appear reasonable in some studies, but hourly or even daily precipitation cannot be reasonably interpolated from widely spaced precipitation gauges because of storm dynamics (Johnson and Hanson, 1995). For hydrology, precipitation information is needed on time scales of at least a few hours to calculate runoff rates and so this interpolation problem is a very serious one. Similar problems exist for temperature, humidity, wind speed and short and longwave radiation. Wind data, so critical to calculation of evaporation and turbulent heat transfer, are even rarer than precipitation data.

Typically, the modeling or working scale is a compromise solution between process representation and the model application. Since, the modeling scale is different than the process scale (i.e., scale that the natural phenomena exhibit) and much larger than the observation scale (i.e., scale at which observations are sampled), scaling techniques are needed to bridge this gap (Blöschl and Sivapalan, 1995). Interpolation and aggregation/disaggregation techniques are the more common methods used. Interpolation techniques estimate patterns from points (i.e., changes of scale in terms of spacing) whereas aggregation methods involve the combination of a number of point values in space to form one average value (i.e., change of scale in terms of support) that correspond to an increase in support scale. Disaggregation methods, on the other hand, are the opposite transformation and estimate patterns from spatial average values.

Hydrological models are very sensitive to scaling issues. The typical modeling approach is to apply the same model structure in several basins whereas the parameters, empirical or not, are varied in the calibration process. This means that the model structure is general but not the parameters. Therefore, a change in scale might involve a change in the parameter values, in particular if these parameters are related to local conditions such as climate and physiography (Bergström and Graham, 1998).

4.4.4 Aggregation Methodologies

Heterogeneity in the landscape has forced hydrologists to conceptualize the physics and seek effective parameter values (Pietroniro and Soulis, 2003). Distributed hydrological models use aggregation methods to account for landscape variability and processes representation; however, a critical point in the application of these models is the choice of element size. In general, increasing the level of discretization increases the accuracy of the simulation, but there should be a level beyond which the model performance cannot be increased (Wood et al., 1988). In addition, the smaller the grid size in which the catchment is divided, the larger is the volume of information needed and the associated computational time.

A typical method for representing landscape heterogeneity is the grid-based approach (e.g., Système Hydrologique Européen, SHE). In this case, the basin is split into a number of usually square elements linked to channel reaches. Each grid is the computational element and has a specific surface elevation given by a digital elevation map. An
example is shown in Figure 4.8, applied to streamflow routing. This approach has the potential to assign distributed fields of gridded meteorological data from atmospheric models, gridded topographic data from digital elevation models and gridded vegetation descriptions from classification of satellite remote sensing radiometric pixels and therefore the capability to predict a variety of distributed processes at each element grid. However, predictions are grid-scale dependent. Refsgaard (1997) concluded, after comparing finer and coarser grids that simulations based on 1000 m or larger grid size, while still accurate, may require recalibration of the parameters due to sub-grid heterogeneity and scale differences with process length scales. In distributed hydrological models, the assumption of areas with similar hydrological behavior is a common method for reducing model complexity. The Representative Elementary Area (REA) approach defined by Wood et al. (1988) assumes that the size of the areal elements is defined by considering processes at smaller scales as insignificant for modeling purposes. Wood et al. (1988) carried out an empirical averaging experiment to assess the impact of changing scale. Since different correlation lengths and spatially invariant precipitation did not significantly change this result, they concluded that the REA was strongly influenced by topography. Even though the concept of universal REA is attractive for modeling purposes (e.g., grid-based models), it had been demonstrated that the size for a model element is dependent on the processes being represented and the type of climate, terrain and vegetation where the model is being applied (Bîstîschl et al., 1995; Woods et al., 1995). These results show that there is no evidence for one universal size of REA and that the size of REA depends on many factors, including

storm duration and variability, flow routing and infiltration characteristics. It is therefore apparent that the size of the REA will be specific to catchment and application.

