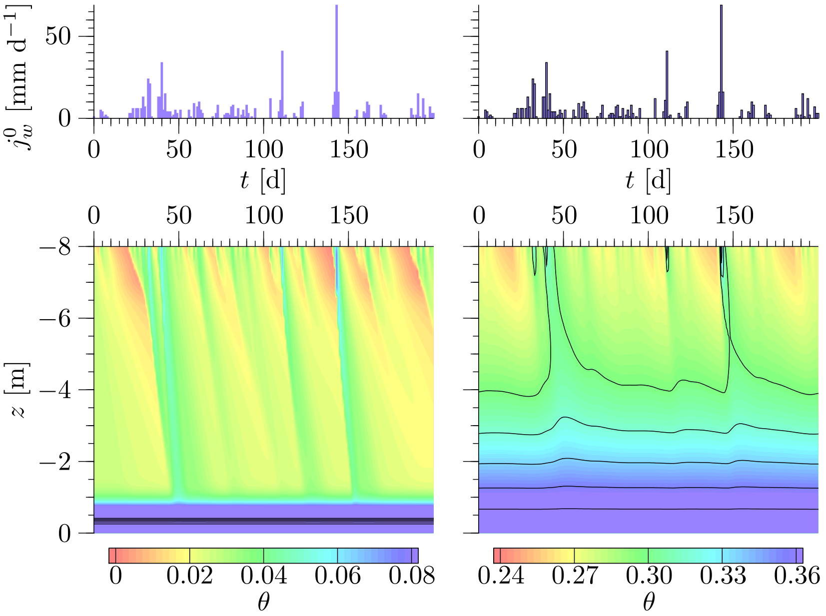


Figure 6.25. Water content $\theta(z, t)$ in sand (**left**) and silt (**right**) for realistic atmospheric forcing derived from a measured rainfall time series (**top**) and an assumed constant evapotranspiration flux of -1 mm d^{-1}. The root model from Figure 6.24 was used and a constant water table at $z = 8$ m was assumed. The black contour lines indicate regions with water saturations where $\Theta = 0.75$, 0.8, 0.85, 0.9, and 0.95. Notice the different color scales for the two graphs.

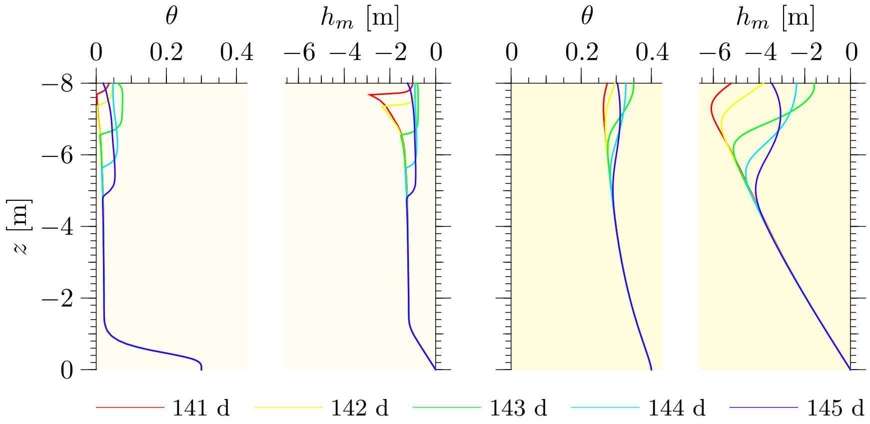


Figure 6.26. Profiles of θ and h_m in the sand (**left**) and silt (**right**) for the most intense rain event during days 141. . . 144. Precipitation rates for these days were 8, 16, 69, and 16 mm, respectively. No rain occurred on day 145.

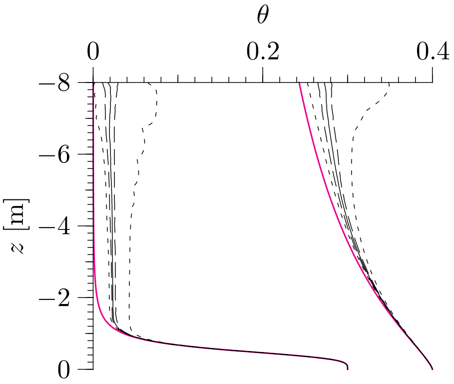


Figure 6.27.
Statistical distribution of θ during the entire period shown in Figure 6.25 for sand and silt. Minimum and maximum are given by dotted lines, median by solid lines, and first and third quartile by dashed lines. Magenta lines represent θ in static equilibrium.

an average flux of 6.48 mm d^{-1} with a maximum of 34 mm d^{-1}. The initial stage of this second group, between days 21 and 30, also nicely illustrates how smaller pulses are overtaken by successively larger ones. This effect becomes particularly clear in Figure 6.26 which focuses on the strongest rain event in this series. In the silt, the phenomenology is basically similar to that in the sand. However, the infiltration fronts are rather unincisive due to the high soil water dispersivity and the associated very rapid spreading.

Regions of high water saturation and their fluctuations are of particular importance for various soil functions like the aeration of plant roots and the redox potential with the associated chemical reactions and microbiological transformations. Prominent chemical reactions are for instance the reduction and mobilization of iron and manganese and the related release of toxins like arsenic. Important microbiological transformations that are bound to anaerobic environments include denitrification and methanogenesis with the associated emission of the greenhouse gases N_2O and CH_4. Regions with high water saturations are indicated by the black contour lines in Figure 6.25. We notice that in the sand, these regions are limited to a thin layer within some 0.5 m of the water table and that they are hardly affected by the infiltration flux. In the silt, on the other hand, high-saturation regions depend strongly on the infiltration flux. They extend several meters above the water table and actually reach right to the surface during phases of heavy precipitation.

Finally, we consider the statistical distribution of $\theta(z)$ (Figure 6.27). For both soils, 50% of the values fall within a rather narrow band around the median value. Since time is sampled uniformly, this also translates into θ being approximately equal to the mean value during half of the time. Again for both soils, excursions to lower water contents are much smaller than to higher values which apparently reflects the relatively seldom heavy showers. The minimum values reached in the profile are higher or, near the water table, equal to those corresponding to static equilibrium. This first of all is a manifestation of the net infiltration regime. Such a regime does not guarantee $\theta \geq \theta^{\text{static}}$, however. Forcing shallower profiles with the same time series actually produces values of θ that are smaller than θ^{static}, particularly in the silt. A further prerequisite is that the time scale on which the system relaxes to static equilibrium after an infiltration event is longer than the time between events. In a deep profile, the water content near the surface is lower on average, hence it must be increased stronger to sustain any given infiltration flux. Consequently, more water is stored, drainage and root water uptake can thus be sustained longer than in a shallower profile.

6.4
Inverse Estimation of Hydraulic Properties

Thus far, we considered the evolution of a known system from a given initial state under a prescribed external forcing. Mathematically, this corresponded to solving a partial differential equation with given parameters, initial and boundary conditions. This is the so called *forward problem*.

A major difficulty with natural porous media are the required hydraulic properties $\theta(h_m)$ and $K(\theta)$ which are very difficult to determine. There exist various experimental methods in the lab (Figure 3.21) as well as in the field to measure $\theta(h_m)$ and $K(\theta)$ rather directly. However, these are invariably difficult and error-prone. An alternative approach is to monitor the water flow in response to an appropriate external forcing and to deduce $\theta(h_m)$ and $K(\theta)$ from this. This is the so called *inverse problem* where, mathematically, PDE, initial and boundary conditions, and the solution are given while the parameters of the PDE are to be found.

As a specific example, consider some hydraulic experiment where the water content θ_i^{meas} is measured at location $\mathbf{x}$ at times $\{t_i,\ i = 1, \ldots, n\}$. Next, consider a parameterization $\mathbf{p}$ of the hydraulic properties. For instance, this may be the Mualem-van Genuchten parameterization (3.45) and (3.56) with $\mathbf{p} = \{\theta_r, \theta_s, \alpha, n, K_0, a\}$, assuming the soil is uniform. Finally, let $\theta^{\text{mod}}(t_i; \mathbf{p})$ be the volumetric water content simulated for location $\mathbf{x}$ and time t_i. Now, solving the inverse problem consists of choosing $\mathbf{p}$ such that measured and simulated values are in optimal agreement. Often, this is achieved by minimizing the *cost function*

$$\chi^2(\mathbf{p}) := \sum_{i=1}^{n} \left[\frac{\theta_i^{\text{meas}} - \theta^{\text{mod}}(t_i; \mathbf{p})}{\sigma_i} \right]^2 , \tag{6.40}$$

where σ_i is the uncertainty of the θ_i^{meas}. A standard procedure for such a minimization is the *Levenberg-Marquardt* algorithm described in Section B.2 in the Appendix.

One can show that the cost function (6.40) is optimal for the case where θ_i^{meas} are statistically independent with the error described by a Gaussian distribution with mean 0 and variance σ_i^2. Then, at the minimum, χ^2 is a random variable with χ^2-distribution, thence the name, with

$$\langle\chi^2_{\text{min}}\rangle = n - m \qquad \text{and} \qquad \text{var}(\chi^2_{\text{min}}) = 2[n-m] \ , \tag{6.41}$$

where n and m are the dimensions of the data array $\boldsymbol{\theta}^{\text{meas}}$ and of the parameter array $\mathbf{p}$, respectively.

Obviously the concept can be generalized easily to accommodate any assortment of data, for instance θ_i, h_{m_i}, and $\mathbf{j}_{w_i}$ for any set of generalized locations $\mathbf{x}_i$, typically locations in time and space. Denote the corresponding array with $\mathbf{d}$ and let C the associated covariance matrix with $C_{ij} = \langle[d_i - \langle d_i\rangle][d_j - \langle d_j\rangle]\rangle$. Also, the parameter vector may be augmented to allow for multiple materials. Finally, let $\mathcal{M}(\mathbf{x};\mathbf{p})$ denote the simulation results for locations $\mathbf{x}$ and parameter array $\mathbf{p}$. The generalized form of (6.40) then becomes

$$\chi^2(\mathbf{p}) = \big[\mathbf{d} - \mathcal{M}(\mathbf{x};\mathbf{p})\big]^{\mathrm{T}}\mathsf{C}^{-1}\big[\mathbf{d} - \mathcal{M}(\mathbf{x};\mathbf{p})\big] \tag{6.42}$$

and may again be minimized with the *Levenberg-Marquardt* algorithm. Before discussing a number of important issues involved in the solution of an inverse problem, we look into two examples.

Example: Multi-Step Outflow Method Based on previous work by *Zachmann et al.* [1981] who considerer draining soil columns, *Kool et al.* [1985] and *Parker et al.* [1985] suggested to perform *one-step outflow* experiments for determining $\theta(h_m)$ and $K(\theta)$. The experiment started with a fully water-saturated soil column. At time 0, the matric head h_m at the lower end was reduced to some value $h_m < 0$ and time series of the resulting outflow of water were recorded. *Toorman et al.* [1992] demonstrated that in addition measuring a time series of the matric head somewhere within the soil colum greatly improved the accuracy of the estimated parameters. Only with this modification the method became useful for a wider range of soils. As a next improvement, *Eching et al.* [1994] and *van Dam et al.* [1994] suggested to reduce the matric head in several smaller steps instead of a single large one and demonstrated that the results were reliable without additional instruments in the soil column. This so called *multi-step outflow* method has since become the standard for estimating $\theta(h_m)$ and $K(\theta)$. A typical setup is shown and explained in Figure 6.28.

Results from a multi-step outflow run with a sand column of 0.1 m height are shown in Figure 6.29. The experiment starts from an initially saturated soil column and a correspondingly positive pressure p_w at the lower end. This is subsequently reduced in discrete steps down to a value that corresponds to a matric head of -0.3 m. In response, water flows out of the column. With each step, the water flow $q_w(t)$ increase rapidly and tails off ever more slowly. Clearly, the rapid increase stems from the region next to the lower boundary which reacts quickly to a change of the pressure. In contrast, the tailing is determined by the equilibration of the entire soil column and dominantly reflects the actual hydraulic conductivity. As expected, this tailing becomes more pronounced as the matric head decreases and the soil becomes drier.

After the desired minimum value of the matric head has been reached, water is allowed to infiltrate back into the soil column be raising p_w, again in discrete

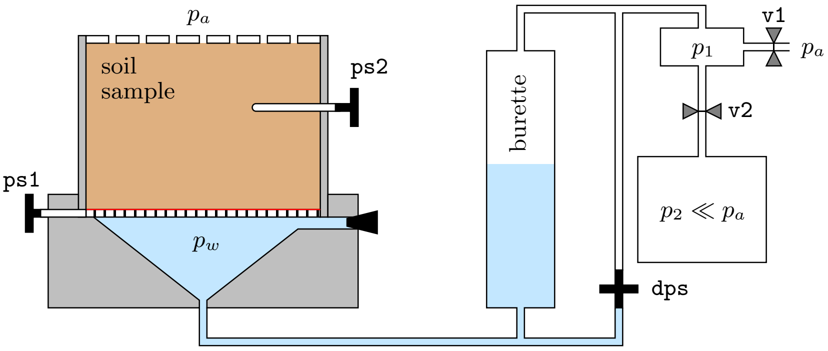


Figure 6.28. Typical setup for multi-step outflow experiment. The soil sample is restrained by two plates to prevent swelling. The upper plate is permeable to air and allows the air pressure within the soil sample to remain at p_a. The lower restraint consists of a thin membrane (red) with very small pores on top of a coarse-grained porous plate. The membrane prevents air from the soil sample to enter the vacuum system and a plate provides the mechanical support. The pressure in the porous plate is measured by sensor `ps1`. Since the hydraulic resistance of the membrane is much larger than that of the plate, `ps1` essentially measures p_w. The signal from `ps1` is used to control p_w, the external forcing, through valves `v1` and `v2`. They connect to two pressure reservoirs, ambient air with p_a and the vacuum container with $p_2 \ll p_a$. The outflow is calculated from the height of the water table in the burette which is measured by the differential pressure sensor `dsp`. As additional quantity, the matric head at some location within the soil sample is measured with pressure sensor `ps2`.

steps. Recalling the hysteresis of the soil water characteristic shown in Figure 3.22 (page 60), we expect that this infiltration follows a path that is different from that of the previous drainage even if the forcing is identical. Indeed, the peaks of the water flow are considerably smaller than those on the drainage branch. Necessarily, we thus require a parameterization of the hydraulic properties that includes hysteresis. A simple extension of the Mualem-van Genuchten parameterization (3.45) and (3.56) was proposed by *Kool and Parker* [1987] who assumed that the scaling parameter α depends on the hysteresis branch while all the other parameters are independent. A motivation for such an assumption may be deduced from Figure 3.20 (page 58) which indicates that the drainage branch is determined by the distribution of bottle-necks while the imbibition branch depends on the cavities. The implication then is that the distribution of volumes associated with the bottle-necks and cavities are related to each other by a simple scaling of the corresponding pore sizes. This scaling is given by α. Indeed, αh_m is the quantity that occurs in the Mualem-van Genuchten parameterization and $[\alpha h_m]^{-1}$ may be interpreted as the dimensionless pore-size thus leading to the above implication.

Once the experimental data are available and the general form of the parameterization is presumed, the corresponding parameters $\mathbf{p}$ may be estimated. Provided that the measurements are independent, (6.42) reduces to

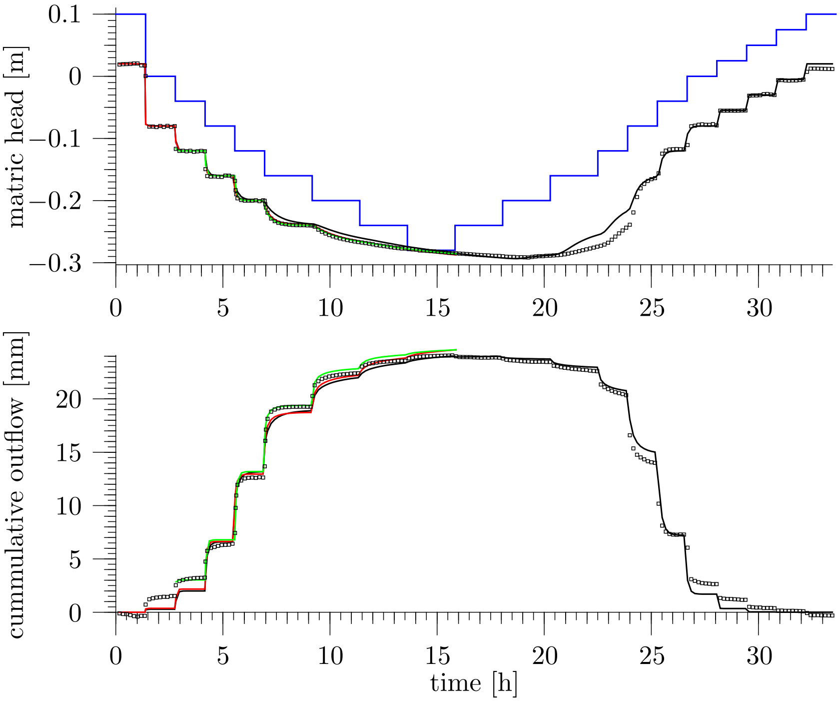


Figure 6.29. Typical result of a multi-step outflow experiment for a coarse-textured sand with grain sizes between 0.25 and 0.63 mm. Symbols show the measured matric head h_m 2 cm below the upper plate (top) and the cumulative outflow Q_w (bottom) in response to the external forcing by the matric head at the lower plate (blue line). Results from inversions using the entire dataset (black), the drainage only (red), and part of the drainage (green) are drawn as lines. The corresponding parameters are given in Table 6.3.

$$\chi^2(\mathbf{p}) = \sum_{i=1}^{n_Q} \frac{[Q_{w_i}^{\text{meas}} - Q_w^{\text{mod}}(t_i;\mathbf{p})]^2}{\sigma_{Q_i}^2} + \sum_{i=1}^{n_h} \frac{[h_{m_i}^{\text{meas}} - h_m^{\text{mod}}(t_i;\mathbf{p})]^2}{\sigma_{h_i}^2} \;, \tag{6.43}$$

where n_Q and n_h are the number of measurements of cumulative outflow and of matric head, respectively. Inversion of the data was done with the code ESPHIM by *T. Zurmühl*. In the following, we consider three subsets of data and their inversions: the complete drainage-imbibition, drainage only, and the dry part of drainage. The resulting parameters are given in Table 6.3 and the corresponding functions are shown in Figure 6.30.

As a preliminary, we notice that the plateau of Q_w that follows the transient phase after each step reflects the soil water characteristic $\theta(h_m)$. Indeed, with the outflow vanished, h_m is constant within the soil column – varying only by some centimeters with the height above the lower plate, here by a maximum of 0.1 m – and the cumulative outflow is proportional to $\theta_s - \theta$. On the other hand, the transient phase is determined by the amount of water that has to leave the soil column, i.e., by the soil water capacity, and by the hydraulic conductivity that

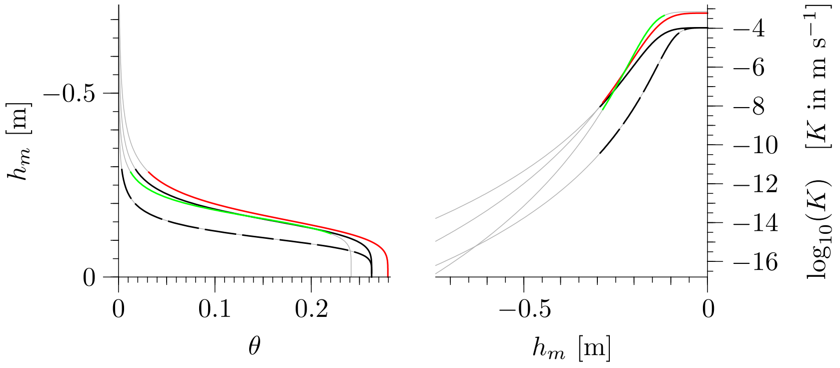


Figure 6.30. Mualem-van Genuchten parameterization with parameter values (Table 6.3) obtained from different subsets of the experiment shown in Figure 6.29. The black curves are from the inversion of a full drainage-imbibition run with dashes indicating imbibition. The red line is for a non-hysteretic formulation inverted from the drainage data alone and the green line is the same model but only for the dry part of drainage. The gray lines show extrapolations of the parameterization into regions not covered by data.

determines the maximal flux. The time scale of the transient phase may thus be expected to be set by the inverse of the soil water diffusivity $D_w = K_w/C_w$ that was defined previously by (6.35).

After these preliminaries, consider the inversion of the entire data set, the black line in Figure 6.29. Apparently, the model is unable to describe the data correctly since already the plateaus of $Q_w(t)$ are not captured correctly. Hence, the parameterization of $\theta(h_m)$ cannot be accurate. Notice that, depending on the matric head, the model produces both, under- and over-predictions for the soil water capacity. Looking at simulated matric head, we also notice rather large deviations from the measurements, mostly in the form of a longer tail. This hints at too small a value of the hydraulic conductivity.

Table 6.3. Optimal values of Mualem-van Genuchten parameters for hydraulic properties $\theta(h_m)$ and $K(\theta)$ shown in Figure 6.30 and obtained from inverting the data from Figure 6.29. Notice that the parameters θ_r and θ_s cannot both be estimated from a multi-step outflow experiment alone, only their difference. The negative sign of α reflects the convention that $h_m < 0$ for bound water.

data used for inversion	$\theta_s - \theta_r$	α_d [m^{-1}]	α_w [m^{-1}]	n	a	K_0 [m s^{-1}]
drainage & imbib.	0.263	−6.33	−9.34	5.32	0.86	$1.05{\cdot}10^{-4}$
drainage only	0.279	−6.18		4.77	2.10	$5.98{\cdot}10^{-4}$
part of drainage	0.241	−6.04		6.38	1.44	$7.08{\cdot}10^{-4}$

Next, we look at the inversion of the outflow data alone, the red curve in Figure 6.29, to see if the problem is with the parameterization of the hysteresis. Clearly the agreement between with the data is much better now, with the matric reproduced almost perfectly. Apparently, the parameterization of the hysteresis is indeed too restrictive. However, the outflow is still not represented accurately. This indicates that the van Genuchten parameterization is not flexible enough for this soil material.

Excluding the wet part of the drainage curve from the inversion, the green curve in Figure 6.29, finally leads to a satisfactory approximation of the measured matric head and of the cumulative outflow. However, the price for this success is a rather small domain of validity.

Finally, we look at the parameters and at the curves shown in Figure 6.30. We first notice that the conductivity K_0 at saturation is comparable for the two inversions based on outflow data alone but that it differs by a factor of 6 to 7 from the value obtained for the complete inversion. However, as Figure 6.30 reveals, the discrepancy quickly vanishes as h_m becomes more negative. Looking at $\theta(h_m)$, we find that the three inversions yield curves that are quite similar for intermediate values of h_m but that they deviate significantly towards more extreme values. We notice that both, $\theta(h_m)$ and $K(h_m)$, become uncertain as the range of actually measured values is left. This is of course typical and well-known for all extrapolations.

The outcome of the inversion exercise with the data from Figure 6.29 is characteristic. It clearly demonstrates the constraints imposed on the solution of an inverse problem by parameterizations with insufficient flexibility. While this issue has been clearly identified its solution is still an active field research. The current most promising approaches indeed favor the more flexible parameterizations offered by spline approximations. With all the uncertainties imposed by an unsuitable parameterization, we realize that the difference between the drainage branch and the imbibition branch is very much larger than the differences between the various inversions. Looking at the outflow data in Figure 6.29, we find that the model actually underestimates the hysteresis. From this, we conclude that neglecting hysteresis in describing water flow through natural soils, at least through coarse-textured ones as studied here, will lead to much worse misrepresentations than those which result from the different inversions.

In closing, we notice the minimum pressure p_w at the lower plate as a fundamental limitation of the multi-step outflow suction method described so far. Since p_w cannot be lower than the vapor pressure $p_w^{\mathrm{vap}} = \mathcal{O}(1\ \mathrm{kPa})$, the lowest potential $\psi_m = p_w - p_a$ available is about $-100\ \mathrm{kJ\,m^{-3}}$. (Recall that energy density is dimensionally equal to pressure and normal air pressure is 100 kPa.) Actually, practical limitations set in much earlier and measurements with $p_w < 80$ kPa are already difficult and the method is mostly applied for $\psi_m > -20\ \mathrm{kJ\,m^{-3}}$. What typically happens is that water degasses below the lower plate and thereby feigns a higher outflow rate and, through the reduced contact area, a lower hydraulic conductivity. An attempt to circumvent this limitation was the multi-step outflow *pressure* method where the setup is modified such that $p_w \gg p_w^{\mathrm{vap}}$ and air pressure p_a in the sample is increased instead. Then, $\psi_m = p_w - p_a$ can be set to arbitrarily low values. Compared to nature, where p_a is always very near to the ambient air pressure, the water phase in a pressure experiment may be in a different state: For low values of ψ_m, p_w is still well above the vapor

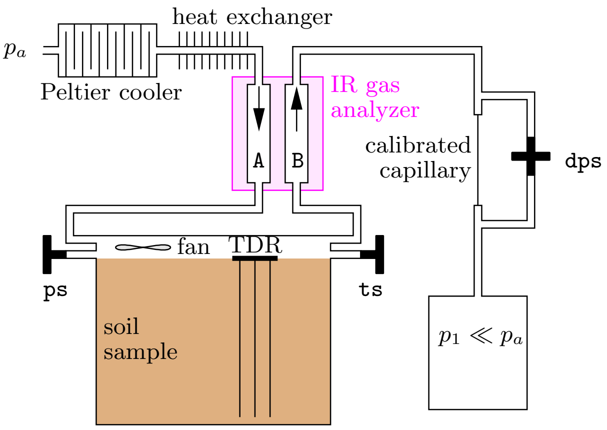


Figure 6.31. Conceptual setup for evaporation experiment. Air whose chemical potential is controlled with a Peltier cooler passes over a soil sample. The head space above the sample is well-mixed by a small fan and pressure and temperature are measured by sensors `ps` and `ts`, respectively. Air flow q_a is measured with the differential pressure sensor `dps` as the pressure drop across the calibrated capillary. Water flow is calculated from q_a and the measured difference of the vapor concentrations in cells `A` and `B`. The TDR (time-domain reflectometry) probe provides additional information on the water content in soil sample. The real setup is considerably more complicated than shown here. There are for instance duplicate Peltier coolers with automatic de-icing and additional gadgets for the periodic online-calibration of the gas analyzer. The entire system is computer-controlled. [*Diploma thesis of K. Schneider, 2005*]

pressure of the bulk phase whereas in nature it would be below. Hence, the significance of pressure measurements is questionable.

Example: Evaporation Method Evaporating water from the soil surface instead of sucking it out from the bottom offers the full range of ψ_m under near-natural conditions. Indeed, Kelvin's equation shows that the water potential in the air phase can be made arbitrarily negative by reducing relative humidity. This equation, introduced and discussed on page 247, relates the ratio between the partial pressure p_v of water in the vapor phase and the vapor pressure at temperature T to the chemical potential ψ_m, which is identical to the matric potential used so far. It may be written as

$$\frac{p_v(\psi_m)}{p_v(0)} = \exp\left(\frac{m_w \psi_m}{\rho_w \mathrm{R} T}\right) , \tag{6.44}$$

where m_w and ρ_w are molar mass and density of water, respectively, and $\mathrm{R} = 8.3144\ \mathrm{J\,mol^{-1}K^{-1}}$ is the universal gas constant.

A conceptual setup of an evaporation experiment is shown in Figure 6.31. The first step is the conditioning of the incoming air. This is accomplished with controlled Peltier coolers which freeze out the water vapor at a prescribed temperature. The lower this temperature, the lower the vapor pressure in the cooler. This determines the total amount of water in the air and with it, after warming back to room temperature, the relative humidity and finally, with (6.44), the potential ψ_m. This conditioned air is then conducted through the thoroughly mixed head space above the soil sample where it takes up water. The uptake is determined by measuring the difference of the water concentrations between inflow and outflow. Since these concentrations are in general very small, an accurate instrument is required, as for instance a differential infrared spectrometer. Together with the volume flow of air, which is obtained from the pressure drop across a calibrated capillary, the water flow out of the soil sample can be calculated. Finally, the hydraulic properties are again obtained by inverting these measurements for a prescribed form of the parameterization.

While the evaporation method makes the full range of ψ_m accessible, it does have a few drawbacks and fundamental limitations. The major drawback is the time it takes to run such an experiment. While the evaporation rate is initially only limited by the air flow through the head space, the hydraulic conductivity of the soil sample becomes the major bottle-neck at later times. As can be gathered from Figure 6.4, this is particularly severe for coarse-textured materials. More fundamental limitations arise from salt accumulation at the surface which modifies the hydraulic properties and from the evolution of cracks already in samples with moderate clay contents. While both effects clearly also occur in natural surface soils, and could there be considered as part of their effective hydraulic properties, this is no longer true for samples that originate from greater depths. Finally, we notice that modeling is much more demanding for an evaporation experiment than for a suction experiment since, certainly in the dry range, the vapor flux in the soil sample must be taken into account.

While inversion is an attractive approach to estimate hydraulic properties it also raises a number of important issues, most of which are active research fields. Here, we will only touch upon a few of them.

An important practical aspect is the required computational effort. Invariably, an inverse problem is much harder to solve than the corresponding forward problem. It starts from an initial parameter array $\mathbf{p}_0$ and then follows a certain path through parameter space to the final optimal array $\mathbf{p}^{\text{opt}}$. The number of steps on that path and the cost of each step depend on the algorithm, but they involve at least one solution of the forward problem for each step and often as many as $n + 1$, where n is the dimension of $\mathbf{p}$. As an indication, a reasonably well-conditioned inverse problem with a good initial guess may demand $\mathcal{O}(10^2 n)$ runs of the forward problem, while an ill-conditioned problem can require well over $10^5 n$ runs.

Inverse problems are notorious for having multiple solutions either because of noisy data or, more serious, because the experimental design does not allow to identify all the parameters. Here, we will only be concerned with algorithmic consequences and discuss further implications below. The difficult thing with multiple solutions is to know whether the problem at hand indeed has this deficiency. A simple approach would be to map the χ^2-surface for the entire physically admissible parameter space. The costs for this are exorbitant, however, even for moderately complicated models and moderately large dimensions of $\mathbf{p}$. A more practical, albeit not completely save, approach is to use different starting points $\mathbf{p}_0$ and to check if they yield different optimal parameters $\mathbf{p}^{\text{opt}}$. Once the inverse problem of interest is found to possess multiple solutions and the experimental design cannot be modified to remove the ambiguities, then special algorithms are used which can find the global minimum. Examples are simulated annealing [*Press et al.* 2002] and genetic algorithms [*Levine* 1996]. These globally optimizing methods are invariably very expensive and should be considered as a last resort for otherwise intractable problems.

Once the optimal parameters have been obtained, the natural question to arise is: Is the model able to fit the data and if so, what is the probability for the goodness of the fit? The first step in addressing this question is to look at the residual, in the notation of (6.42) at $\mathbf{d}-\mathcal{M}(\mathbf{x};\mathbf{p}^{\text{opt}})$, and to check if there occur characteristic structures that could indicate deviations of the data from the model predictions. A more quantitative approach is to compare the actual value of χ^2_{min} with its expectation and variance as given in (6.41). To actually calculate the probability for the goodness of the fit, rather strong assumptions are required, in particular that the data errors indeed follow a Gaussian distribution. Then, the probability $P(\chi^2_{\text{min}})$ can be calculated from the χ^2-distribution with $n-m$ degrees of freedom. As an aside, we comment that while the estimated parameters are not very sensitive to the distribution of measurements errors the explicit statistical quantities are.

Given an appropriate model, we want to know the accuracy of the estimated parameters and the correlation between them. For the highly nonlinear problems considered here, this is a difficult question that can only be addressed adequately with Monte Carlo simulations, which are much more expensive than solving the inverse problem. A much cheaper alternative is available if the distribution of the data errors is Gaussian. Then, one can show [e.g., *Press et al.* 2002] that also the parameters $\mathbf{p}^{\text{opt}}$ follow a Gaussian distribution and that its covariance matrix C equals a^{-1}, where a is the curvature of the χ^2-surface at $\mathbf{p}^{\text{opt}}$. The elements of a are

$$a_{ij} = \frac{1}{2}\frac{\partial^2\chi^2}{\partial p_i \partial p_j} \ . \tag{6.45}$$

Hence, the variance of parameter p_i, under the premiss that all the other parameters are fixed at their optimal value, is

$$\text{var}(p_i) = C_{ii} \tag{6.46}$$

and the coefficient of correlation between parameters p_i and p_j is

$$\rho_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}} \ . \tag{6.47}$$

Some algorithms automatically yield the matrix a because it is also required for the optimization. This is in particular the case for the popular *Levenberg-Marquardt* algorithm. If this is not the case, it can easily be estimated from a finite differences approximation of (6.45). We reemphasize that these handy results are only correct to the extent that (i) the model is a correct description of the data and (ii) the distribution of the data errors is Gaussian.

Very often in soil physics, and in the environmental sciences at large, we are not in the lucky position to have a correct model at hand. More typical is the case where the model is a more or less crude approximation of some complicated process. While the statistical tests will immediately signal the failure of the model, we may still wish to continue and use the model. An example of such a failure is shown in Figure 6.29 where the modeled tails of the outflow peaks are consistently longer than the measured ones. Parameters obtained in such case are "effective" in the sense that all inadequacies of the model – dynamics, parameterizations, initial and boundary conditions – are folded into the parameters. Such results must be used with utmost care: Processes similar to those considered for the inversion may be represented approximately correctly by such a combination of an inadequate model and optimally matched parameters. However, it may fail completely if the parameters are used in a different context. *Bayer et al.* [2005] demonstrated for instance that a multi-step outflow experiment with a weakly heterogeneous soil column could be inverted rather satisfactorily with the (wrong) assumption that the soil was uniform. However, when the column was turned upside down and the multi-step outflow experiment was repeated, to model was not able to predict the data.

We mention in closing that the forward and the inverse problem may be understood as special cases of a more general problem formulation. To this end, we recognize that {*PDE*, *parameters*, *initial conditions*, *boundary conditions*, *solution*} is a redundant set of information about a particular process at a particular site. The general problem consists of deducing the redundant subset from any given complete and nonredundant subset. The important question as to what experimental design is optimal, or at least sufficient, for the identification of a given set of parameters of a given type of model has not yet been addressed to any depth, however.

6.5
Preferential Flow

From a phenomenological perspective, preferential flow refers to nonuniform movement of water and in particular to the more rapid flow in some subregions of the soil. In this sense, flow in heterogeneous formations as illustrated in Figures 6.11–6.13 is already called preferential. There, water is gated by less permeable layers and funneled through more permeable regions. For inclined and cross-bedded layers this may lead to a strong focusing of the flow to narrow channels. Indeed, the term *funnel flow* was coined as an example of preferential flow [*Kung* 1990*a*,*b*]. An extreme example of funnel flow is the capillary barrier sketched in Figure 6.10.

Natural soils are almost always nonuniform. Preferential flow in the above sense may thus be expected as the rule rather than as the exception [*Flury et al.* 1994], the definition is thus not very useful. A more fruitful approach is to consider the dominant flow processes and to distinguish *matrix flow*, which is dominated by capillary forces and described by the Richards equation (6.2), from *preferential flow* that is dominated by other forces. Examples include flow through large and continuous voids, so called macropore flow, instabilities that lead to fingered flow, and flow in incompletely wetted soils.

6.5.1
Macropore Flow

Large continuous voids – desiccation cracks, wormholes, old root channels – are common in the surface layer of most soils as illustrated by Figure 3.6. For voids with radii approaching or even exceeding the capillary length given by (3.9), some 2.7 mm for water in a perfectly wettable soil, gravity becomes the dominating force, flow velocities increase rapidly, and the Navier-Stokes equation has to be used instead of the Buckingham-Darcy flux law. A further, no less fundamental issue is the existence of an REV which is typically impeded by the highly anisotropic shape of the voids and the large extent in at least one direction.

Already for pores with diameters well below the capillary length, a difficulty arises: With the high velocity and correspondingly short transit times the assumption of a local hydraulic equilibrium between the macropore and the surrounding finer-textured matrix is no more warranted. As a consequence, Richards' equation cannot represent the dynamics anymore even if Stokes flow is still a reasonable approximation. For detailed studies of such processes, one has to abandon the continuum approach and to take recourse in pore-scale representations.

Macropore flow has been identified as the culprit behind contaminations of groundwater with pesticides and other agrochemicals which are typically decomposed or at least retained in the biologically active top soil layers. Corresponding observations in drainage water were the first indirect reports on macropore flow [*Lawes et al.* 1882]. Despite the intensive study ever since, understanding of the phenomenon remains qualitative and heuristic representations mark the state of the art.

6.5.2
Flow Instabilities

Displacement of a high-viscosity fluid with an invading low-viscosity fluid has been studied for a long time [*Saffman and Taylor* 1958; *Homsy* 1987]. The respective infiltration front was found to be unstable and to disintegrate into fingers giving rise to beautiful patterns. An apparently similar, albeit

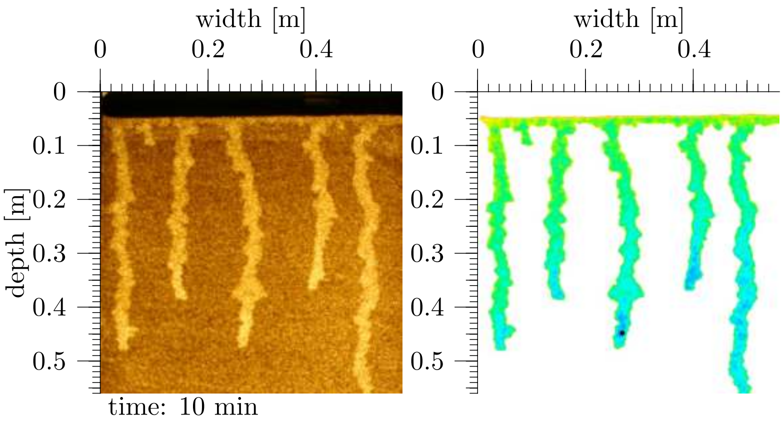


Figure 6.32. Flow instability induced at transition from fine-textured sand (dark top layer) to coarse-textured sand observed with light transmission in a Hele-Shaw cell with a thickness of 3 mm. The sand is initially dry then a uniform and constant infiltration flux is imposed. The image on the right is normalized and calibrated and shows the water saturation. [Image courtesy of F. Rezanezhad]

in detail quite different phenomenon occurs in soils when water infiltrates an initially dry, water-wet, coarse-textured medium (Figure 6.32). Notice in particular that for pure fluids, instability occurs with the "more fluid" invading the "less fluid" while in porous media there are other factors, in particular wettability and permeability of the matrix, which modify the behavior such that also a water front invading dry soil may become unstable. Looking at Figure 6.32, the significance of fingered flow for groundwater recharge and for the transport of solutes through the soil are apparent: (i) A larger fraction of the infiltrating water reaches a depth where it is no longer susceptible to evaporation, hence groundwater recharge increases. (ii) Dissolved substances that enter the soil with the water traverse the biologically active top layers more rapidly, hence there is less time for the immobilization or decomposition of contaminants, and consequently a higher risk of groundwater contamination. (iii) Solutes that reside outside of the flow channels first have to diffuse into them before they get eluted, thus have a considerably higher residence time. Consequently, flow instabilities have been studied extensively, both experimentally [*Glass and Nicholl* 1996; *DiCarlo* 2004] and theoretically [*Eliassi and Glass* 2001].

The first striking feature of unstable flow in porous media is that these so-called *viscous fingers* remain rather stable in the presence of the capillary forces exerted by the dry surroundings. This is in stark contrast to fingers in pure fluids, without a porous matrix, where interfacial tension leads to a rapid coarsening of the finger pattern [*Sharon et al.* 2003]. The crucial ingredients for understanding this are (i) the observed saturation overshoot in the tip of the finger – notice the deeper blue of the finger tips in the right

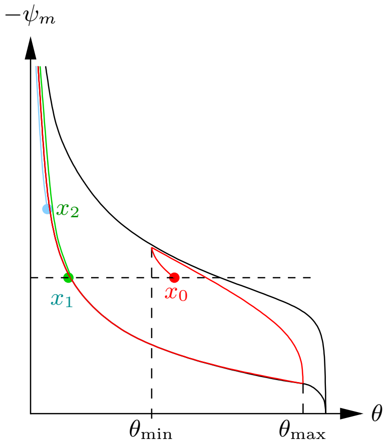


Figure 6.33.
Evolution of hydraulic states during passage of finger tip. At location x_0, where the center of the tip passes through, (θ, ψ_m) moves from the initial very dry state through the maximum $(\theta^{\text{tip}}, \psi_m^{\text{tip}})$ towards $(\theta^{\text{core}}, \psi_m^{\text{core}})$. During the same time, the state at x_1 evolves monotonically towards $(\theta^{\text{fringe}}, \psi_m^{\text{core}})$. As the states at x_0 and x_1 approach the same potential ψ_m^{core} on different hysteresis loops, the radial water flux ceases even though the water contents are quite different. This prevents further rapid radial growth of the finger and stabilizes it. There is still a gradient between x_1 and x_2, however, and the finger continues to expand. This is severely hindered by the low conductivity in this dryer range and by the limited supply of water.

image of Figure 6.32 – and (ii) the hysteresis of the soil-water characteristic as shown for instance in Figure 3.22.