The complexities of the environment and data availability have seen many researchers favor aggregated computational units. The main reason for this is to reduce the increasing computational time, especially for larger basins and finer spatial resolution, and to restrict the number of parameters to be determined. Hydrological Response Units (HRUs) are one of the more common aggregation approaches where the model units are defined according to the hydrological behavior. These units can be related to natural landscape units and are characterized from an understanding of how hydrological processes interact with biophysical characteristics of the catchment. Therefore HRUs are usually defined by overlapping maps of different characteristics, such as soils, slope, aspect, vegetation cover, etc. (Flügel, 1995; Beven, 2000; Pemery et al., 2007). Dornes et al. (2008b) explored the concept of warranted complexity in HRU delineation and showed that process interactions determine the required number of HRU for successful simulation of hydrological processes and streamflow. An example of HRU application to a mountain catchment is shown in Figure 4.9, where the main HRU are due to slope, aspect and vegetation (Dornes, 2013). In agricultural catchments, soil type, vegetation cover, cultivation practice and drainage will determine the principle HRU. Grouped Response Units (GRUs; Kouwen et al., 1993) are an alternative for describing spatial variability, where areas with similar land cover, soils, etc., are grouped for parameter estimation purposes with no requirement for grids or sub-basins to be hydrologically homogeneous. Kuchment et al. (1996) used a finite-element
schematization in a physically based rainfall runoff model for representing the main channel network and landscape mosaic. By combining topographic, soil and land use maps, they divided a basin into 445 finite elements, giving an average area for a single element of about 7.5 km² and providing 99 finite elements along the river network but still preserving the general pattern of steep and gentle slopes. Effective parameters were found through model calibration against streamflow. Another methodology is Representative Elementary Watershed (REW; Reggiani and Schellekens, 2003), defined as the smallest elementary unit into which a catchment can be discretized whilst still being representative of other sub-entities of the catchment. The REW is assumed to be composed of five sub-regions: the unsaturated zone, saturated zone, concentrated overland flow zone, saturated overland flow zone and channel zone. The main difference with the traditional approaches (i.e., point mode equations) is that the governing equations derived from the REW approach are applicable directly at the catchment scale, and hence they have been derived in a comprehensive manner for the whole catchment or REW, as opposed to being derived separately for different processes (Lee et al., 2007). Although the REW approach is proposed as an alternative method, its application is still limited because of fundamental differences between this concept and catchment function in nature.

The major disadvantage in models using aggregation methods based on similarities (e.g., HRUs, GRUs), is the way in which each unit is considered to be spatially homogeneous. In general, within the computational element the physical properties are conceptualized and sometimes, effective parameter values or sub-HRU probability distribution functions are used to account for sub-grid variability. There are several methods currently used to attempt to include sub-grid heterogeneity into distributed modeling efforts. One includes replacement of the most important dependent variables in the governing equations by probability distribution functions (pdfs). Becker and Braun (1999) applied areal distribution functions of soil water holding capacity to represent spatial heterogeneities distinguishing between agricultural and forested HRUs. However, they concluded that additional scaling laws are required for describing lateral flows between landscapes. Faria et al. (2000) examined the forest canopy influence on snow-cover depletion. They found that the frequency distribution of SWE under boreal canopies fit a log-normal distribution; and the highest canopy density had the most variable snow water equivalent. The relationships between the spatial distributions of SWE and melt energy promoted earlier depletion of the snow cover than if the melt energy were uniform, with the strongest effect in heterogeneous or medium-density canopies. Another example that includes the explicit incorporation of parameterization of sub-grid variability through the use of a depletion curve into the snowmelt model was made by Luce et al. (1999) and Luce and Tarboton (2004). Analogous conclusions were described by Pomeroy et al. (2004), where one of the major scaling problems in applying point-scale equations over large areas is the spatial association between driven variables and/or parameters, which can result in spatial correlations and covariance amongst the terms of a physically based equation. A different approach is the up-scaling of point-scale hydrologic conservation equations to the computational grid areas. They mainly seek to scale the governing equations so that they accurately represent the phenomena at the larger modelling scale. It is based on the "coarse-graining" approach, which states that mechanisms important in one scale are not important in either a much larger or much smaller scale (e.g., Kavvas et al., 1998; Kavvas, 1999).

4.4.5 Parameter Calibration

Calibration of hydrological models is intended to estimate model parameters that allow the model to closely match the observed behavior of the real system it represents (Gupta et al., 1998). Traditionally, the calibration of hydrological models has been performed manually by trial-and-error against streamflow observations. The trial-and-error method implies a manual parameter adjustment by running a number of model simulations. Due to its limitations (e.g., subjectivity and time-consuming processes), research into automatic calibration procedures based on the increasing computer power has led to the use of different automatic parameter optimization approaches. These approaches are based in general on optimizing (i.e., minimizing or maximizing, as appropriate) the value of objective functions, which are numerical measures of the difference between observed and simulated data (Sorooshian and Gupta, 1995). Automatic parameter optimization has three advantages over manual calibration: it is faster since it is computer based, it is less subjective and the confidence of the model simulation can be explicitly stated. On the other hand, the difficulty of defining the best objective function or criterion to be optimized, the difficulty in finding the global optimum when many parameters are involved, and the mutual dependency of and the impossibility to distinguish between the different error sources are the main disadvantages of automatic calibration methods (Refsgaard and Storm, 1996).

The calibration procedure also has to deal with different types of uncertainties. For instance, the model parameter values can be determined directly from direct measurement; however, in many situations the parameters are conceptual representations that do not exist in reality. Therefore, sources of uncertainty can be due to: random or systematic errors in the input data and recorded data used for comparison with the simulated output, errors associated with the use of optimal parameter values and errors due to an incomplete or biased model structure. In addition, Duan et al. (1992) illustrated that, even when simple model structures are used and input data error is minimal, the parameter estimation problem is not trivial. It is constrained by many regions of attraction, many local optima, rough response surfaces with discontinuous derivatives, poor and varying sensitivities of the response surface and nonlinear parameter interaction.