We start out with the fact of saturation overshoot and consider the change of the hydraulic state (θ, ψ_m) at three locations – x_0 on the centerline, x_1 at the outer limit of the core, and x_2 in the fringe – during the passage of the finger tip (Figure 6.33). Initially, the soil is dry hence ψ_m strongly negative and θ very small. With the tip approaching x_0, both θ and ψ_m increase there and move along the imbibition branch of the soil water characteristic. They reach maximal values θ^{tip} and ψ_m^{tip}, respectively, with the center of the tip passing over x_0. Then, water content decreases as is evident qualitatively from Figure 6.32 and was measured in detail by *DiCarlo* [2004]. This decrease brings the state on the desorption branch of the soil water characteristic. On this, (θ, ψ_m) approaches the metastable state $(\theta^{\text{core}}, \psi_m^{\text{core}})$, whose time-scale for change is very much longer than the one for reaching it.

Why does such a metastable state exist? The answer lies in the evolution of the hydraulic state at location x_1, at the outer limit of the core. As the tip passes by, water infiltrates radially, driven by the large hydraulic gradient. Consequently, (θ, ψ_m) moves up the imbibition branch and approaches $(\theta^{\text{core}}, \psi_m^{\text{core}})$ monotonically. With the matric potential at x_0 and x_1 approaching the same value, ψ_m^{core}, the radial gradient decreases and with it the corresponding flux: the finger becomes stabilized.

On a time-scale that is much longer than that of the fingers creation, it continues to expand. This is caused by the large radial gradient in the fringe of the finger, between x_1 and x_2. However, with the water content already low at x_1 and further decreasing in the fringe, hydraulic conductivity decreases very rapidly and with it the radial flux.

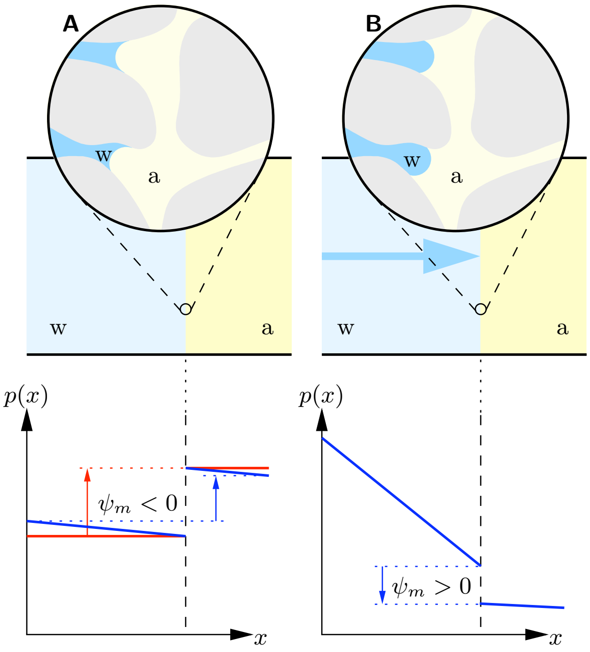


Figure 6.34. Microscopic view of water-air interface (upper row) and pressure (lower row) for infiltration into an initially dry, water-wet medium with different fluxes: (**A**) static equilibrium (red lines) and low flux (blue lines) and (**B**) very high flux. The matric potential $\psi_m = p_w - p_a$ across the interface is related to the mean curvature (Young-Laplace). Notice that the slope of $p(x)$ is related to the flux and to the viscosity of the fluid. Hence it is steeper for water than for air.

The crucial question then is: What causes the saturation overshoot in the finger tip? To answer this, we consider the infiltration front from the microscopic perspective illustrated in Figure 6.34.

First consider the regime sketched in frame **A** of Figure 6.34. In static equilibrium, with the fluids at rest, the pressure is constant within both phases and discontinuous at their interface. The pressure jump, the matric potential $\psi_m = p_w - p_a$ defined in (3.19), is determined by the mean curvature as described by the Young-Laplace equation (3.2). Slowly increasing the pressure gradient in the situation shown in Figure 6.34 at first only leads to a readjustment of the interface but not to a water flux: the interface is "pinned". Further increase leads to a sudden jump of one of the interfaces across the large void and the subsequent rapid invasion of smaller pores ahead until the accumulated pressure is released. This is an instance of the previously encountered *Haines jump*. With the gradient sufficiently high, many temporally overlapping jumps occur at different locations and give rise to a continuous flux. Increasing the gradient further will lead to a proportional increase of the flux as described by the Buckingham-Darcy law. This leads to the flow regime that we considered in most of this chapter so far. Let us call it the *Richards-regime*.

For the phenomena just described, except maybe for the rapid advance during a jump, the interface evolves through quasi-equilibrium states because the time-scale for its adjustment is much shorter than that of the flow. The reason for this is that for slow flow in porous media, interfacial forces are much stronger than viscous forces. The ratio between these two forces is characterized by the *capillary number*

$$\mathrm{Ca} = \frac{\mu v}{\sigma} \, . \tag{6.48}$$

This may be motivated by considering interfacial and viscous forces at a characteristic length ℓ smaller than the capillary length ℓ_c defined by (3.9). We find with (2.11) that the viscous force scales like $\mu[v/\ell]\ell^2$ and with (3.1) that the interfacial force scales like $\sigma\ell$. The ratio of the two yields (6.48). Inserting values for water and assuming the rather high velocity of 1 $\mathrm{mm\,s^{-1}}$ yields $\mathrm{Ca} \approx 10^{-5}$. Hence, the interfacial forces are some five orders of magnitude larger than the viscous forces.

Next, we consider frame **B** of Figure 6.34 which represents a much higher water flux. The velocity here is so high that the water-air interface moves faster than the water-solid interface, despite the fact that water is the wetting fluid (notice Exercise 6.8). As a consequence, the mean curvature of the water-air interface becomes positive. Hence, keeping in mind the very small value of the capillary number, pressure builds up in the water phase such that the matric potential ψ_m becomes positive. As an aside, we notice that this is another instance where ψ_m does not reflect pore size and wettability alone, as is the case for static equilibrium, but that it contains a dynamic component. This is often highlighted by calling ψ_m the *dynamic capillary pressure*. This name is indeed used for all deviations from the static value, irrespective of the origin, which is at times a bit disturbing. We already encountered another example of dynamic capillary pressure in the discussion of (6.7), the water potential that arises with continuous multiphase flow. Notice that these two cases are fundamentally independent of each other. A common feature of all situations where a dynamic component of ψ_m becomes significant is that the relation between θ and ψ_m, as it is given by the soil water characteristic, breaks down because the latter is a relation for static equilibrium. As a consequence, also Richards' equation breaks down since it is based on quasi-static equilibrium states.

Weitz et al. [1987] performed an ingenious experiment to reveal the dependence of ψ_m on velocity. They chose two immiscible liquids with the same viscosity, filled a porous column with one of the fluids and displaced it with the other. They recorded the pressure drop across the column as the interface passed through (Figure 6.35). They found that, depending on the flow velocity, the pressure jump across the interface, i.e., the matric potential ψ_m, was positive, zero, or negative, in accordance with the sketch in Figure 6.34.

We may expect that a higher value of ψ_m also leads to a higher value of the saturation, hence to the observed saturation overshoot in the finger tip. The question then arises: Why does the saturation, and presumably also the matric potential, decrease behind the tip? The additional pressure

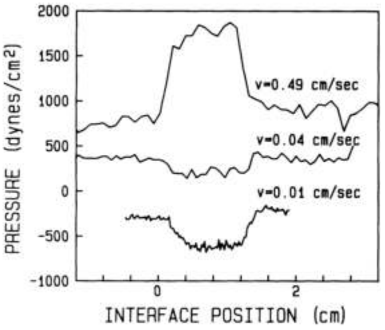


Figure 6.35.
Measured pressure drop across porous column as interface between two immiscible fluids with same viscosity passes through. The dyne is the unit of force in the cgs system and equals 10^{-5} N. (Figure 1 from *Weitz et al.* [1987])

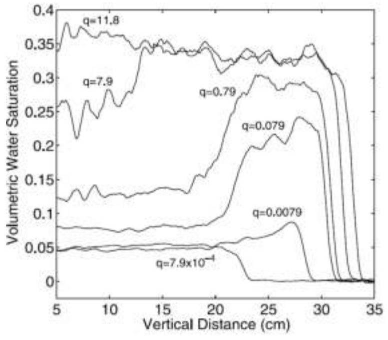


Figure 6.36.
Saturation overshoot during the invasion of water with different fluxes into an initially dry porous column. Fluxes are given in $\mathrm{cm\,min^{-1}}$. (Figure 5 from *DiCarlo* [2004])

is required to overcome the "entrance resistance" posed by the slow wetting of the wetting. Behind the tip, no such resistance is present anymore, the medium is already wet. *DiCarlo* [2004], working with light-transmission in a narrow colum that corresponds to a one-dimensional system, showed that saturation overshoot is strongest for intermediate fluxes (Figure 6.36). This is readily understood: For very low fluxes, flow is essentially in the Richards regime where wetting of the solid surface is fast compared to the speed of the advancing front. Hence not overshoot is generated. Conversely, for very high fluxes the medium is completely saturated already and no overshoot can occur even if the matric potential is positive.

Everything considered so far works beautifully without gravity. What is its impact then? Nothing really essential for the generation and stabilization of viscous fingers. However, it provides a uniform driving force for the flow such that ψ_m far behind the tip remains constant. We may thus assemble the puzzle of viscous fingers in soils: Water flow in the core of the finger is pure gravity flow. In sufficiently coarse-textured media, the corresponding velocities are so high that the speed of wetting of the solid surface at the front becomes limiting. As a consequence water piles up behind the front. From this a hydrostatic pressure results as the surplus water is not hold by capillarity anymore. The pressure increases such that the entrance resistance caused by the slow wetting speed is overcome. The piled-up water, the

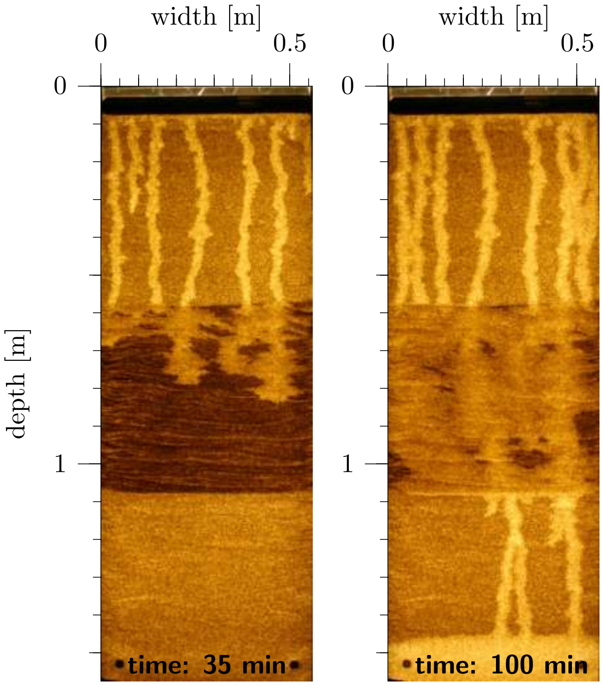


Figure 6.37. Fingered flow in medium with heterogeneous layer at depth interval between 0.56 and 1.05 m sandwiched between two uniform layers. The sole difference between the materials is that the fine-grained fraction was removed before the uniform layers were built up. At the time the image was taken, the flow pattern was still evolving. Eventually, the number of fingers in the lower uniform layer was about the same as in the upper one. The lower end of the cell is closed. A single outlet, the black dot at the lower left corner, allows the water to leave the cell. The outlet in the right corner is closed as may also be deduced from the shape of the water table. Figure 6.32 shows the early stage of the flow pattern with all the flow fingers still in the first uniform layer. [Images courtesy of F. Rezanezhad]

saturation overshoot in the finger's tip, stabilizes the radial growth of the finger through the hysteresis of the soil water characteristic.

Apparently, a prerequisite for flow fingers is a porous medium with large pores, a correspondingly high permeability, and a very steep hydraulic conductivity function. This is typically encountered in coarse-grained granular media (sands) with a rather uniform grain-size distribution. The impact of a wide grain-size distribution is illustrated by Figure 6.37

which shows fingered flow in a layered medium. The middle layer, in the depth interval between 0.56 and 1.05 m, has a grain-size distribution in the range from 0.25 to 1.25 mm. The same material, but previously sieved such that the grains are larger than 0.63 mm, was used to build the adjacent layers. The narrower grain-size distribution of the sieved material also facilitated the building of more uniform layers. In contrast, the middle layer is more heterogeneous with thin fine-grained filaments separating coarser-grained regions. This is a consequence of the inevitable sorting that results from pouring of granular media with a wide grain-size distribution.

Starting out with a dry medium, a constant water flux is imposed at the surface and leads to flow fingers as already found above. Eventually, such a finger hits the finer-grained layer. Since its permeability is much smaller, the finger widens, its velocity decreases and with it also the capillary number Ca. In this experiment, the drop in Ca was sufficient to break down the finger and to initiate a regular infiltration into the middle layer. However, as this infiltration front reaches the underlying uniform layer, the conditions again become favorable for flow fingers and they appear again.

Exercises

6.1 Capillary Bundles Choose parameters for the Brooks-Corey and for the van Genuchten parameterization of the soil water characteristic. Under the premise that the porous medium consists of parallel and circular capillaries, calculate and draw the resulting (i) pore size distribution and (ii) hydraulic conductivity function.

6.2 Stationary Flow, Uniform Soil Consider two uniform soils with different texture, e.g., the sand and silt used in this chapter, with a constant water table at a certain depth. Sketch matric head h_m and water content θ for the dynamic equilibrium during constant infiltration and constant evaporation.

6.3 Stationary Flow, Layered Soil Consider two uniform layers and a constant infiltration flux such that both layers remain unsaturated and that the top layer is the better hydraulic conductor than the bottom layer. What does “better hydraulic conductor” mean for an unsaturated medium? Sketch and explain $h_m(z)$ and $\theta(z)$ across the interface.

6.4 Mulching Gardeners and farmers mulch the soil, i.e., they cover it with straw or bark residues. Similarly, ancient cultures in semi-arid regions, in Yemen for instance, covered their fields with coarse sand. What is or was their motivation?

6.5 Infiltration into Sand and Silt Consider the two materials described in Table 6.1 on page 146. Assume for simplicity that initially, they are both completely dry, i.e., $\theta = 0$ for $t < 0$. What is the propagation velocity of the front that results from the infiltration flux $j_w^0 = 5 \cdot 10^{-6}\ \mathrm{m\,s^{-1}}$?

6.6† Maximum Infiltration (Green-Ampt Problem) Consider a very gently sloping site such that any excess water from a heavy rainfall runs off, but infiltration can still be described in a one-dimensional approximation. Assume the soil to be uniform

and the water content to be constant, $\theta < \theta_s$ for times $t < 0$. Let a heavy rainfall start at $t = 0$ and assume that it is so heavy that runoff sets in immediately. Approximate this by $h_m = 0$ at the soil surface. As a passing remark, $h_m > 0$ in reality since the runoff will lead to a sheet of water with a certain thickness ℓ. However, ℓ will be small, a few millimeters at most, lest erosion sets in and destroys the entire setting. Calculate, as a function of time t, (i) the depth of the infiltration front and (ii) infiltration flux.
Hint: Recall that the infiltration front is self-sharpening and approaches a constant shape after some transition time. Choose a point on that front – half height, 90% height,... – and refer to the corresponding matric head as h_m^f. *Use it to define the driving force. It will turn out to be a parameter that has to be fitted based on experimental data.*

6.7 Multi-Step Outflow Figure 6.29 indicates that the matric head measured by the tensiometer in the soil column continues to decrease for some time after the head at the lower boundary already raises again. Explain!

6.8 Capillary Number and Advance of Viscous Finger With the introduction of the capillary number in (6.48) it was argued that interfacial forces between water and air in a porous medium are orders of magnitudes larger than viscous forces. Nevertheless, in explaining frame B of Figure 6.34 it was stated that the wetting of the matrix, which would appear to be an interfacial phenomenon, is slower than the advance of the air-water interface, which is a viscous phenomenon. How come?

7

Solute Transport

Solutes are a prime issue for the usability of subsurface water, hence also for the availability of freshwater. As pointed out in Chapter 1, water is an almost universal solvent and we typically find a broad spectrum of chemicals in subsurface water. A notorious examples is nitrate, NO_3^-, which stems mostly from fertilizers. It is found to exceed the EU drinking water limit of 50 mg/l in about 15% of the pumping wells in the EU [*European Environment Agency* 2010] with a reported average value of some 15 mg/l in the groundwaters of Western and Central European States [*European Environment Agency* 2007]. In addition, pesticides, again mostly from agricultural sources, and a spread of organics from industrial sources as well as from transportation are found in a large fraction of wells, although mostly below legal limits.

Inputs of chemicals into the soil solution are many. Of global significance are (i) wet and dry deposition from the atmosphere, e.g., for nitrogen and sulfur compounds, (ii) application as agrochemicals, most importantly of nitrogen, phosphorous, and potassium salts as fertilizers but also of organic compounds for pest control, and (iii) natural biogeochemical processes that transform minerals as well as soil organic matter. These extended, so called diffuse sources are distinguished from the more localized point and line sources. Examples for the latter are waste disposal sites, industrial areas, and leaky tanks or sewage pipes.

We notice that solutes are not the only agents to affect the quality of subsurface water. Particularly near the soil surface, around waste water pipes, or beneath open dumps, there typically exists a wide range of bacteria and viruses, which are often more dangerous than most solutes. However, they are filtered out after rather short transport distances through soils.

Transport of solutes through the subsurface – moving from one location to another – in general involves to classes of processes, (i) transport proper and (ii) interactions with themselves other constituents.

Transport proper refers to the displacement of the solute body *with* the mean flow of water (convection) and to spreading *within* the water phase (dispersion). Clearly, a prerequisite for describing convection is a correct solution of the flow problem, which was addressed in Chapters 5 and 6. Dispersion is an even more difficult process since it not only depends on the local mean flow field, but also on its variations in space and time (Chapter 4). At the smallest scale, such variations arise from the porous structure. At increasingly larger scales, the hierarchical architecture of natural porous formations as presented in Section 3.1 lead to increasingly larger variations. Finally, transient flow causes temporal variations that typically lead to a strong dispersion of solute distributions. Phenomenologically, dispersion has two effects: it increases the passage time of a pulse and it lowers the maximum concentration. For the case of a toxic contaminant for instance, this leads to a longer exposure time but also to a lower maximum concentration.

Interactions encompass a wide spectrum of physicochemical and biological processes, sometimes also referred to as biogeochemistry. They range from the preferred location of ionic species near charged surfaces to the precipitation of a mineral phase, from the formation of new compounds between dissolved species to microbial transformations. With respect to transport, interactions may be separated into two groups, (i) sorption and (ii) transformation. *Sorption* encompasses absorption by some other phase or adsorption at some interface. It affects transport by changing the concentration in the water phase and, often more importantly, by changing the effective transport velocity. Adsorption at the solid matrix may reduce it by many orders of magnitude, for instance for heavy metals in soils with average to high pH-values. On the other hand, adsorption at colloids may increase the velocity, relative to that of water, because the larger particles move through larger voids where velocities are higher. This was for instance reported for plutonium [*Kersting et al.* 1999]. *Transformations* also affect the solute concentration by changing its total mass, sometimes permanently. A particularly simple example is radioactive day. More complicated transformations involve the solute and some agent that is localized in the porous medium. These include specific reaction sites at mineral surfaces and bacteria. An example for the latter is microbial denitrification, where NO_3^- is transformed through a number of steps to N_2O and N_2, both of which eventually leave the soil as gases [*Gruber and Galloway* 2008]. Still more complicated, from the perspective of solute transport, are transformations where reactants from different sources start to interact as their respective plumes begin to overlap as is the case in many applications of enhanced natural attenuation approaches to soil and groundwater remediation. The actual mixing of the different plumes is a key aspect for the efficiency of the overall process.

It should be mentioned that, depending on the chemical species and the specific environment, transport mechanisms besides dissolution and transport in water may become dominant. Important pathways are soil erosion, e.g., for phosphorous transport into surface waters, and burrowing animals or ploughing for the mixing of strongly adsorbing chemicals like lead and cadmium. We are not considering these processes here, however.

In this chapter, we focus on solute transport proper, on the movement of a *conservative tracer*, the famous “dyed water molecules”. Such a tracer is an abstraction that allows us to exclude all biogeochemical interactions. Approximations to such a solute are a number of water isotopes like ^{3}HOH and $H_2{}^{16}O$, anions like Cl^- and Br^-, and even some dyes like Fluorescene and Brilliant Blue. As always, there is some fine print to the abstraction of a conservative solute when it comes to real systems. Water molecules can pass across phase boundaries, for instance, or they may and typically do adsorb at mineral surfaces. All this may not happen in exactly the same way neither to the conservative tracer nor to the solute of eventual interest. We do not bother with these higher-order aspects here, however.

As a final introductory comment, notice that while transport with the restriction of a conservative tracer will turn out to be linear, i.e., independent of concentration, whenever the concentration is so low that it does not affect fluid density, this is still a much more difficult issue than fluid flow. First of all, the solution of the flow problem is a mandatory prerequisite. The main reason, however, is that describing solute transport requires correct trajectories of the flow field, which are global properties. In contrast, for flow the gradient of the water potential suffices, which is a local property.

7.1
Transport with Stationary Groundwater Flow

The water content in groundwater is invariant in time. Together with a stationary flow field, solute transport becomes a simple process which closely resembles the situation studied in Section 4.2. This is often a quite reasonable approximation for aquifers as well as for deeper soil layers.

In the following, we will first look into some generic relations that are solely based on the conservation of solute mass. Next, transport in the asymptotic regimes, the near- and the far-field, will be studied in essentially uniform media before a framework is introduced, transfer functions, that facilitates the study of general linear transport processes. Finally, the role of heterogeneity in solute transport is scrutinized.

7.1.1
Generic Relations

Transport almost always involves different regimes between the near- and the far-field. This makes it hard to obtain general formulations for the dynamics and for solutions of transport problems. However, some general relations may be deduced from a formulation of the conservation of solute mass alone. To simplify the following discussion, we focus on a one-dimensional situation. Let x denote the spatial coordinate in the direction of the mean flow.

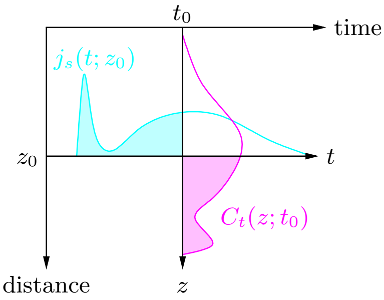


Figure 7.1. Relation between total concentration C_t and solute flux j_s for a solute-pulse input. As shown by (7.4), the shaded areas correspond to the same solute mass.

Relation between Total Concentration and Flux The conservation of mass in a one-dimensional situation may be written as

$$\partial_t C_t + \partial_x j_s = 0 \; . \tag{7.1}$$

Integrating with respect to time from 0 to t then yields

$$C_t(x;t) = C_t(x;0) - \partial_x \int_0^t j_s(\tau;x)\,\mathrm{d}\tau \; . \tag{7.2}$$

Alternatively integrating with respect to space from 0 to x produces

$$j_s(t;x) = j_s(t;0) - \partial_t \int_0^x C_t(\xi;t)\,\mathrm{d}\xi \; . \tag{7.3}$$

Next, consider the special case of the solute-pulse input $j_s^\delta(t;0) = m_0\delta(t)$ into a soil that is initially free of tracer, hence $C_t(x;0) = 0$. Integrating (7.3) from 0 to t then yields (Figure 7.1)

$$\int_0^t j_s^\delta(\tau;x)\,\mathrm{d}\tau = \underbrace{\int_0^t j_s^\delta(\tau;0)\,\mathrm{d}\tau}_{m_0 \text{ for } t>0} - \int_0^x C_t^\delta(\xi;t)\,\mathrm{d}\xi = \int_x^\infty C_t^\delta(\xi;t)\,\mathrm{d}\xi \; , \tag{7.4}$$

where the superscript δ serves as a reminder of the special case.

Relations (7.2)–(7.4) come in handy when a change of perspective from the concentration- to the flux-view is desired. Indeed, we will find that this is sometimes an elegant way to solve apparently difficult problems.

Probability Density Functions for Travel Distance and Travel Time Expressing solute mass by number of molecules and invoking the law of large numbers, we find for the *travel distance pdf*

$$p_x(x;t) = \frac{C_t^\delta(x;t)}{m_0} \; , \tag{7.5}$$

where $p_x(x;t)$ is shorthand for the transition pdf $p_x(x;t|0;0)$ and $C_t^\delta(x;t)$ is the distribution of the total concentration that evolved from the initial concentration $C_t^\delta(x;0) = m_0\delta(x)$ after time t.

In analogy, the travel time pdf is introduced by considering a medium where the initial tracer concentration vanishes, hence $C_t(x;0) = 0$, and total mass m_0 per unit area is injected across the input plane at $x = 0$ by the solute flux $j_s^\delta(t;0) = m_0\delta(t)$. At location x, this leads to the flux $j_s^\delta(t;x)$, which is often referred to as the *breakthrough curve*. Again expressing mass by number of molecules and invoking the law of large numbers, we obtain the travel time pdf

$$p_t(t;x) = \frac{j_s^\delta(t;x)}{m_0} \; . \tag{7.6}$$

We notice that an extension of (7.5)–(7.6) to more dimensions is straightforward even though the resulting formulation for (7.6) will be a bit clumsy since different possible flow directions have to be accounted for.

Using (7.5)–(7.6) in conjunction with (7.2)–(7.3) readily yields the relations

$$p_x(x;t) = -\partial_x \int_0^t p_t(\tau;x)\,\mathrm{d}\tau \quad \text{and} \quad p_t(t;x) = -\partial_t \int_0^x p_x(\xi;t)\,\mathrm{d}\xi \tag{7.7}$$

between the two pdfs.

As will be demonstrated later, in Section 7.1.4, for stationary flow, and at most linearly interacting solutes, transport in an arbitrary medium is completely described by the two pdfs p_x and p_t. It is thus convenient to use them instead of concentrations and fluxes. In addition, it is sometimes useful to work with the corresponding cdfs (cumulative distribution functions)

$$P_x(x;t) := \int_{-\infty}^x p_x(\xi;t)\,\mathrm{d}\xi \qquad \text{and} \qquad P_t(t;x) := \int_0^t p_t(\tau;x)\,\mathrm{d}\tau \tag{7.8}$$

instead of the pdfs.

Remarks The concentration $C_t(x)$ for fixed time t is a useful quantity because the corresponding integral $\int_0^x C_t(\xi)\,\mathrm{d}\xi$ represents the solute mass contained in the depth interval $[0,x]$. In contrast, $C_t(t)$, for fixed location x, is hardly ever useful since its interpretation depends on the water flux and on the solute transport process. Hence, x is a true variable while we may think of t more of as a parameter. To distinguish the two roles, the variables are separated by a semicolon.

Similarly, the solute flux $j_s(t)$ through fixed location x is useful since the integral $\int_0^t j_s(\tau)\,\mathrm{d}\tau$ gives the solute mass that has passed location x during time interval $[0,t]$. In contrast, $j_s(x)$ for fixed t is of limited value. Hence, the roles of t and x are reversed with respect to those in the concentration perspective.

Concentration $C_t(x;t)$ and flux $j_s(t;x)$ are complementary representations of a transport phenomenon, one focusing on space the other one on time (Figure 7.1). They are complementary in the sense that if one is given, either theoretically or measured with sufficient spatial and temporal resolution, the other follows immediately from (7.2) and (7.3), respectively.

While the previous assertion is rather obvious, a question of more practical interest is: Given $C_t(x;t_0)$ at a certain fixed time t_0, e.g., from sampling a tracer experiment, could we calculate the flux at time t_0 or the concentration $C_t(x;t_1)$ at some later time t_1? The answer is no, because we would also need to know $\partial_t C_t(x;t_0)$. The analogous question about the flux – $j_s(t;x_0)$ given at some location x_0 and a projection to some other location wanted – clearly also has a negative answer. This situation only changes if the flux law (4.50) is known or postulated.

The two quantities $C_t(x;t)$ and $j_s(t;x)$ correspond to two natural measuring strategies: obtaining $C_t(x;t_0)$ from soil coring at a fixed time t_0 and $j_s(t;x_0)$ from outflow measurements at a fixed location x_0. There exist a number of other methods whose results can only be interpreted after appropriate assumptions on the form of the flux law are made. Examples are time series of $C_w(x_0;t_i)$ at fixed location x_0 for times t_i, $i = 1,\ldots,n$, as they are obtained from sampling the soil solution with suction cups or from TDR-measurements (time-domain reflectometry).

7.1.2
Near-Field: Stochastic Convection (SC)

For sufficiently short transport distances, we envisage the flow field to be represented by an ensemble of parallel stream-tubes (Figure 4.6 on page 93). Neglecting hydrodynamic dispersion within the tubes, particles move with a constant velocity with the ensemble described by the pdf $g_v(v)$ of the flow field. In analogy to the short-time limit (4.16) of the Taylor-Aris dispersion, the travel distance pdf of a particle then becomes $\delta(x - vt)$ and its travel time pdf $\delta(t - x/v)$. The ensemble of particles is thus described by the travel distance pdf

$$p_x(x;t) = \int_0^\infty \delta(x - vt) g_v(v)\,\mathrm{d}v = \frac{1}{t} g_v\left(\frac{x}{t}\right) \tag{7.9}$$

and the travel time pdf

$$p_t(t;x) = \int_0^\infty \delta(t - x/v) g_v(v)\,\mathrm{d}v = \frac{x}{t^2} g_v\left(\frac{x}{t}\right) . \tag{7.10}$$

The two pdfs evolve in a rather simple way. Consider (x_0, t_0) and (x, t) such that $v = x/t = x_0/t_0$ and look at the spatial distribution of particles as it changes with time. Since the velocity of each particle is constant, their order remains the same. Thus, for any fixed value of v, the same number of particles has traveled farther than vt. Hence, the cumulative distribution function $P_x(x;t)$ depends only on v, that is on the ratio x/t and multiplying both arguments with the same factor t_0/t leaves it invariant,

$$P_x(x;t) = P_x\left(x\frac{t_0}{t};t_0\right) . \tag{7.11}$$

In analogy, by multiplying with x_0/x we also obtain

$$P_t(t;x) = P_t\left(t\frac{x_0}{x};x_0\right) . \tag{7.12}$$

Recalling that $p_\diamond(\diamond) = \partial_\diamond P_\diamond(\diamond)$, with $\diamond \in \{x,t\}$, these scaling relations may also be expressed as

$$p_x(x;t) = \frac{t_0}{t}p_x\left(x\frac{t_0}{t};t_0\right) \quad \text{and} \quad p_t(t;x) = \frac{x_0}{x}p_t\left(t\frac{x_0}{x};x_0\right) , \tag{7.13}$$

respectively. This demonstrates that in the SC-limit, concentration distributions or breakthrough curves are just stretched in space and time with the amplitudes scalled such that mass is conserved.

7.1.3
Far-Field: Convection-Dispersion (CD)

Inserting the flux law (4.52) into (7.1) yields for the far-field in a one-dimensional uniform medium with constant water content and stationary flow the convection-dispersion equation (CDE)

$$\partial_t C_w + v\partial_x C_w - D\partial_{xx} C_w = 0 , \tag{7.14}$$

where we dropped the subscript of D_{eff} and used that θ, v, and D are all constant. First notice that the solute flux $j_s = \theta v C_w - \theta D \partial_x C_w$ also satisfies (7.14). This may be verified by direct substitution. Hence, we also have

$$\partial_t j_s + v\partial_x j_s - D\partial_{xx} j_s = 0 . \tag{7.15}$$

Furthermore, with (7.5)–(7.6), also the travel distance pdf p_x and the travel time pdf p_t satisfy the CDE. Hence

$$\partial_t p_x + v\partial_x p_x - D\partial_{xx} p_x = 0 , \tag{7.16}$$

and

$$\partial_t p_t + v\partial_x p_t - D\partial_{xx} p_t = 0 . \tag{7.17}$$

Before proceeding with solutions of the CDE, we scrutinze the meaning of (7.14) and (7.16) in greater detail, with the gained insight being readily transferable to (7.15) and (7.17). First look at (7.14) which describes the evolution of some initial concentration distribution $C_t(x;0)$ with time. This initial distribution may have an arbitrary shape and (7.14) merely states that it is translated with constant velocity v and smeared in a diffusion-like manner determined by the coefficient D. Hence, it describes the movement of an ensemble of solute particles. In contrast, (7.16) describes the evolution of the pdf $p_x(x;t|0;0)$ for the transition $0 \to x$ during time t, i.e., of the function that describes the evolution of concentration distributions. While the initial condition for (7.14), the concentration at time 0, may be arbitrary, it is fixed for (7.16) as $p_x(x;0|0;0) = \delta(x)$ since a particle cannot move in no time. However, as the available transit time t increases, a particle can move to increasingly greater distances and its location becomes increasingly uncertain. This is described by $p_x(x;t|0;0)$ being translated with constant velocity v and being smeared in a diffusion-like manner determined by the coefficient D. Despite their formal similarity, equations (7.14) and (7.16) hence have quite a different meaning. We mention that an equation for the transition probability is referred to as a *Fokker-Planck* equation.

Travel Distance PDF To calculate an explicit solution for the CDE (7.14) consider a one-dimensional, unbounded, initially tracer-free medium to which, at time $t = 0$ a unit solute mass is added at $x = 0$. The solution of this problem clearly also solves the Fokker-Planck equation (7.16), hence we denote it with p_x. The problem may then be formulated as

$$\begin{aligned} \partial_t p_x + v\partial_x p_x - D\partial_{xx} p_x &= \delta(x)\delta(t) \\ p_x(x;t) &= 0,\ t < 0 \\ \lim_{|x|\to\infty} p_x(x;t) &= 0\ , \end{aligned} \tag{7.18}$$

where the source term $\delta(x)\delta(t)$ describes the solute or probability input. Before proceeding, we notice that, given our insight into the underlying processes, the easiest way to get the solution of this problem would be to guess it – a translated Gaussian – and to verify it by inserting it back. However, since this approach will not work for more complicated processes, we explore the formal approach already for this simple situation.

We choose to employ integral transforms for solving this partial differential equation. Using (A.71), first transform the problem into the Laplace space conjugate to time t and arrive at

$$\begin{aligned} s\widehat{p}_x + v\partial_x \widehat{p}_x - D\partial_{xx}\widehat{p}_x &= \delta(x) \\ \lim_{|x|\to\infty} \widehat{p}_x(x;s) &= 0\ , \end{aligned} \tag{7.19}$$

where $\widehat{p}_x(x;s)$ is the Laplace transform of $p_x(x;t)$ with respect to t, (A.77) was used for transforming $\partial_t p_x$, and (A.74) for $\delta(t)$. Notice that the initial condition, the second equation in (7.18), is directly included in the transform of $\partial_t p_x$. This is a linear, essentially ordinary differential equation for $\widehat{p}_x(x)$ with parameter s and could be solved already at this point. However, we proceed one step further and, using (A.109), transform the problem into the Fourier space conjugate to x to obtain

$$s\widetilde{\widehat{p}}_x + ivk\widetilde{\widehat{p}}_x + Dk^2\widetilde{\widehat{p}}_x = 1\ , \tag{7.20}$$

where $\widetilde{\widetilde{p}}_x(k;s)$ is the Fourier-Laplace transform and i is the imaginary unit. Notice that in analogy to the Laplace transform which directly absorbed the initial condition, the Fourier transform absorbs the boundary condition. Hence, we end up with an algebraic equation that is readily solved and yields

$$\widetilde{\widetilde{p}}_x(k;s) = \frac{1}{s + ivk + Dk^2} \ . \tag{7.21}$$

The solution is not useful in this space, however, and we need to transform it back, again step by step. First employing (LT.4) from Section A.3.5 to invert the Laplace transform with respect to s produces

$$\widetilde{p}_x(k;t) = \exp(-kt[Dk + iv]) \ . \tag{7.22}$$

This solution, still in Fourier space, is already useful for calculating the evolution of the spatial moments. Indeed the n-th spatial moment may be calculated as

$$\langle x^n(t)\rangle = i^n \frac{\partial^n}{\partial k^n} \widetilde{p}_x(k;t)\Big|_{k=0} \ , \tag{7.23}$$

which may be quickly verified by inserting definition (A.109) of the Fourier transform and comparing the result with definition (A.19) of the moments. For expectation and variance of the travel distance, this yields

$$\langle x(t)\rangle = vt \qquad \text{and} \qquad \mathrm{var}\big(x(t)\big) = 2Dt \ , \tag{7.24}$$

where $\mathrm{var}(x) = \langle x^2\rangle - \langle x\rangle^2$. Notice that $\mathrm{var}(x) \propto t$ as expected for a CD-process. Solving (7.24) for the transport parameters again leads to the *method of moments* for estimating the parameters, which was introduced with (4.26)–(4.27).

Staying with the travel distance moments, recall definition (4.15) of the microscopic Peclet number, $\mathrm{pe} = r_0 v/D_m$, and define the *macroscopic Peclet number* in analogy as

$$\mathrm{Pe} = \frac{\langle x\rangle v}{D} = \frac{2\langle x\rangle^2}{\mathrm{var}(x)} \ , \tag{7.25}$$

where $\langle x\rangle$ is chosen as pertinent length, and the parameters v and D are taken from (7.24). Notice that the two definitions are only formally similar but that their significance is quite different. The microscopic Peclet number pe is a constant that characterizes the fundamental processes together with the porous medium. In contrast, the macroscopic Peclet number Pe is a heuristic characterization of the transport phenomenon. Furthermore, it is not a constant but increases linearly with transport distance, testifying to the fact that transport in the far-field becomes increasingly convection-dominated. Still, we may employ it to obtain an order of magnitude estimate for the characteristic *mixing length* λ. Defining λ to be the mean transport distance for $\mathrm{Pe} = 1$, we obtain from (7.25)

$$\lambda = \frac{D}{v} = \frac{\mathrm{var}(x)}{2\langle x\rangle} \ . \tag{7.26}$$

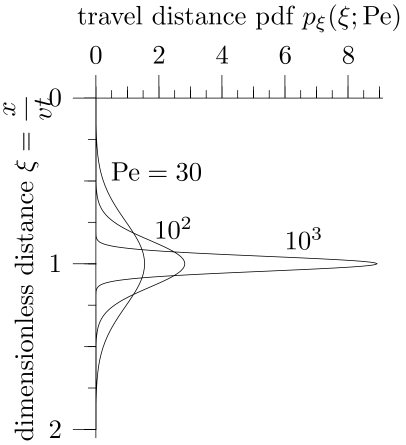


Figure 7.2. Dimensionless travel distance pdf (7.28) for different values of Pe. Distance x is along the direction of the mean flow which is typically horizontal in aquifers and vertical in soils. Notice: Pe is a convolute of effective transport parameters v and D and, with $\langle x\rangle = vt$, of travel time. For a fixed medium, $\mathrm{Pe} \propto t$. In the dimensionless representation shown here, the pdf becomes more localized with time. This is a consequence of $\langle x\rangle$ increasing proportional to time while the spreading increases only proportional to $\sqrt{t}$.

We recognize λ as the *dispersivity* that was already introduced with (4.38). Notice that for a uniform medium with stationary flow, λ is a constant since v and D are constant. Experimentally, λ may be estimated from the right-hand side of (7.26) with the required moments gained from a corresponding tracer experiment. We will find later, in Section 7.1.6, that λ is proportional to the characteristic length of the underlying hydraulic structure that causes the relevant spatial variations of the flow field.

To finally obtain the result in real space, the inverse Fourier transform with respect to k is applied to (7.22). As expected, this yields a Gaussian with mean vt and variance $2Dt$,

$$p_x(x;t) = \frac{1}{2\sqrt{\pi Dt}} \exp\Big(-\frac{[x-vt]^2}{4Dt}\Big) \; . \tag{7.27}$$

It is sometimes useful to have (7.27) in dimensionless form. This is readily found by a transformation of variables. Introducing the dimensionless distance $\xi = x/[vt]$, we write for the probability of finding a particle within interval $\mathrm{d}x$ of x as

$$p_x(x;t)\,\mathrm{d}x = p_x(vt\xi;t)\,vt\mathrm{d}\xi \; . \tag{7.28}$$

Using (7.27) with definition (7.25) of the Peclet number and $\langle x\rangle = vt$ then yields (Figure 7.2)

$$p_\xi(\xi;\mathrm{Pe}) = \sqrt{\frac{\mathrm{Pe}}{4\pi}} \exp\Big(-\frac{\mathrm{Pe}}{4}[\xi-1]^2\Big) \; . \tag{7.29}$$

Travel Time PDF As a preliminary remark, we recall that $C_t(x;t)$ and $j_s(t;x)$ are complementary descriptions of transport phenomenon and the same is true for the corresponding pdfs $p_x(x;t)$ and $p_t(t;x)$. Hence, $p_t(t;x)$ does *not* just equal (7.27) interpreted as a function of t; this would be a severe conceptual error.

Instead of repeating the previous calculations to arrive at the travel time pdf p_t, we choose the alternative route of employing (7.7) and arrive at

$$p_t(t;x) = \frac{x+vt}{4\sqrt{\pi D t^3}} \exp\Big(-\frac{[t-x/v]^2}{4Dt/v^2}\Big) \ . \tag{7.30}$$

For the far-field, the CLT requires p_t to be a Gaussian which is not satisfied by (7.30), however. Indeed, (7.18) does not include a provision that it should only be applicable in the far-field. Consequently, the pdfs (7.27) and (7.30) formally also yield descriptions of the near-field. From a physical perspective, this is nonsense, however! We know from the previous discussion of Taylor-Aris and of hydrodynamic dispersion that the coefficient D is an effective quantity *that only develops as the solute is transported.* For short transport distances, the value of D would be much smaller, right at the source actually equal to the molecular diffusion coefficient, and that it increases, at first linearly in the near-field, to asymptotically reach its final value in the far-field. We notice in passing that this same argument also applies to the travel distance pdf although there – (7.27) and (7.29) – we are not forced to appreciate the issue because the pdf already is Gaussian. Still, near the source and as long as the ensemble of solute particles has not uniformly sampled the flow field, the value of D will be smaller than later in the far-field.