4.4.6 Regionalization of Model Parameters

Regionalization methods imply the transference of model parameters from a basin that is expected to behave similarly to the basin of interest. The similarity measure can be based on spatial proximity, basin attributes or similarity indices (Blöschl, 2005). There are several regionalization techniques; however, nearly all the studies follow the same approach (e.g., Blöschl and Sivapalan, 1995; Abdulla and Lettenmaier, 1997; Fernandez et al., 2000; Littlewood, 2003). Typically, regionalization techniques involve the
definition of relationships between calibrated model parameters and basin attributes. The most common methods are the bi-variate and multivariate regression methods between parameters and basin attributes, and the definition of cluster or groups of basins in hydrologically homogeneous areas where a priori defined parameters can be applied. The difficulty is that the relationships are likely to be weak due to parameter equifinality since many parameter sets might produce similar simulations. For example, Kuczera and Mroczkowski (1998) suggested that the problem of parameter identifiability in conceptual catchment models (where parameters do not exist in nature) is due to the existence of multiple optima and high correlation amongst model parameters. This makes the regionalization of conceptual model parameters in ungauged basins virtually impossible. Hydrological regionalization studies have so far shown limited success and in general depend on the degree of similarity between the basins and on the type of the data used in the regional analysis (Littlewood, 2003). Fernandez et al. (2000) addressed this issue by performing a regional calibration approach where parameters were identified by both minimizing model biases and maximizing goodness of fit of relationships between parameters and basin characteristics. Regional calibration techniques were also performed by Hundecha and Bárdossy (2004) using a semi-distributed conceptual model in 95 sub-basins of the Rhine basin where the coefficients of the relationships between basin attributes and parameters were calibrated rather than the model parameters; however, a limitation of these methods could be the large number of coefficients to be calibrated.

Alternatively, Parijka et al. (2007) proposed an iterative regional calibration method as a solution to the dimensionality of the calibration problem where local information such as streamflow data was combined with regional information such as an a priori distribution of the model parameters from gauged basins in the area in one objective function. Götzinger and Bárdossy (2007) showed that regionalization methods using conditions imposed on the parameters by basin characteristics in distributed conceptual models were the ones that performed best due to the reduction of parameter space. Meier and Blüschl (2004) after comparing several regionalization methods in 308 Austrian basins found that methods based on spatial proximity performed better than regression methods based on basin attributes. Goswami et al. (2007) demonstrated that the regionalization of rainfall-runoff model parameters that were calibrated against regional pooling of streamflow data of twelve basins in France was the one that performed best amongst three methods involving calibration, concluding that the assessment of regional homogeneity and analysis of data are very important for regionalization approaches using calibration methods. In any case, regionalization is of restricted utility in sparsely gauged regions where it can best work where physically based parameters are regionalized over large areas (Dornes et al., 2008a).

4.4.7 Improving Hydrological Model Realism and Performance

One of the consequences of using sophisticated hydrological models or a detailed spatial model discretization is the increase in the number of unknown model parameters with their associated uncertainties that, when propagated through the model, increase the bounds of predictive uncertainty (Atkinson et al., 2003). Hydrological model performance has been hampered both by a lack of information on the catchment and its hydrological characteristics (Sivapalan et al., 2003) and by ubiquitous misconceptions on the operation of the hydrological cycle that have persisted in many hydrological models. A process approach to hydrological prediction was proposed as a method to improve model performance (Pomeroy et al., 2005); however, persistent errors in the process descriptions of many models continue to hamper their performance and require empirical calibration or regionalization of model parameters. These misconceptions cause systematic deviations of model simulations from actual hydrological process operation. Applicability of physically based modeling approaches, which in theory would enable the parameters to be derived from field measurements, has been restrained by heterogeneity of process responses and unknown scale-dependence of parameters. Prior information is thus limited and it is recognized that models and/or parameters must sometimes be identified through inverse modeling (Kavetski et al., 2003). While calibration to streamflow observations has been used to “correct” these deviations, the problems of equifinality cause uncertainty in parameter identification (Beven and Freer, 2001). The reliance on calibration, when streamflow observations are missing, creates a problem for hydrological prediction in ungauged basins; however, when observations are available, the reliance on calibration supports the continuing persistence of deficient modeling approaches. This dilemma is artificial as the need for calibration is due partly to conceptual errors in hydrological model structure, form and resolution and partly to an inability to identify values for certain parameters. Successful hydrological model prediction of multiple endpoints without calibration has been demonstrated using flexible model structures that are appropriate for the catchment being simulated (Pomeroy et al., 2007) and parameters identified using detailed biophysical observations at research catchments (Gelfan et al., 2004; Fang et al., 2010; 2013; Pomeroy et al., 2013).