As a consequence of the above discussion, we only consider the asymptotic limit of (7.30) – for large transport distances – as travel time pdf for the CD process. For large distances, and correspondingly large times, $p_t(t;x)$ as given by (7.30) becomes sharply peaked around $t = x/v$. The asymptotic limit is thus found by replacing t with x/v everywhere except in the numerator of the exponential. This yields

$$p_t(t;x) = \sqrt{\frac{v^3}{4\pi D x}} \exp\Big(-\frac{[t-x/v]^2}{4Dx/v^3}\Big) \ , \tag{7.31}$$

which is a Gaussian with mean and variance given by

$$\langle t(x)\rangle = \frac{x}{v} \qquad \text{and} \qquad \mathrm{var}\big(t(x)\big) = \frac{2Dx}{v^3} \ , \tag{7.32}$$

respectively. This again allows the determination of effective parameters v and D from temporal moments – measured in an appropriate tracer experiment – by the method of moments.

The dimensionless form of (7.31),

$$p_\tau(\tau) = \sqrt{\frac{\mathrm{Pe}}{4\pi}} \exp\Big(-\frac{\mathrm{Pe}}{4}[\tau-1]^2\Big) \ , \tag{7.33}$$

is obtained in analogy to (7.28)–(7.29) by introducing the dimensionless travel time $\tau = vt/x$ (Figure 7.3).

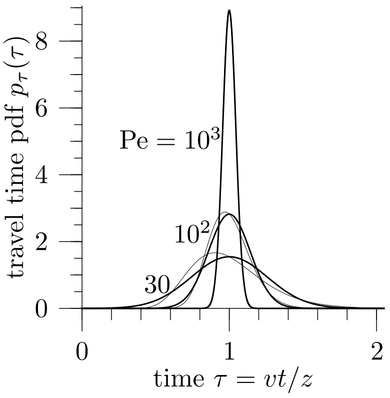


Figure 7.3.
Dimensionless travel time pdf (7.33) for different values of the Peclet number. In analogy to Figure 7.2, Pe for a fixed medium now is proportional to travel distance x and $p_\tau(\tau)$ becomes more localized with increasing x. Thin lines are dimensionless representations of (7.30).

Identification of Transport Regimes Spreading of a solute distribution in the near-field is much faster than in the far-field: the variance of the corresponding pdf is proportional to is strongly affected by the t^2 or x^2 in the SC-limit versus proportional to t and x in the CD-limit. Identification of the transport regime is thus of prime importance. This might sound like a trivial task: just determine the mixing length ℓ and judge if the travel distance is large compared to it, hence if Pe $\gg 1$. Finding ℓ is not a simple matter, however, since it depends on the extent of hydraulic structures and on their hydraulic conductivity. Both of which are difficult, and expensive, to determine. Hence, one often performs a tracer experiment to address the issue directly.

Given an experiment with a highly localized initial tracer distribution what measurements should then be performed to identify the transport regime? A first step would be to determine the tracer distribution $C_t(x;t_1)$ after some time t_1. If this distribution is not Gaussian, then the pulse has not yet reached the far-field. If it is Gaussian, further measurements are required because, as can be deduced from (7.9), for an arbitrary $C_t(x;t_1)$ there exists a velocity pdf $g_v(v)$ such that an SC-process lead to exactly this concentration distribution. Obviously, we need to measure $C_t(x;t_2)$, at some later time $t_2 > t_1$ in order to distinguish between SC and CD. The quantity to look for is the time-dependence of var(x) which is quadratic for SC and linear for CD (Figure 7.4). The analogous argument also holds if the solute flux $j_s(t;x_1)$ is measured at some depth x_1. Again, to identify the regime, fluxes must be measured at least at two depths.

As a practical issue, it should be noted that measured concentrations and fluxes in a real experiment always exhibit quite a large uncertainty that results from the sampling procedure itself and, often more importantly, from the heterogeneity of the aquifer or soil. Hence, it is usually next to impossible to determine if a given distribution is Gaussian or not. This situation automatically demands two measuring depths or times which are separated sufficiently. Since estimation of the variance is much more robust

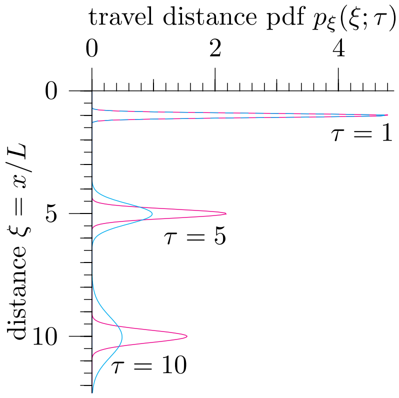


Figure 7.4. Evolution of two travel distance pdfs that are identical at time $\tau = 1$ with an SC- (cyan) and a CD-process (magenta). Parameters are such that at time $\tau = 1$, the pdfs equal the Gaussian (7.29) with $\mathrm{Pe} = 300$.

that estimation of the distribution function, this then also suffices for the identification. Obviously, all this is under the premise that the domain of interest is in fact macroscopically uniform.

7.1.4
Transfer Functions

Up to now, we only considered the transport of initially narrow solute pulses and the respective pdfs. Solute inputs are generally of a much more complicated form, however. These range from the widely separated decaying pulses that result from the application of solid fertilizer, which is dissolved by subsequent rainfall, to the highly irregular flux due to the dry and wet deposition of chemicals from the atmosphere. Fortunately, for the important class of conservative substances, we do not have to solve the transport problem over and over again. Instead, we invoke a central result of the theory of linear systems which ascertains, that a stationary linear system $\mathcal{S}$, which transforms some input function h into the output function $\mathcal{S}(h)$, is completely defined by the output function $\mathcal{S}(\delta)$ which results from the narrow-pulse input δ. In the following, this result is demonstrated for the flux of a conservative solute and then generalized to other situations.

Solute Flux Consider the transport of a conservative solute with a constant flow of water through a porous medium that is uniform and rigid. For simplicity, assume everything to be one-dimensional. Let $j_s(t;x_1)$ be the flux at location x_1 in response to the flux $j_s(t;x_0)$ at location x_0, where x_1 is downstream of x_0 with $\ell = x_1 - x_0$. Finally let $\mathcal{S}_\ell$ be the operator which describes the possibly very complicated transport process from x_0 to x_1 such that $j_s(t;x_1) = \mathcal{S}_\ell(j_s(t;x_0))$. This system is linear and stationary in space and time,

$$\mathcal{S}_\ell\big(\alpha j_s(t-t_0;x)\big) = \alpha j_s(t-t_0;x+\ell) \ , \tag{7.34}$$

because (i) the individual molecules of a conservative solute are transported independently (linearity) and (ii) the water flow is constant and the medium

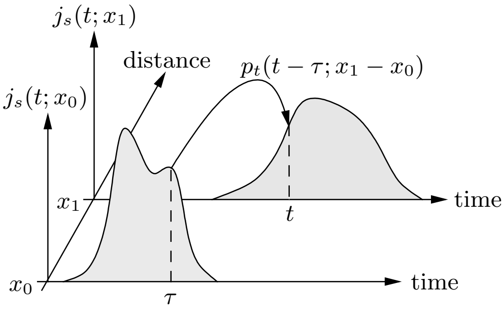


Figure 7.5.
Projection of solute flux $j_s(t;x_0)$ through location x_0 to downstream location $x_1 = x_0 + \ell$ using transfer function $p_t(t;x_1 - x_0)$ as formulated by (7.37).

is rigid and uniform (stationarity). The response of system $\mathcal{S}_\ell$ to the narrow-pulse input $\delta(t)$ is

$$p_t(t;\ell) = \mathcal{S}_\ell(\delta(t)) \;, \tag{7.35}$$

where $p_t(t;\ell)$ is the travel time pdf for travel distance ℓ. Next, decompose the arbitrary input flux $j_s(t;x_0)$ into a succession of narrow pulses as

$$j_s(t;x_0) = \int_{-\infty}^{\infty} j_s(\tau;x_0)\delta(t-\tau)\,\mathrm{d}\tau \;. \tag{7.36}$$

Applying $\mathcal{S}_\ell$, using (7.34) and the fact that $\mathcal{S}_\ell$ operates on t but not on τ inside the integral, yields for the solute flux at distance $x_1 = x_0 + \ell$

$$\begin{aligned} j_s(t;x_1) &= \mathcal{S}_\ell\Big(\int_{-\infty}^{\infty} j_s(\tau;x_0)\delta(t-\tau)\,\mathrm{d}\tau\Big) = \int_{-\infty}^{\infty} j_s(\tau;x_0)\mathcal{S}_\ell\big(\delta(t-\tau)\big)\,\mathrm{d}\tau \\ &= \int_{-\infty}^{\infty} j_s(\tau;x_0)p_t(t-\tau;\ell)\,\mathrm{d}\tau =: j_s(t;x_0) * p_t(t;x_1-x_0) \;, \end{aligned} \tag{7.37}$$

where $*$ is the convolution operator (Figure 7.5). This demonstrates that the travel time pdf $p_t(t;\ell)$ indeed provides a complete description of the transport process: The flux at location x_1 is given by the convolution of the input flux at x_0 and the pdf $p_t(t;x_1-x_0)$. We may also say that (7.37) provides a projection of the flux at x_0 to downstream location x_1.

As a final step, we recognize that for a causal system $p_t(t<0;\ell) = 0$ because there cannot be a response prior to its cause. Thus (7.37) may also be written as

$$j_s(t;\ell) = \int_{-\infty}^{t} j_s(\tau;x_0)p_t(t-\tau;\ell)\,\mathrm{d}\tau \;. \tag{7.38}$$

The interpretation of this is: During the short time interval $[\tau, \tau + \mathrm{d}\tau]$, the number of solute molecules passing x_0 is $N_A j_s(\tau;x_0)\,\mathrm{d}\tau/m_s$, where m_s is the molar mass of the solute. The probability for each of these molecules to reach x_1 at time t is determined by the transition probability density $p_t(t;x_1|\tau;x_0)$. Hence, the number of molecules passing x_1 at time t after they were at x_0 at any time $\tau \le t$ is $\int_{-\infty}^{t} N_A j_s(\tau;x_0)p_t(t;x_1|\tau;x_0)\,\mathrm{d}\tau/m_s$. Replacing the conditional pdf by $p_t(t-\tau;\ell)$ and multiplying with m_s/N_A to get the solute mass flux leads to (7.38).

Using $p_x(x-\xi;t_1-t_0)$, the simplification of the transition probability $p_x(x;t_1|\xi;t_0)$ for a stationary system, we arrive in analogy at

$$C_t(x;t) = \int_{-\infty}^{\infty} C_t(\xi;0)p_x(x-\xi;t)\,\mathrm{d}\xi \ , \tag{7.39}$$

for the total concentration at time t that results from the initial concentration $C_t(x;0)$.

Generalization Transfer functions are strictly applicable only to linear and stationary systems. This, however, includes a much larger class of processes than just conservative solutes in uniform media. It encompasses all transport processes where the solutes undergo at most linear interactions, for instance linear rate-limited exchange with multiple phases. As long as the flow is constant, this class also includes all transport regimes – near-, intermediate-, and far-field – in arbitrarily heterogeneous formations. However, transfer functions are in general neither useful for studying solutes with highly nonlinear interactions, e.g., heavy metals, nor for transport with nonlinear and highly transient flow as is typical in soils near the surface where water flow is driven by strongly fluctuating rainfall.

Some further aspects are worth noticing: (i) Whenever the transport process can be formulated explicitly, the corresponding transfer function can be calculated by solving the equations for a narrow-pulse input. (ii) The transfer function – $p_t(t;\ell)$ or $p_x(x;t)$ – is an implicit description of the transport process and is experimentally accessibly. This is true not just for a single depth but potentially for several or even all depths. Transport may thus be described even if the underlying process is not known or very complicated [*Roth and Jury* 1993]. (iii) While transfer functions are experimentally accessible, we notice that the time required to determine them increases roughly proportionally to the size of the system. With typical flow velocities on the order of 10^2 m y^{-1} for aquifers and 1 m y^{-1} for soils, this approach becomes quickly prohibitive for large systems.

Apparently, there is nothing in (7.35)–(7.38) that limits them to solute transport. Hence, the same line of thought is applicable to any stationary linear system. Different disciplines just tended to give different names to the kernel $p_t(t;\ell)$. It is called *transfer function* (soil physics, electrical engineering), *Green's function* (mathematics), *impulse response function* (mechanical engineering), or *point spread function* (optics), to just name some.

7.1.5
A First Glimpse at Heterogeneity

In an attempt to capture characteristic transport features of natural soils and aquifers, we consider a porous medium that consists of two contrasting materials. One is highly permeable and sustains the externally forced flow of water while the other one is so impermeable that the water phase is practically immobile. Examples where such a model is applicable can be found on all scales and include transport over short distances where interstitial water is relevant, porous carbonate pebbles in a sand matrix, and silt and clay lenses in a sand matrix. In such media, solutes are transported with the water flow in the mobile region while the exchange with the immobile region is predominantly by diffusion.

A detailed description of such a two-region model would involve the geometry with the distribution of the two materials. Water flow and solute transport in the mobile region would then be calculated together with the diffusive exchange between and within the mobile and immobile regions. This leads to a rather complicated formulation which calls for a numerical solution and will be addressed in Section7.1.6. Short of this, we study a blunt approximation where the concentration in each of the two regions is equilibrated in the direction perpendicular to the flow. Thus there are no lateral concentration gradients in the two regions, but the concentration may be discontinuous at the interface between them. The two regions are coupled by a first-order kinetics that is driven by the concentration jump. While such an approximation cannot represent diffusive exchange precisely, it maintains its essential feature of a linear and "soft" coupling between the concentrations in the two regions. We further assume that, at the scale of interest, the two materials are so thoroughly intertwined that the detailed structure disappears after averaging over an appropriate volume. Finally, the medium shall be macroscopically uniform with the water flux and the total water content θ constant. This allows a one-dimensional representation, the so-called mobile-immobile model (MIM) for solute transport [*Coats and Smith* 1964; *van Genuchten and Wierenga* 1976].

Dynamics of the Mobile-Immobile Model Denote the volumetric water content in the mobile region by θ_{m} and in the immobile region by θ_{im}, hence $\theta = \theta_{\mathrm{m}} + \theta_{\mathrm{im}}$. In analogy, let the concentration be C_{m} in the mobile region and C_{im} in the immobile region, with total concentration $C_t = \theta_{\mathrm{m}} C_{\mathrm{m}} + \theta_{\mathrm{im}} C_{\mathrm{im}}$. Next, we assume for simplicity that solute transport in the mobile region is a convection-dispersion process, i.e., we are far from any boundary at the scale of the local mixing length λ. The solute flux then becomes $j_s = \theta_{\mathrm{m}} v C_{\mathrm{m}} - \theta_{\mathrm{m}} D \partial_x C_{\mathrm{m}}$, where $v = j_w / \theta_{\mathrm{m}}$ is the convection velocity and D the local dispersion coefficient. Notice that for a given water flux j_w, the velocity v increases as the mobile water content decreases. Finally, we model the exchange between the two regions by the first-order kinetics $\partial_t C_{\mathrm{im}} = -\omega [C_{\mathrm{im}} - C_{\mathrm{m}}]$ with rate parameter ω. Hence, ω^{-1} is a characteristic time for the exchange between the two regions. Combining all this with the formulation (7.1) for the conservation of mass leads to

$$\begin{aligned} \partial_t C_{\mathrm{m}} + \frac{\theta_{\mathrm{im}}}{\theta_{\mathrm{m}}} \partial_t C_{\mathrm{im}} + v \partial_x C_{\mathrm{m}} - D \partial_{xx} C_{\mathrm{m}} &= 0 \\ \partial_t C_{\mathrm{im}} &= -\omega [C_{\mathrm{im}} - C_{\mathrm{m}}] \, . \end{aligned} \tag{7.40}$$

In the next step, we choose a characteristic length ℓ for the domain of interest and introduce the dimensionless variables

$$\xi := \frac{x}{\ell} \quad \text{and} \quad \tau := \frac{tv}{\ell} \; . \tag{7.41}$$

They transform (7.40) into

$$\begin{aligned} \partial_\tau C_\mathrm{m} + [\mathrm{R}-1]\partial_\tau C_\mathrm{im} + \partial_\xi C_\mathrm{m} - \frac{1}{\mathrm{Pe}}\partial_{\xi\xi} C_\mathrm{m} &= 0 \\ \partial_\tau C_\mathrm{im} &= -\Omega[C_\mathrm{im} - C_\mathrm{m}] \end{aligned} \tag{7.42}$$

and lead to the introduction of the dimensionless quantities

$$\mathrm{Pe} = \frac{v\ell}{D} \; , \qquad \mathrm{R} = 1 + \frac{\theta_\mathrm{im}}{\theta_\mathrm{m}} = \frac{\theta}{\theta_\mathrm{m}} \; , \qquad \Omega = \frac{\omega\ell}{v} \; . \tag{7.43}$$

These are the *macroscopic Peclet number* Pe, the *retardation factor* R, and the *rate parameter* Ω. Notice that Ω may be interpreted as the ratio between the mean travel time ℓ/v in the mobile region and the relaxation time ω^{-1} for the equilibration between mobile and immobile region.

First consider the case $\Omega \to \infty$, i.e., instantaneous equilibrium. Apparently, this enforces $C_\mathrm{m} = C_\mathrm{im}$ and reduces (7.42) to

$$\mathrm{R}\partial_\tau C_\mathrm{m} + \partial_\xi C_\mathrm{m} - \frac{1}{\mathrm{Pe}}\partial_{\xi\xi} C_\mathrm{m} = 0 \; . \tag{7.44}$$

We recognize the similarity with the convection-dispersion equation (7.14), which, expressed with dimensionless variables, reads

$$\partial_\tau C_w + \partial_\xi C_w - \frac{1}{\mathrm{Pe}}\partial_{\xi\xi} C_w = 0 \; . \tag{7.45}$$

Introducing the *retarded time* $\tau' = \tau/\mathrm{R}$ transforms the two equations into each other and demonstrates that the MIM model in the limit of an instantaneous interaction, i.e., for $t \gg \omega^{-1}$ reduces to a retarded convection-dispersion model. Hence solutions of the latter are also solutions of the former provided time is replaced by its retarded analog, τ/R or t/R. The reason for this is simple: With an instantaneous interaction the particles are equally distributed in the two regions. The fraction of time they spend in the mobile region, where they are subject to transport, is only $\theta_\mathrm{m}/[\theta_\mathrm{m} + \theta_\mathrm{im}]$. Thus they require on average more time to travel a certain distance, exactly by the factor of R more.

Travel Time PDF We calculate the travel time pdf for the MIM-model in analogy to (7.18)–(7.19) and consider the narrow pulse input $j_s^\delta(\tau;0) = \delta(\tau)$ into an initially tracer-free soil. Transforming (7.42) into Laplace space then yields

$$\begin{aligned} s\widehat{C}_\mathrm{m} + s[\mathrm{R}-1]\widehat{C}_\mathrm{im} + \partial_\xi \widehat{C}_\mathrm{m} - \frac{1}{\mathrm{Pe}}\partial_{\xi\xi} \widehat{C}_\mathrm{m} &= 0 \\ s\widehat{C}_\mathrm{im} &= -\Omega[\widehat{C}_\mathrm{im} - \widehat{C}_\mathrm{m}] \end{aligned} \tag{7.46}$$

and, inserting $\widehat{C}_{\mathrm{im}}$ from the second equation into the first one,

$$\Big[s + [\mathrm{R}-1]\frac{s\Omega}{s+\Omega}\Big]\widehat{C}_{\mathrm{m}} + \partial_\xi \widehat{C}_{\mathrm{m}} - \frac{1}{\mathrm{Pe}}\partial_{\xi\xi}\widehat{C}_{\mathrm{m}} = 0 \; . \tag{7.47}$$

Notice that in dimensionless coordinates $j_s = C_{\mathrm{m}} - \mathrm{Pe}^{-1}\partial_\xi C_{\mathrm{m}}$, hence $\widehat{j}_s = \widehat{C}_{\mathrm{m}} - \mathrm{Pe}^{-1}\partial_\xi \widehat{C}_{\mathrm{m}}$. Thus, if $\widehat{C}_{\mathrm{m}}$ satisfies (7.47) so does $\widehat{j}_s$ and, taking into account the specified boundary flux, also the transform $\widehat{p}_t$ of the travel time pdf. We may thus also formulate

$$\begin{aligned}\Big[s + [\mathrm{R}-1]\frac{s\Omega}{s+\Omega}\Big]\widehat{p}_t + \partial_\xi \widehat{p}_t - \frac{1}{\mathrm{Pe}}\partial_{\xi\xi}\widehat{p}_t &= 0\\ \widehat{p}_t(s;0) = 1, \qquad \widehat{p}_t(s;\infty) &= 0 \; .\end{aligned} \tag{7.48}$$

This ordinary differential equation has the solution

$$\widehat{p}_t(s;\xi) = \exp\Big(\frac{\xi\mathrm{Pe}}{2}\Big[1 - \sqrt{1 + \frac{4\alpha(s)}{\mathrm{Pe}}}\Big]\Big) \tag{7.49}$$

with

$$\alpha(s) = s + [\mathrm{R}-1]\frac{s\Omega}{s+\Omega} \; . \tag{7.50}$$

Similar to (7.22)–(7.24), where we calculated the moments of the travel distance in Fourier space, we now obtain the travel time moments directly in Laplace space from

$$\langle t^n \rangle = [-1]^n \frac{\partial^n \widehat{p}_t(s)}{\partial s^n}\Big|_{s=0} \; . \tag{7.51}$$

This yields for the first two moments of the dimensionless travel time

$$\langle \tau \rangle = \mathrm{R}\xi \qquad \text{and} \qquad \mathrm{var}(\tau) = 2\Big[\frac{\mathrm{R}^2}{\mathrm{Pe}} + \frac{\mathrm{R}-1}{\Omega}\Big]\xi \tag{7.52}$$

or in dimensional form

$$\langle t \rangle = \mathrm{R}\frac{x}{v} \qquad \text{and} \qquad \mathrm{var}(t) = \frac{2\mathrm{R}^2}{v^3}\Big[D + \frac{\mathrm{R}-1}{\mathrm{R}}\frac{v^2}{\omega}\Big]x \; . \tag{7.53}$$

To interpret this, recall (i) the travel time moments of the pure convection-dispersion given by (7.32) and (ii) the retarded time $t' = t/R$ introduced for the case of instantaneous equilibrium. Apparently, the mean travel time is independent of the rate parameter ω! Hence the expectation of the retarded time, $\langle t' \rangle = \langle t/\mathrm{R} \rangle = x/v$, equals that of the pure convection-dispersion. However, the variance of the travel time depends on ω. First notice that $\mathrm{var}(t) \propto x$, hence we may identify the MIM model as an effective convection-dispersion process. Next, we compare $\mathrm{var}(t') = \mathrm{var}(t/\mathrm{R}) = \mathrm{var}(t)/\mathrm{R}^2$ with the corresponding expression in (7.32) and are led to define the effective dispersion coefficient

$$D_{\mathrm{eff}} := D + \frac{\mathrm{R}-1}{\mathrm{R}}\frac{v^2}{\omega} \; . \tag{7.54}$$

Dispersion in this model thus stems from two independent sources: from the hydrodynamic dispersion in the mobile region, represented by D, and from the interaction, represented by R and ω. The latter contribution is noteworthy because (i) it diverges as $\omega \to 0$ and (ii) its dependence on v identifies it as a "diffusion-like" dispersion similar to the one encountered in Taylor-Aris dispersion. This makes perfect sense since MIM exchange indeed is "diffusion-like". Hence, whenever the second term in (7.54) becomes important, the dispersivity introduced in (4.38) becomes a rather useless quantity since it is no more constant, not even approximately. For hydrodynamic dispersion to dominate also in the MIM model, we obviously require $D \gg [\mathrm{R}-1]v^2/[\mathrm{R}\omega]$ or, for this case using $\lambda = D/V$,

$$\lambda \gg \frac{\mathrm{R}-1}{\mathrm{R}}\frac{v}{\omega} \ . \tag{7.55}$$

This states that the distance traveled during the characteristic time $[\mathrm{R}-1]/[\mathrm{R}\omega]$ for the equilibration of the interaction is much smaller than the hydrodynamic mixing distance λ.

Until now, we only considered the travel time pdf in Laplace space. Inversion of (7.49) with (7.50) is not trivial. It is greatly facilitated by the generalized convolution theorem (A.103) – as shown in (A.104)–(A.108) – and yields

$$p_t(\tau;\xi) = p_t^{\mathrm{c}}(\tau;\xi)\mathcal{A}(\tau) + \int_0^\tau p_t^{\mathrm{c}}(\tau';\xi)\mathcal{A}(\tau')\mathcal{T}(\tau',\tau-\tau')\mathcal{B}(\tau-\tau')\,\mathrm{d}\tau' \tag{7.56}$$

with

$$\begin{aligned}\mathcal{A}(\tau) &= \exp(-\Omega[\mathrm{R}-1]\tau) \ ,\\ \mathcal{B}(\tau) &= \exp(-\Omega\tau) \ ,\\ \mathcal{T}(\tau_1,\tau_2) &= \Omega\sqrt{[\mathrm{R}-1]\tau_1/\tau_2}\,\mathrm{I}_1\big(2\Omega\sqrt{[\mathrm{R}-1]\tau_1\tau_2}\big) \ ,\end{aligned} \tag{7.57}$$

where $p_t^{\mathrm{c}}(\tau;\xi)$ is the travel time pdf for the mobile region – for instance obtained from a tracer that is conservative with respect to the interaction – and I_1 is the modified Bessel function of the first kind of order 1. Although this solution appears rather complicated we recognize some important features. Most importantly: the travel time pdf of the interacting solute can be expressed in terms of the pdf p_t^{c} of the conservative tracer. However, in contrast to the situation with instantaneous equilibration, the two pdfs are not related by a simple scaling but by a more general linear transformation whose form depends on R and Ω.

The pdf p_t^{c} consists of two terms which constitute two fractions of the solute mass, depending on whether a molecule encountered the immobile region or not. The first term is proportional to the pdf of the conservative tracer, hence represents the molecules that never entered the immobile region. The amplitude of this fast fraction decreases exponentially with time, given by the

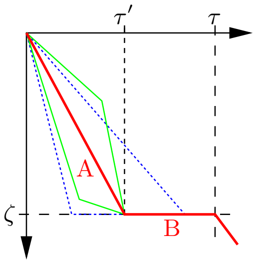


Figure 7.6.
Conceptual space-time diagram for MIM model. To travel distance ξ during time τ, a molecule travels to ξ during time τ' (A), transits into the immobile region where it stays for time $\tau - \tau'$ (B), and moves back into the mobile region and immediately crosses depth ξ. Thin lines represent alternative paths for the transition $(0,0) \rightarrow (\xi, \tau)$ with the same residence time τ' in the mobile region (green) and with different τ' (blue).

factor $\mathcal{A}(t) = \exp(-\Omega[\mathrm{R}-1]\tau)$, hence the fast fraction decreases exponentially with transport distance (see Exercise 7.5).

The second term results from molecules that encountered the immobile region. Formally, this is written as an integral transform of p_t^{c} with kernel $\mathcal{A}(\tau')\mathcal{T}(\tau, \tau - \tau')\mathcal{B}(\tau - \tau')$ with the integration accounting for all possible times $0 < \tau' < \tau$ a particular molecule spent in the mobile region. Since it is irrelevant whether the time spent in a region is split into several intervals or not, the complex sequence of transitions between the two regions is replaced by a two step process (Figure 7.6): (A) the molecule travels to ξ during time τ' without seeing the immobile region and (B) immediately before crossing the monitoring plane at ξ, it transits into the immobile region and stays there for time $\tau - \tau'$. Upon release from the immobile region, it immediately crosses the monitoring plane. We now look at the probabilities associated with each of these steps. From the definition of p_t^{c}, we have that $p_t^{\mathrm{c}}(\tau'; \xi)\,\mathrm{d}\tau'$ is the probability for a molecule to travel distance ξ during time interval $[\tau', \tau' + \mathrm{d}\tau']$, if the interaction is absent. With interaction, this probability is reduced by the factor $\mathcal{A}(\tau')$ which accounts for transitions to the immobile region that have already occurred by time τ'. Thus, $p_t^{\mathrm{c}}(\tau'; \xi)\mathcal{A}(\tau')\,\mathrm{d}\tau'$ is the probability for a molecule to do step A and thus to be available for a transition to the immobile region. In analogy, the probability density for a molecule which entered the immobile region to be still there after time τ is $\mathcal{B}(\tau)$. With probability $1 - \mathcal{B}(\tau)$ it has made the transition back to the mobile region at earlier times. Since the total travel time to distance ξ must be τ, and τ' has already passed during the first step, the probability density for a molecule to be available for the transition to the mobile region at the required time is $\mathcal{B}(t - \tau')$. Clearly, not all the molecules that are available for a transition, either to the immobile or to the mobile region, will also make it. The probability density for the transitions is described by the function $\mathcal{T}(\tau', t - \tau')$, which depends on the residence times in the two regions. The probability for a molecule to make the transition $(0,0) \rightarrow (\xi, \tau)$ with the residence time in the mobile region in the interval $[\tau', \tau' + \mathrm{d}\tau']$ is thus $p_t^{\mathrm{c}}(\tau'; \xi)\mathcal{A}(\tau')\mathcal{T}(\tau', t - \tau')\mathcal{B}(t - \tau')\,\mathrm{d}\tau'$. Integrating over all possible times τ' in the mobile region yields the second term in (7.56).

The dimensionless travel time pdf (7.56) and the corresponding cdf are illustrated in Figure 7.7 for different values of the rate parameter Ω. For

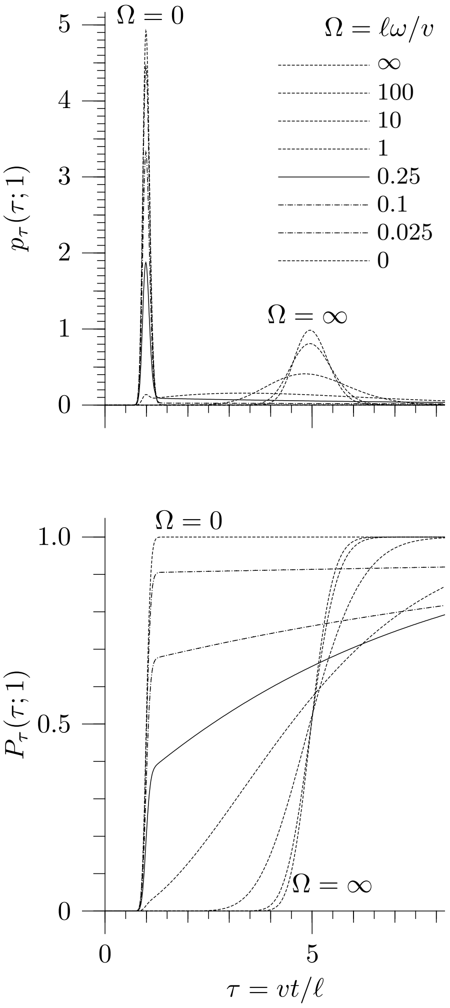


Figure 7.7.
Travel time pdf $p_t(\tau;\xi)$ and cdf $P_t(\tau;\xi)$ for distance ℓ, $\xi = 1$, for MIM model with different rate parameters $\Omega = \omega\ell/v$. Notice that $\Omega[\mathrm{R} - 1]$ is the inverse of the characteristic residence time of a solute molecule in the mobile region with respect to the exchange with the immobile region. Peclet number and retardation factor are fixed with $\mathrm{Pe} = v\ell/D = 300$ and $\mathrm{R} = \theta/\theta_\mathrm{m} = 5$. While p_t and P_t are mathematically equivalent, they are related by (7.8), this is not necessarily the case experimentally when minimal detection limit, finite accuracy, and temporal resolution are taken into account. Experimental limitations become more manifest as Ω decreases.

$\Omega = 0$, the mobile and the immobile region are decoupled. Hence, only the mobile region is active in the transport which is described by the convection-dispersion equation (7.45). Expectation and variance of the travel time for distance ξ then become $\langle\tau\rangle = \xi$ and $\mathrm{var}(\tau) = 2\xi/\mathrm{Pe}$, respectively. Notice that this appears to contradict (7.53) which yields $\langle\tau\rangle = \mathrm{R}\xi$ and $\lim_{\Omega\to 0}\mathrm{var}(\tau) = \infty$. This manifests the singularity that results from the increasing long and thin tail of p_t as $\Omega \to 0$.

As a preliminary to discussing the case $\Omega > 0$, we notice that $\Omega = \ell\omega/v$ is a conglomerate of rate parameter ω, flow velocity v, and travel distance ℓ. Hence, the range of possible values reflects a range of situations in the dimensional realm.

For $0 < \Omega \ll 1$, the two regions are very weakly coupled and the probability for a transition between them is very small. Correspondingly, only a small fraction of the solute mass will be transferred into the immobile region where it will stay for a long time, however. This leads to the observed very long tail of p_t.

As Ω increases, the fast peak at $\tau = \xi$ rapidly decreases since the chance for a molecule to pass entirely through the mobile region vanishes exponentially with Ω. Instead, a new peak occurs at $\tau = \mathrm{R}\xi$, at the expected travel time, becomes more pronounced for $\Omega \gg 1$, and asymptotically approaches the Gaussian shape expected from (7.44).

We notice in passing that the singularity at $\Omega = 0$ found above is more than a mathematical peculiarity and has direct experimental consequences. First consider the travel time pdf p_t which, experimentally, is determined from measuring the solute flux after travel distance ℓ. Since every instrument has a finite detection limit, there exists a critical value of Ω below which the tail will not be detected anymore and only the fast peak will be manifest. As a consequence, estimates of the travel time moments will be completely wrong – the expectation by a factor R, the variance by much more – and thus also the deduced transport parameters. At a first glance, the problem could be solved by measuring P_t instead of p_t, since it records the mass fraction in the remaining tail,

$$P_t(\tau;\xi) = \int_0^{\tau} p_t(\tau';\xi)\,\mathrm{d}\tau' = 1 - \int_{\tau}^{\infty} p_t(\tau';\xi)\,\mathrm{d}\tau' = 1 - \int_{\tau}^{\infty} \frac{j_s^{\delta}(\tau';\xi)}{m_0}\,\mathrm{d}\tau' \,. \tag{7.58}$$

However, the corresponding measurement is limited by the finite absolute accuracy with which the solute flux can be determined. Which one of the methods is to be preferred depends on the specifications of the instruments.

Travel Distance PDF In a MIM model, the travel distance pdf $p_x(x;t)$ depends sensitively on the initial distribution of the tracer between the two regions, particularly for small values of the rate parameter ω. Here, we consider the experiment performed to obtain $p_t(t;x)$ – the flux input $j_(t;0) = m_0\delta(t)$ through the surface at $x = 0$ – but now look at it from the perspective of the travel distance pdf $p_x(x;t)$. The actual calculations are most efficiently done in Laplace space by inserting (7.49) with (7.50) into the Laplace transform of (7.2). The result is then transformed back to normal space by a numerical Laplace inversion. In the following, we only look at graphical representations of the solution.

Clearly, the phenomenology observed with the travel time pdf is also manifest in the travel distance pdf (Figure 7.8). For $\Omega \ll 1$, most of the solute mass is contained in the fast peak which shows the greatest travel distance. With increasing Ω, the mass is transferred into the tail which, for $\Omega \gg 1$, approaches a single peak that corresponds to equilibrium interaction. Such a transfer becomes manifest at any one particular site, with ω fixed, when we follow the solute pulse in time as is illustrated in Figure 7.9. For

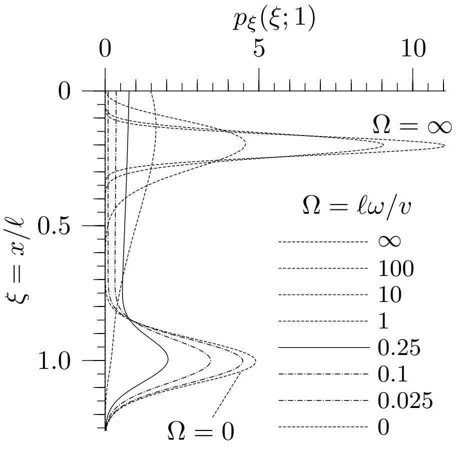


Figure 7.8. Travel distance pdf $p_x(\xi;\tau)$ after time $\tau = 1$, for MIM model with different rate parameters $\Omega = \omega\ell/v$. Peclet number and retardation factor are fixed and equal to the values used in Figure 7.7 (Pe = 300 and R = 5).

times $\tau \ll 1$ which are much shorter than the characteristic residence time in the mobile region, a large fraction of the solute mass has not yet encountered the immobile region and thus travels with the mean velocity v of the mobile region. The small fraction that traverses into the immobile region travels much slower, eventually with v/R, and thus leads to the long tail all the way up to the surface at $\xi = 0$. For times $\tau \gg 1$, the fast pulse disappears, the tail detaches from the surface, and a Gaussian pulse evolves.

We notice that the apparent velocity $\langle\xi\rangle/\tau$ and the apparent dispersion coefficient $\mathrm{var}(\xi)/[2\tau]$ are not constant as they are when we calculate them from the moments (7.53) of p_t. Instead the velocity decreases from v to v/R and, as a quick calculation shows, the dispersion coefficient increases.

Figure 4.8 on page 98 shows an experiment that may be understood in terms of mobile and immobile regions. Apparently, the tracer pulse leaves behind a long and slowly moving tail as it progresses. Recalling the experiment – impermeable glass beads, narrow grain size distribution, saturated medium – one would not immediately think of immobile regions. However, as illustrated in Figure 7.10, such immobile water may well be expected in the narrow spaces between the grains that are not in the main flow paths. Indeed, places within a distance ℓ of flow paths are accessible through molecular diffusion on a time scale ℓ^2/D, where D is the coefficient of diffusion. The time scale for convective transport is ℓ/v. Equating the two yields $\ell_\mathrm{m} = D/v$ for the characteristic extent of the mobile region around a flow path. While the MIM model as developed above just contains one single time scale for the exchange, ω^{-1}, the real system will exhibit a continuous range of exchange times. This does not affect the qualitative behavior, however.

Generalizations We first notice that the mathematical formulation (7.40) can be easily adapted to more general cases of immobile regions. The most important of these is that of a solute that is adsorbed by the solid matrix. For

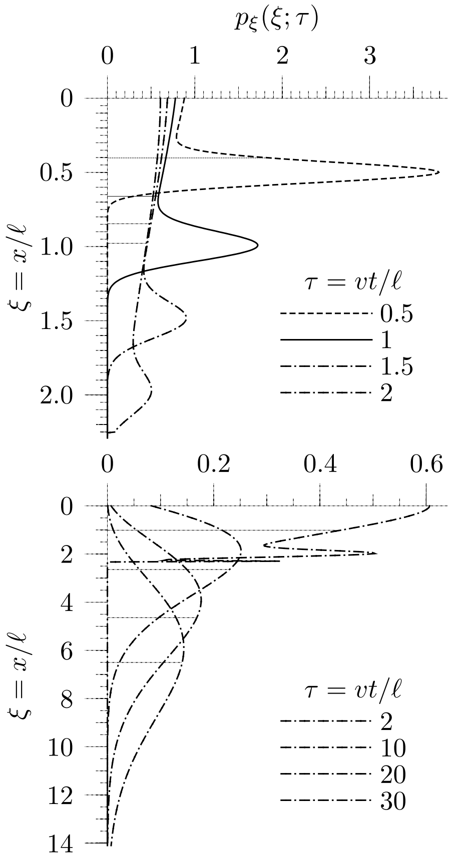


Figure 7.9. Travel distance pdf $p_x(\xi;\tau)$ for MIM model for $\Omega = 0.25$ after short (**top**) and long (**bottom**) travel times. Again, Peclet number and retardation factor are chosen as Pe = 300 and R = 5. The value of Ω is chosen such that $\tau = 1$ is the characteristic residence time in the mobile phase with respect to the exchange. The thin horizontal lines indicate the expectation value $\langle\xi\rangle$ for each of the distributions. The curve for $\tau = 2$ is shown in both graphs for easier comparison. Notice the different scaling of the two graphs.

a large class of solutes, adsorption may be approximated as a linear process such that in equilibrium

$$C_s^{\text{eq}} = K_d C_w \tag{7.59}$$

is the concentration in the adsorbed phase where C_w is the concentration in the water phase and K_d is the distribution coefficient. For practical reasons, C_s is customarily measured as mass of adsorbed chemical per mass of soil matrix. Hence, the total concentration becomes

$$C_t = \theta C_w + \rho_b C_s \tag{7.60}$$

where ρ_b is the bulk density introduced in (3.13). Again assuming a first order kinetics for the exchange between adsorbed (immobile) and water phase (mobile) we obtain for the dynamics, in analogy to (7.40),

$$\begin{aligned} \partial_t C_w + \frac{\rho_b}{\theta}\partial_t C_s + v\partial_x C_w - D\partial_{xx} C_w &= 0 \\ \partial_t C_s &= -\omega[C_s - K_d C_w] \ . \end{aligned} \tag{7.61}$$

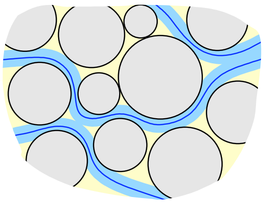


Figure 7.10.
Sketch of immobile water in saturated, uniform, granular media. Water flow will occur along a network of flow paths (dark blue lines). The region within $\ell_\mathrm{m} = D_\mathrm{molec}/v$ of a flow path may be considered as mobile (light blue) whereas those beyond act as immobile (light yellow).