4.5 Coupled Hydrological Land Surface Schemes

Recent development of coupled hydrological land surface schemes has blurred the lines between land surface and hydrological models, offering complementary advantages of the vertical and horizontal flux focus of each approach, the physical rigor of the LSS and the catchment conceptualization of the hydrological models.

An example of a HLSS is the MESH model. As part of the MEC (Modélisation Environnementale Communautaire) developed by Environment Canada, the MESH (MEC–Surface and Hydrology; Pietroniro et al., 2007) is a stand-alone land surface hydrological model configuration of MEC that couples a LSS (CLASS; refer to Section 4.3.2) with hydrological routing schemes. Representation of spatial heterogeneity is based on a mosaic approach using the Group Response Unit (GRU) concept of hydrological landscape units (Soulis et al., 2000). The routing scheme was developed by Soulis et al. (2000; 2005) and is shown in Figure 4.10. It includes the adaptation of CLASS to sloped terrain drainage functions and its coupling to the routing scheme of the WATFLOOD model (Kouwen, 1993). This involved the
inclusion of physically based transfer functions between the soil column and the micro-drainage system within each GRU. The fundamental drainage element is conceptualized by an assembly of sloped blocks connected to a stream and with the drainage system. Each block has a typical length \( L \) perpendicular to the receiving stream of length \( L_v \). The \( L \) is the distance between the divide of the element (GRU) and the stream and is equal to \( \frac{1}{2} \) of the drainage density, \( D_{ch} \) defined as \( EL_0/A \) where \( L_0 \) is the element area. Thus, a GRU is viewed as a mosaic of slope tiles, with average dimensions \( L \) and \( L_v \) and average slope \( A \), drained by a system of micro-channels. Excess surface water drains to the micro-drainage system as overland flow, \( q_{over} \), represented by Manning's equation.

Figure 4.11 shows how the interflow or horizontal near-surface flow occurs through the soil matrix and the macropore structure, leaving the control volume through the seepage face. The conceptualization interflow as shown here was introduced by Soulsis et al. (2000) and uses a shallow aquifer flow model, assuming that interflow occurs almost entirely when soil moisture is between saturation and field capacity. However, rather than solving the Richard's equations with the added complexity of highly variable hydraulic conductivities in the upper soil layer, the shallow aquifer was forced to fit a simpler power law that relates the total outflow at the seepage face and the average volumetric moisture content stored in a control volume, \( \theta \). This approach assumes an initial condition where the seepage face is fully saturated. With time, the water table drops below the surface of the face and the interflow becomes a mixture of saturated and unsaturated flow. Behind and above the water table, saturation declines in both time and space.

The gravitational movement of water between the soil layers is governed by a finite difference solution of Richard's equation for unsaturated flow in porous media. The relation between horizontal and vertical hydraulic conductivity in slopes is assumed to be less than 10%, so the Dufil–Forschheimer approximation is valid and the \( V_z \) can be calculated using a one-dimension Richard's equation. Variation of the hydraulic conductivity with depth follows an exponential form similar to TOPMODEL, whereas the variation of hydraulic conductivity in unsaturated conditions uses the Clapp–Hornberger soil physics.

River or streamflow routing in MESH is based on a storage routing method originally implemented in the WATFLOOD model (Kouwen, 1988). This is a simple technique since storage is calculated as a function of the outflow alone. The implementation is based on the continuity equation for each river reach where the inflow consists of overland flow, interflow, baseflow and channel flow from all contributing upstream basin elements, whereas outflow is related to the storage through the Manning formula. Channel cross-section area is related to storage by dividing the storage by the channel length, and channel storage is calculated using a relation such that the channel cross-section area is given as a function of drainage area. The roughness coefficient incorporates a channel shape and width-to-depth ratio as well as Manning's \( n \).

Currently, hydrological land surface schemes rely on probabilistic, tiled or grouped representation of sub-grid heterogeneity and assume that river basins can be represented as a sequence of grids. However these simplifications of nature river basin processes are often inadequate for simulating ecosystem interactions and estimating contributing areas for runoff generation and cannot provide realistic simulations of the water cycle at sub-basin scales. Future development of coupled hydrological land surface schemes is expected to include greater spatial representation of sub-grid micro-climates, vegetation patterns, precipitation dynamics, horizontal surface and subsurface hydrological fluxes, and varying contributing areas for runoff generation. These developments will be required for these schemes to provide credible multiscale simulations from basin scales to continents and for linkages with numerical ecosystem models. It is hoped that such developments can spur the further development of quantitative ecohydrology as a biogeochemistry.
References