The dimensionless variables (7.41) again lead to (7.42) with the sole difference that the retardation factor is now given by

$$\mathrm{R} = 1 + \frac{\rho_b K_d}{\theta} \; . \tag{7.62}$$

Obviously, the discussions following (7.42) also apply to the new model. This formal analogy between the mobile-immobile two-region model and linear kinetic adsorption was first pointed out by *Nkedi-Kizza et al.* [1984].

For a second generalization we notice that the dimensionless form (7.45) of the convection-dispersion equation in Laplace space may be cast into the form (7.49) by choosing

$$\alpha^{\mathrm{CD}}(s) = s \; . \tag{7.63}$$

We recognize that s, the conjugate of time t, only appears in $\alpha(s)$ and thus speculate that the transport process, in this case convection-dispersion, determines the general form given by (7.49) while the interaction, here MIM, determines the form of $\alpha(s)$. Indeed, *Roth and Jury* [1993] showed that this is correct for a general linear flux law of the form (4.50) and a general linear interaction model as long as the two are independent. Specifically, they showed that in Laplace space the travel time pdf $\widehat{p}_t(s;\xi)$ of a linearly interacting solute may be expressed in terms of the pdf $\widehat{p}_t^{\mathrm{c}}(s;\xi)$ for a solute that is conservative with respect to the interaction as

$$\widehat{p}_t(s;\xi) = \widehat{p}_t^{\mathrm{c}}(\alpha(s);\xi) \; . \tag{7.64}$$

Apparently, the MIM model with an underlying convection-dispersion is a special case. The practical significance of (7.64) is that transport of an interacting solute may be predicated from (i) an experiment with a conservative tracer and (ii) interaction parameters determined in an independent experiment.

Critique The mobile-immobile model yields important conceptual insights into the effect of heterogeneity on solute transport. However, it is not particularly useful for quantitative studies and for real geologic formations. Its fundamental deficiency is that the effective dispersion process is diffusion-like as is testified by $D \propto v^2$ in (7.54). Such a behavior is not corroborated by experimental evidence. Further issues of a more practical nature include that (i) distinguishing only two regions and a single coupling rate is inadequate for many situations, (ii) relating the model parameters – the “immobile fraction” θ_{im} or $\rho_b K_d$ and the effective rate constant ω – to observable properties of the subsurface architecture is difficult and (iii) the richness of natural architectures, including hierarchical heterogeneities, cannot be represented.

7.1.6
Heterogeneous Media

As a preliminary, we recall the separation of the subsurface architecture into *structure* and *texture* as introduced in Section 3.3.2. In the current context, the structure includes those large-scale features of the formation that need to be represented explicitly in order to arrive at a faithful description of solute transport. In contrast, the texture includes all those small-scale features for which a statistical representation suffices, which may thus be incorporated into the hydrodynamic dispersion. Notice that we will often use “microscopic” or “local” to refer to texture and “macroscopic” to refer to structure.

Model Formulation We consider stationary flow in an aquifer whose hydraulic structure is given by the effective conductivity field $K(\mathbf{x})$, which is furthermore assumed to be isotropic. Given appropriate boundary conditions, the flow field $\mathbf{v}(\mathbf{x})$ is obtained by solving (5.22) and invoking Darcy’s law (3.30) in the form $\mathbf{j}_w = -K\nabla h_w$ together with $\mathbf{v} = \mathbf{j}_w/\phi$.

The texture of the aquifer, is irrelevant for the flow field, its contribution is implicitly incorporated in the effective conductivity. Texture is relevant for solute transport, however, since the small-scale velocity variations are the cause of dispersion. In general, it is far from obvious what the optimal parameterization for this dispersion is. For formations with a scale-gap, i.e., with characteristic sizes of textural features being much smaller than the scale at which the structure is resolved, invoking the convection-dispersion model with an effective dispersion tensor is appropriate. On the other hand, if the texture is known to contain units that are qualitatively different from the structure, for instance clay lenses in a sandy aquifer, the MIM model may provide a better parameterization.

Choosing the convection-dispersion parameterization for the local dispersion yields

$$\partial_t[\phi(\mathbf{x})\,C_w] + \nabla\cdot[\phi(\mathbf{x})\,\mathbf{v}(\mathbf{x})\,C_w] - \nabla\cdot[\phi(\mathbf{x})\,\mathsf{D}_{\mathrm{eff}}(\mathbf{x})\,\nabla C_w] = 0 \tag{7.65}$$

which was already derived in (4.53) and is now written such as to emphasize that ϕ, $\mathbf{v}$, and $\mathsf{D}_{\mathrm{eff}}$ are spatially variable. The last step to complete the model is the specification of $\mathsf{D}_{\mathrm{eff}}$. We choose it as a superposition of molecular diffusion, described by the Millington-Quirk parameterization (4.55), with

$\theta = \phi$, and of the hydromechanic dispersion (4.56). In cartesian coordinates, this superposition may be written as

$$D_{ij}(\mathbf{x}) = [\lambda_\ell - \lambda_t]\frac{v_i(\mathbf{x})v_j(\mathbf{x})}{|\mathbf{v}(\mathbf{x})|} + \Big[\lambda_t|\mathbf{v}(\mathbf{x})| + \phi^{1/3}D_m\Big]\delta_{ij} \ , \tag{7.66}$$

where λ_ℓ and λ_t are the longitudinal and the transverse dispersivity, respectively, and D_m is the coefficient of molecular diffusion in pure water.

Two-Dimensional Single-Scale Media In general, (7.65) cannot be solved analytically, even if the parameter fields are constant in time. Hence, we take recourse to numerical simulations and to this end consider the two-dimensional single-scale media already studied in Section 5.4 with flow fields shown in Figure 5.12 on page 116.

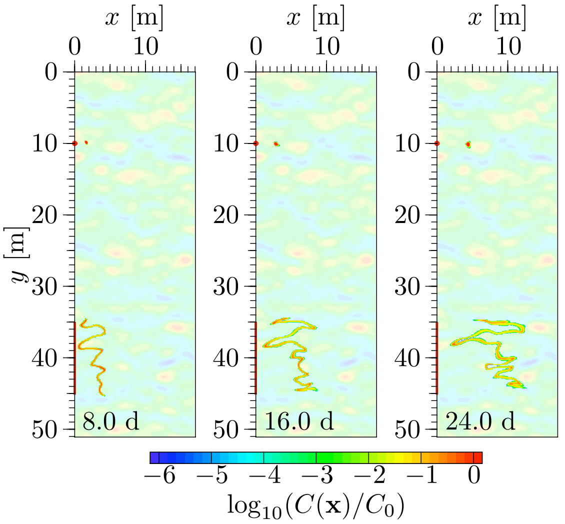
Hydraulic Properties We recall that the hydraulic conductivity of the media used in Section 5.4 is lognormally distributed, i.e., $Y = \log(K/K_0)$ is normally distributed, and that Y is a realization of a weakly stationary random function with isotropic Gaussian autocovariance and correlation length $\ell_Y = 1$ m. The degree of heterogeneity is determined by the variance of Y. Some parameters of the fields are given in Table 5.1 on page 114. The size of the flow domain is chosen as 102.3 m × 51.1 m, some $102\ell_Y \times 51\ell_Y$. The numerical solution is calculated for a spatial resolution of 0.1 m, which leads to a grid with 1024×512 nodes.

Water Flow The water flow is driven by a regional hydraulic gradient of about 0.01 with the left and right boundaries each at a constant hydraulic head. The upper and the lower boundary are impermeable. The velocity field is calculated from $\mathbf{j}_w(\mathbf{x})$ assuming a constant porosity $\phi = 0.25$. Such an assumption is for instance valid for a porous medium that is geometrically similar at each location, at least in a statistical sense, with only the characteristic length varying in space. This is referred to as *Miller-similarity*, due to *Miller and Miller* [1956] who first introduced the concept for flow through porous media. Examples of formations for which Miller-similarity is a reasonable approximation include sandy alluvial deposits.

Solute Transport For solute transport, $D_m = 2 \cdot 10^{-9}$ m^{2}s^{-1} is chosen as the coefficient of molecular diffusion. With the Millington-Quirk factor $\phi^{1/3}$, this leads to $D_m^{\text{eff}} = 1.26 \cdot 10^{-9}$ m^{2}s^{-1}. For hydromechanic dispersion, three cases are studied, namely $\lambda_\ell = 0$, 0.005 m, and 0.025 m and each with $\lambda_t = \lambda_\ell/10$. Notice that λ_ℓ is chosen to be substantially smaller than the spatial resolution of the simulation, 0.1 m, since it represents sub-scale dispersion that originates in the texture.

Numerical Simulation The transport problem is solved by tracking a large number of particles, $1 \dots 32 \cdot 10^6$, using the algorithm of *Roth and Hammel* [1996]. Boundary conditions are such that particles are reflected at all boundaries, except for the righthand boundary where they leave the flow domain.

Figure 7.11. Normalized concentrations for an initial point source (red dot) and for an initial line source (red line) perpendicular to the mean flow. The variance of $\log_{10}(K/K_0)$ is 0.15 and travel times are 8, 16, and 24 days. Dispersion is by molecular diffusion only. The background represents the flow field, more precisely $\log(|\mathbf{j}_w(\mathbf{x})|)$, with red and blue indicating low and high values, respectively.



Weakly Heterogeneous Medium We first concentrate on the medium with $\mathrm{var}(\log_{10}(K/K_0)) = 0.15$ and study the transition between different dispersion regimes for two initial distributions at the lefthand boundary – a point source at $y = 10$ m and a line source between $y = 35$ m and 45 m – and for different microscopic dispersion processes.

Molecular Diffusion For a medium that is very uniform microscopically, molecular diffusion is the only relevant microscopic dispersion process. Exercise 7.6 shows that for the flow fields considered here, this is a good assumption if the medium is uniform at a scale of about 5 mm. Of course it is heterogeneous at smaller scales, where the pore space would come into view, as well as at larger scales where the hydraulic structure is represented explicitly. Normalized concentration distributions for short travel times are shown in Figure 7.11.

In a uniform flow field, molecular diffusion leads to an isotropic Gaussian transition pdf whose radius – standard deviation – for time t is $[2D_m^{\mathrm{eff}} t]^{1/2}$. Inserting $D_m^{\mathrm{eff}} = 1.26 \cdot 10^{-9}$ m^{2}s^{-1} yields a radius of 0.07 m after 24 days. Hence, the distribution covers some 14% of ℓ_Y. Since the conductivity field is very smooth, compare Figure 5.11 on page 114, we expect that the size of the pulse is approximately the same in the uniform and in the heterogeneous medium. Indeed, the distribution originating from the point source remains highly localized and its spreading is hardly visible at the given resolution of 0.1 m. Hence, we may envisage the particles to remain on their initial streamline for some time. This regime is very much akin to the near-field of Taylor-Aris dispersion discussed in Section 4.1.2. The major differences are that (i) streamlines are contorted and (ii) the velocity on any particular streamline is not constant. As a consequence, particle distributions cannot be calculated from the initial velocity distribution anymore and the

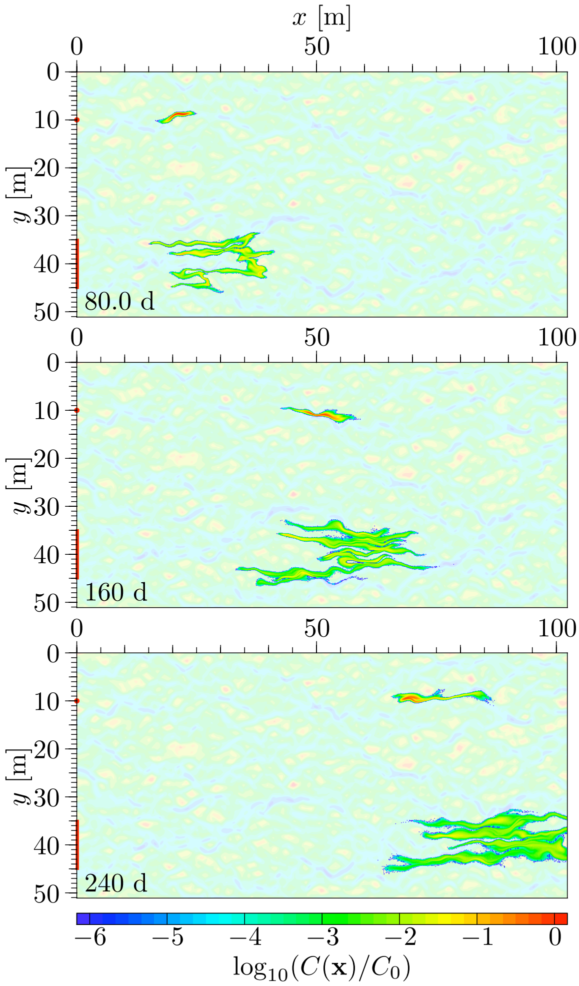


Figure 7.12. Evolution of concentration distributions from Figure 7.11 for 80, 160, and 240 days.

simple stochastic-convective scaling (7.13) is no longer valid. However, for sufficiently short times – certainly for travel distances smaller than ℓ_Y – we may expect simple stochastic-convection to yield a reasonable approximation. This is illustrated qualitatively by the concentration distributions produced by the line source. For very much longer times and stationary hydraulic properties, in the statistical sense, we may hypothesize that a single particle moving along a streamline will eventually sample the velocity field in a representative manner. Indeed, this hypothesis is the basis of the *stochastic continuum theory*, a popular perturbation approach for weakly heterogeneous media that is based on early work by *Phythian* [1975] and was further developed by *Gelhar and Axness* [1983] and *Dagan* [1984].

The evolution of concentration distributions for longer times is illustrated by Figure 7.12. We first notice that the distribution from the point source develops a highly anisotropic shape. To understand this, we again calculate the radii of the isotropic distributions that would result from molecular diffusion alone. This yields 0.13 m for $t = 80$ d and 0.22 m for $t = 240$ d, hence a coverage of ℓ_Y by 26 and 44%, respectively. While the width of the pulse is of the expected size, its length is almost two orders of magnitudes larger. Again, the situation is very similar to Taylor-Aris dispersion: The comparatively small transverse spreading leads to transitions between streamlines at somewhat greater distances. This in turn changes the velocity of a particle which spreads the distribution along the direction of the local mean flow much faster than microscopic dispersion. A further increase in the rate of spreading may be expected when the flow channel that contains the pulse forks asymmetrically into two channels. This apparently has not yet happened with the pulse studied here but it certainly will as it widens enough to cover an entire channels.

The same evolution as for the point source is also manifest in the distribution from the line source. Its longitudinal growth is not as striking, though, since different parts of the line have been on different streamlines right from the start. Hence, they already encountered the full range of velocities. However, we notice that the distribution evolves into a set of somewhat isolated thin filaments that are aligned with the mean flow. An intuitive explanation for this is that diffusion brings the particles into high- and low-velocity regions with equal probability. However, longitudinal spreading is more effective in the high-velocity regions simply because velocities and thus also their differentials are higher.

Finally, we emphasis the fundamental importance of molecular diffusion. While it is a slow mixing process for larger distances, it effectuates transitions between streamlines. This in turn is the key for the more effective mixing processes, in particular for Taylor-Aris-like dispersion, within flow channels and for the hydromechanic dispersion that results from the macroscopic network of channels. Without molecular diffusion, an initial point distribution would obviously never spread.

Hydrodynamic Dispersion Next, we study the influence of the magnitude of local dispersion and consider hydromechanic dispersion in addition to molecular diffusion (Figure 7.13). This corresponds to assuming a larger characteristic size for the features that make up the texture of the medium. As a preliminary, we estimate the mean dispersion tensor that accounts for the combined processes. Using (7.66), it may be approximated as

$$D_{\ell,t}^{\text{eff}} = \phi^{1/3} D_m + \overline{v}\lambda_{\ell,t} \ , \tag{7.67}$$

where the subscripts ℓ and t indicate longitudinal and transverse direction, respectively. With the mean velocity $\overline{v} = 4 \cdot 10^{-6}\ \mathrm{m\,s^{-1}}$ – estimated from the displacement of the original line source shown in Figure 7.12 – we obtain for the effective dispersion coefficients $\{D_\ell^{\text{eff}}, D_t^{\text{eff}}\}$ approximately $\{21, 3.2\} \cdot 10^{-9}\ \mathrm{m^2\,s^{-1}}$ for $\lambda_\ell = 0.005$ m and $\{101, 11\} \cdot 10^{-9}\ \mathrm{m^2\,s^{-1}}$ for $\lambda_\ell = 0.025$ m.

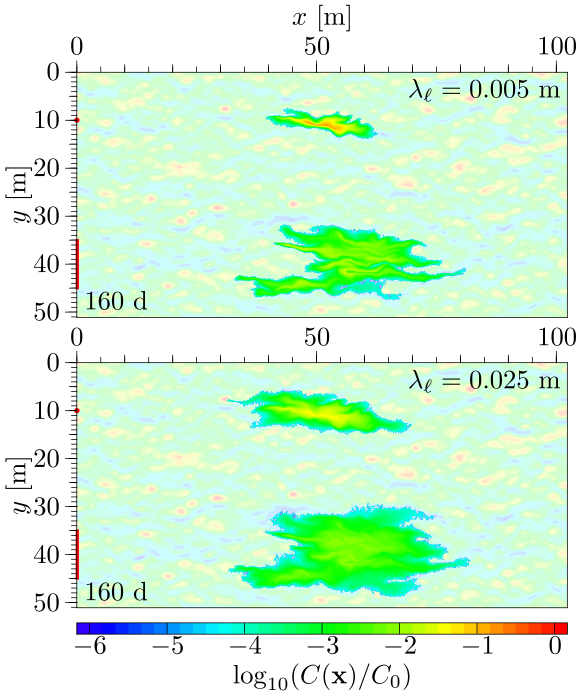


Figure 7.13. Normalized concentration distribution for transport in same velocity field as used for Figure 7.12. The microscopic dispersion process is represented by molecular diffusion and, additionally, by hydromechanic dispersion with $\lambda_\ell = 0.005$ m (upper) and 0.025 m (lower), respectively.

In a corresponding uniform medium, an initial point-like distribution would thus evolve into an anisotropic Gaussian distribution. After 160 d, its major radii would be $\{0.77, 0.30\}$ m and $\{1.67, 0.56\}$ m, respectively. Clearly, the distributions in Figure 7.13 are much broader than this due to the macroscopic dispersion process discussed above.

What is the impact of the significantly increased microscopic dispersion? We first recall that the onset of the macroscopic dispersion is determined by the time scale of transverse mixing. For a mixing distance ℓ this is given by $\ell^2/[2D_t^{\mathrm{eff}}]$. Hence, increased spreading is expected for times that are shorter than with pure molecular diffusion by the factor D_t^{eff}/D_m. For $\lambda_\ell =$ 0.005 m, this factor is 2.6, for 0.025 m it is 8.9. We thus anticipate that the duration of the initial phase of the dispersion process decreases with increasing microscopic heterogeneity. On the other hand, dispersion at the macroscopic scale is dominated by the spatial structure of the velocity field and no more by local mixing. The later phase of dispersion is thus expected to depend only weakly on the value of D_t^{eff} with the main impact being a stronger smoothing with increasing local dispersion. This is corroborated qualitatively by Figure 7.13 and it will be further supported below.

Transition Probability and Ergodicity Recall that the normalized concentration distribution that results from an initial uniform concentration distribution in some arbitrary region Ω may be regarded as the probability density for the transition from Ω to some location $\mathbf{x}$ by time t. Hence, $p_{\mathbf{x}}(\mathbf{x}; t|\Omega; 0) = C(\mathbf{x}; t)/C_0$, where $C(\mathbf{x}; 0) = C_0$ for $\mathbf{x} \in \Omega$ and zero otherwise.

Consider the distribution originating from the point source at $\mathbf{x}_0$. It corresponds to the travel distance pdf $p_{\mathbf{x}}(\mathbf{x}; t|\mathbf{x}_0; 0)$. Recall that this pdf can be measured with an appropriate tracer experiment. The important question arising here is: Once $p_{\mathbf{x}}(\mathbf{x}; t|\mathbf{x}_0; 0)$ is determined for a specific location $\mathbf{x}_0$, what do we learn about transport in the entire flow domain? The answer to this has several facets:

1. For a stationary system – constant flow and time-invariant transport properties – the pdf $p_{\mathbf{x}}(\mathbf{x}; t|\mathbf{x}_0; 0)$ provides a complete description of the influence of the concentration $C(\mathbf{x}_0; t)$ on the entire flow domain. Indeed, following Section 7.1.4, the concentration at an arbitrary location $\mathbf{x}$ may be written as $C(\mathbf{x}; t) = \int_{-\infty}^{t} C(\mathbf{x}_0; \tau) p_{\mathbf{x}}(\mathbf{x}; t - \tau|\mathbf{x}_0; 0)\, \mathrm{d}\tau$.

2. In a heterogeneous medium, $p_{\mathbf{x}}(\mathbf{x}; t|\mathbf{x}_0; 0)$ will *not* describe the transport from any location other than $\mathbf{x}_0$. Such a process would constitute a different system with a different transition pdf. This is obvious from the evolution of the line source in Figure 7.11, which may be regarded as a superposition of pulses.

3. The situation is simpler in the far-field, after the initial distribution has spread over a region that is very much larger than the correlation length ℓ_Y of the hydraulic structure. The concentration distributions are decomposed into several filaments as is apparent from Figure 7.12 and Figure 7.13. However, these filaments reflect the local flow channels and are as such independent of details of the initial distribution $C(\mathbf{x}; 0)$. Hence, shifting $C(\mathbf{x}; 0)$ a bit would not affect the location and extent of the filaments.

While the prediction of the concentration at any one location is impossible without a detailed simulation, an average statistical description thus appears feasible. To this end it is useful to introduce the notion of a *macroscopically uniform* field for which the statistical moments, calculated over a sufficiently large region, are translation-invariant. This concept is in complete formal analogy to that of the representative elementary volume (REV) introduced in Section 3.3.1. Indeed, we may refer to the minimal region that makes a field macroscopically uniform as its REV.

Returning to the predictability of transport in a heterogeneous medium, we consider a macroscopically uniform medium with stationary flow in a conductivity field with finite variance and correlation length. Invoking the central limit theorem we may conclude that eventually, in the so-called *asymptotic limit*, the average plume is transported and dispersed according to the convection-dispersion model with some effective value for the velocity and for the dispersion tensor. Calculating such effective parameters for an ensemble of hydraulic structures, and indeed also the transition from the near- to the far-field, is the aim of stochastic continuum theory [e.g., *Dagan* 1984, 1989]. It is based on the ergodicity hypothesis that is used to construct an ensemble whose statistics – mean, distribution, and covariance function of $Y = \log(K/K_0)$ – equals the spatial statistics of the given realization. Based on the assumption that $\mathrm{var}(Y)$ is small, a perturbation solution is obtained for the moments of the transition pdf. We again emphasize, however, that such a theory only yields results for the ensemble and will not be able to predict local concentrations and fluxes. Still, it has been found to successfully describe the transport in weakly heterogeneous media as has for instance been demonstrated by *Dentz et al.* [2000] for the Borden aquifer.

Apparent Transport Parameters Short of actually entering the stochastic continuum theory, we take a heuristic look at the evolution of the spatial moments of the particle distribution as it is provided by the numerical simulations above and represented in Figures 7.11–7.13. To this end, we employ the *method of moments* and the global apparent parameters defined by (4.27). The moments of interest here are the x-component of the center of gravity and the variances in x- and y-direction. Mean and variance in x-direction are obtained as

$$\langle x \rangle = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{and} \qquad \mathrm{var}(x) = \frac{1}{n-1}\sum_{i=1}^{n}[x_i - \langle x \rangle]^2 \, , \tag{7.68}$$

respectively, where x_i is the x-component of particle i's position. The variance in y-direction is calculated in analogy.

Before proceeding, we recall that the moments (7.68) can be calculated for any particle distribution and likewise numbers for v and D can be obtained from (4.26) or from (4.27). All this is independent of the underlying transport process. However, the interpretation of the parameters v and D does depend on the process. For instance, calculating the spreading of a distribution with $\sqrt{2Dt}$ requires that the underlying process is a convection-dispersion. If this can be ascertained *a priori*, then the parameters are called "effective". If this is not the case, as for instance for the macroscopic transport parameters of a heterogeneous medium, they get the label "apparent". This label may be changed *a posteriori*, once it has been ascertained that the parameters conform to the convection-dispersion model, i.e., D is found to be constant.

For a macroscopically uniform medium, we anticipate that a stationary flow field will also be macroscopically uniform, although, recalling Figure 5.12 on page 116, probably with a much larger REV than the underlying conductivity field. Then, after a sufficient transport distance, the method of moments would again yield effective parameters. The crux here is the "sufficient distance", which may be impractically large. We will find in the following that $10^2\ell_Y$ is an optimistic lower limit even for the rather well-behaved conductivity fields considered here. This is readily understood in the framework of the MIM model discussed in Section 7.1.5, which shows that

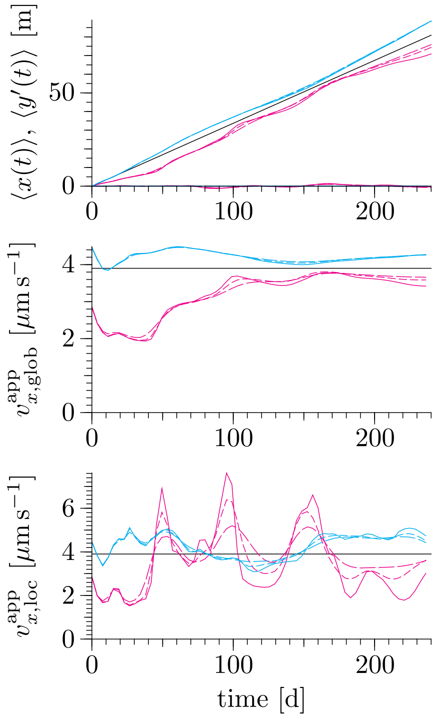


Figure 7.14.
Movement of particle distribution's center of gravity, $\langle x(t)\rangle$ and $\langle y'(t)\rangle = \langle y(t) - y(0)\rangle$ (**top**), global apparent velocity $v^{\mathrm{app}}_{x,\mathrm{glob}} = \langle x(t)\rangle/t$ (**middle**), and local apparent velocity $v^{\mathrm{app}}_{x,\mathrm{loc}} = \partial_t\langle x(t)\rangle$ (**bottom**) for the point source (**magenta**) and for the line source (**cyan**) shown in Figures 7.11–7.13. Movement in y-direction is minute and hardly visible in the graph. The black line represents the mean x-component of the entire velocity field. The three curves for each source represent the three cases of microscopic dispersion considered above, namely, molecular diffusion alone (solid) and hydrodynamic dispersion with $\lambda_\ell = 0.005$ m (short dashes) and $\lambda_\ell = 0.025$ m (long dashes).

macroscopic dispersion in a heterogeneous medium is dominated by (i) the hydraulic contrast between different regions, in (7.54) expressed by the retardation parameter R, and (ii) the mixing time, there the rate parameter ω. A final, more practical issue is the characterization of the flow field, which is traditionally done through the correlation length ℓ_y of the underlying log-conductivity field. This is motivated by the hope that ℓ_y may be measurable for natural formations. Whether or not this will become feasible eventually, we notice that there is a large set of natural formations whose hydraulic characteristics are not represented well by just the correlation length. This includes undulating formations as they typically occur in aquifers that originate in ancient braided rivers.

After these preliminaries, we look into the movement of the center of gravity and the deduced apparent velocities (Figure 7.14). First, we notice that the center of gravity moves along a fairly straight line with minimal deflection in y-direction. Fluctuations in x-direction, along the mean flow, are larger and indicate that the distributions traverse high- and low-flow regions. As a consequence, the velocities in x-direction show considerable variation, by about ±5% for the global velocity from the line source and about a factor of two larger for the one from the point source. The cause for this difference in the magnitude of the fluctuations is the sampling region of the two distributions. The line source extends over 10 m perpendicular to the mean flow. Hence, the particles originating there are distributed over $10\ell_Y$ and thus sample the field of x-velocities representatively. The remaining fluctuations result from large-scale structures of the flow field and from the still imperfect sampling. In contrast, particles from the point source initially sample only a single point where the velocity may be much higher or much lower than the mean. For the case we happen to look at here, the pulse starts in a low-flow region. Its mean global velocity by $t \approx 45$ d is about $1.9 \cdot 10^{-6}$ $\mathrm{m\,s^{-1}}$, half of the mean velocity of the entire field. Looking at Figure 7.11, we find that this low-flow region extends to $x \approx 7$ m where a high-flow region follows. Calculating the travel time to the transition with the previously estimated initial mean velocity yields about 43 d, which fits nicely with the observation in Figure 7.14.

We notice the difference in the level of fluctuations between global and local estimates of the apparent velocity, $v_{x,\mathrm{glob}}^{\mathrm{app}}$ and $v_{x,\mathrm{loc}}^{\mathrm{app}}$, respectively. This is readily understood when we realize that the two represent completely different averages. The global estimate reflects the mean velocity along the entire path of all the particles. In contrast, the local estimate results from the mean velocity at the current locations of the particles. Hence, the two sample completely different regions. For a macroscopically uniform flow field, the two velocity estimates will asymptotically approach each other as well as the mean velocity $\langle v \rangle$ of the entire field, which is yet another estimate. This is at least the case when molecular diffusion is the only microscopic dispersion process. Indeed, diffusion is independent of the velocity field and will spread the particle distribution indiscriminately over an ever increasing region. Eventually, it will sample a region that is large enough to represent the macroscopically uniform flow field.

Next, we consider the evolution of the macroscopic dispersion and describe it, again based on (4.27), with the apparent dispersivities

$$\lambda_\ell := \frac{D_\ell}{v_x} = \frac{\mathrm{var}\big(x(t)\big)}{2\big\langle x(t)\big\rangle} \quad \text{and} \quad \lambda_t := \frac{D_t}{v_x} = \frac{\mathrm{var}\big(y(t)\big)}{2\big\langle x(t)\big\rangle} \ . \tag{7.69}$$

Here we used that the mean flow is in x-direction and the sub- and superscripts are dropped for a leaner notation (Figure 7.15). Furthermore, the global estimate $D = \frac{1}{2t}\mathrm{var}(x(t))$ was used instead of the local estimate $D = \frac{1}{2}\partial_t \mathrm{var}(x(t))$. The former yields more stable results, although at the price of a slower convergence. This was already observed for the case Taylor dispersion (Figure 4.4 on page 86). Finally, notice from the comparison with (7.67) that we fold the contribution of molecular diffusion into the apparent dispersivities. This may be justified because in practical situations, the two processes cannot be separated. In addition, D_m/v is typically some orders of magnitudes smaller than λ^{app}.

First consider the longitudinal dispersivity calculated from the line source. It increases approximately linearly and reaches a maximum after some 50 d, corresponding to a travel distance of some 20 m. This is the near-field

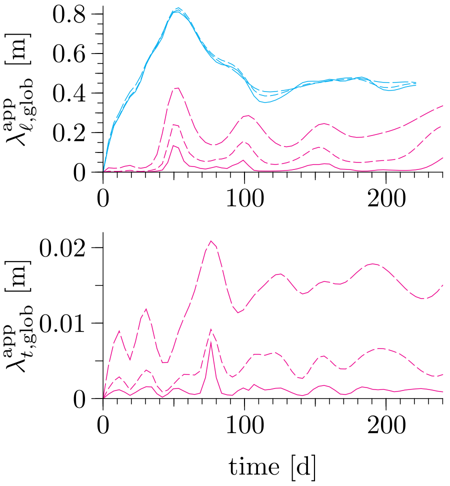


Figure 7.15. Apparent global dispersivities for the longitudinal (**top**) and the transverse (**bottom**) direction obtained with (7.68) for the simulations shown in Figures 7.12–7.13 for $\mathrm{var}(\log_{10}(K)) = 0.15$. Colors and dashes are the same as in Figure 7.14. The contribution from molecular diffusion is minute, $\phi^{1/3} D_m/\bar{v} = \mathcal{O}(10^{-4}\ \mathrm{m})$, and cannot be represented on the scale used here.

(stochastic convection) regime. Indeed, inserting (4.28) into (7.69) leads to $\lambda_\ell = \alpha\langle v\rangle t$, where $\alpha = \mathrm{var}(v/\langle v\rangle)$ is a characteristic of the flow field. In Taylor-Aris dispersion, the SC-regime ends when particles have diffused a sufficient distance from their original streamline and thus encounter significantly different velocities. With the type of smooth velocity field used there, D^{app} approaches its asymptotic value monotonically, as is illustrated in Figure 4.4 on page 86. In a heterogeneous medium, the changing velocity along the streamlines provides an additional and independent path out of the SC-regime. It is this path that also leads to the oscillations of $\lambda^{\mathrm{app}}_\ell$. To understand this, we first realize that the value of $\mathrm{var}(x)$ is dominated by the distance between the front and the rear end of the distribution. These in turn consist of particles that stayed in the fastest and slowest flow channel, respectively. As long as they stay there, the pulse will continue to spread. However, as it progresses, the probability for these channels to end increases. Since the fastest channels tend to end in slow regions and vice versa, the spreading of the pulse will end rather abruptly and it will actually get compressed. This explains the maximum of $\lambda_\ell(t)$.

The unfolding of dispersion becomes more apparent when we consider the evolution of the point source, which we already discussed qualitatively in the context of Figure 7.11. We may distinguish three regimes:

1. For short times, $t < 30$ d, dispersion is dominated by the microscopic processes which here correspond to a convection-dispersion. This is indicated by the constant value of λ_ℓ. Inspection of the data reveals that the values obtained – $7.7 \cdot 10^{-4}$ m, 0.0055 m, and 0.024 m – indeed conform to the expectations. Also the values of the transverse dispersivity are approximately constant and of the correct order, although they fluctuate quite a bit more.

2. For longer times, the pulse starts to oscillate both in the longitudinal and in the transverse direction. The two are actually anti-correlated, i.e., when the pulse expands laterally, it tends to shrink transversally, and vice versa. This fits nicely with expectations for a movement through narrow high-flow and wider low-flow regions. We also notice that for the case with molecular diffusion only, the pulse is compressed almost back to its original tiny size after each expansion. For the cases with additional hydromechanic dispersion, this is not the case and the minimum spreading grows continuously with time while the amplitude of the oscillation decreases. This is readily understood since the microscopic dispersion affects the transition between streamlines which in turn cause the irreversible macroscopic mixing.

3. For very long times, much longer than the 240 d simulated here, we may anticipate that λ_ℓ approaches the same asymptotic value as does the distribution from the line source. This follows from the fact that eventually, both distributions will sample an REV of the velocity field. Indeed, such an approach is hinted at for the case with the largest value of the microscopic transverse hydromechanic dispersivity of 0.0025 m. In a uniform medium, the pulse would have a transverse radius of 0.68 m by $t = 240$ d, which is of the order of ℓ_Y. In contrast, the apparent dispersivity of the pulse with pure molecular diffusion is still far from the asymptotic value and it will take a long time to get there. To also grow to a width of 0.68 m, some 8.6 y would be required during which time the pulse would move some 1.1 km.

Moderately Heterogeneous Medium We consider a conductivity field that has the identical form as the one used before but now with a larger magnitude of the heterogeneity, namely $\mathrm{var}(\log_{10}(K/K_0)) = 0.75$, corresponding to $\sigma_Y^2 = 4.0$. We again study the particle distributions that result from an initial point and line source, respectively (Figure 7.16).

Apparently, dispersion is very much stronger even though the correlation length of the conductivity field is the same. The reason for this is that the velocity gradients are much larger such that particles on different streamlines drift away from each other at a higher rate. Also notice the thin filaments which testify to the small spreading from molecular diffusion and to the persistence of narrow flow channels.

In analogy to the case of the weakly heterogeneous medium, we summarize the spreading of the particle distributions by the apparent global velocities and dispersivities (Figure 7.17). We notice that, because of the stronger dispersion, a faithful estimate of the statistical moments requires a much larger domain than in the previous case.

As in the case of weak heterogeneity, the apparent velocity appears to approach the mean velocity of the entire field. Deviations from this value are larger, however, reflecting the stronger variation of the velocity field. As

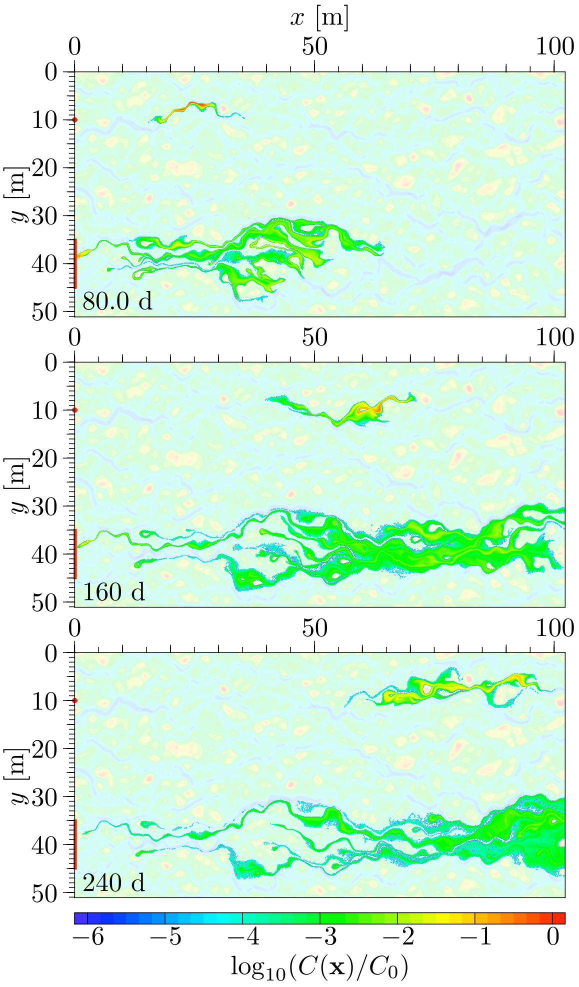


Figure 7.16. Evolution of tracer distributions emanating from a point and a line source in a medium whose conductivity field is identical to the one used for Figure 7.12 except for the larger magnitude of the heterogeneity, for which $\mathrm{var}(\log_{10}(K/K_0)) = 0.75$, corresponding to $\sigma_Y^2 = 4.0$. The regional hydraulic gradient is again 1/102.3 and microscopic dispersion is by molecular diffusion only.

expected, the differences between the different microscopic dispersion regimes are rather moderate for the line source with is broadly sampling of the velocity field right from the start. They become considerably larger for the point source once the spreading of the distribution becomes significant and the asymptotic regime has not yet been reached.

The distribution from the initial line again exhibits the initial stochastic convective phase during which $\lambda_\ell^{\mathrm{app}}$ increases roughly linearly. It ends after some 40 d, comparable to the case of the weakly heterogeneous medium. However, the dispersivity increases faster, because of the larger velocity gradients, such that the maximum value attained is larger by a factor of about 5. A further difference is that the overshoot is smaller, almost negligible. This may be explained by the larger longitudinal extent which has the particles sample the velocity field more representatively upon lateral diffusion.

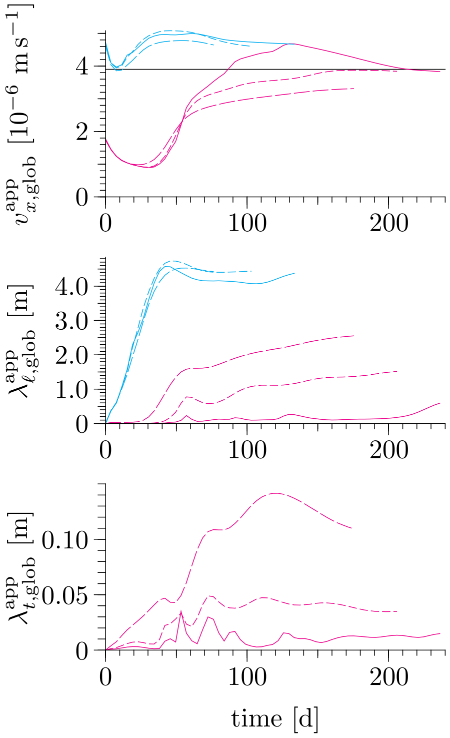


Figure 7.17. Apparent global velocities (**top**) and dispersivities for the longitudinal (**middle**) and the transverse (**bottom**) direction for the simulation shown in Figure 7.16 and for the cases with microscopic hydrodynamic dispersion with $\lambda_\ell = 0.005$ m and 0.025 m, respectively. Colors and dashes are the same as in Figures 7.14–7.15. The curves are only shown for times for which no significant fraction of the particles has reached the outlet end. The black line in the top graph represents the mean x-component of the entire velocity field.

The distribution from the initial point shows a practically negligible longitudinal dispersion for some time, until the distribution has spread sufficiently in the transverse direction. After that, spreading increases rapidly as hydromechanic dispersion sets in. In contrast to the weakly heterogeneous case, the fluctuation of the apparent dispersivity is very small, which is again explained by the rapid longitudinal spreading and the correspondingly better sampling of the velocity field.

We finally comment that the statistical moments only provided a very rough description of the actual particle distribution. For instance, we obtain $[2\lambda_\ell^{\mathrm{app}}\overline{v}t]^{1/2}$ for the radius in longitudinal direction. With $\overline{v} = 4\ \mu\mathrm{m\,s^{-1}}$ and $t = 160$ d this yields some 21 m which is a very rough representation indeed. It obviously indicates neither the very long filaments behind and in front of the main distribution nor the regions with practically negligible concentration right in the middle of the plume. Depending on the particular question at hand – think of first arrival from a nuclear depository or, in contrast, of the final cleanup of some noxious chemical – our interest may be in exactly such filaments, in their extent and specific localization. An example of great current interest is the arsenic contamination of large portions of the groundwater in Bangladesh where it is found that wells that yield clean water may be located a very short distance, tens of meters, from wells producing contaminated water [*van Geen et al.* 2003].

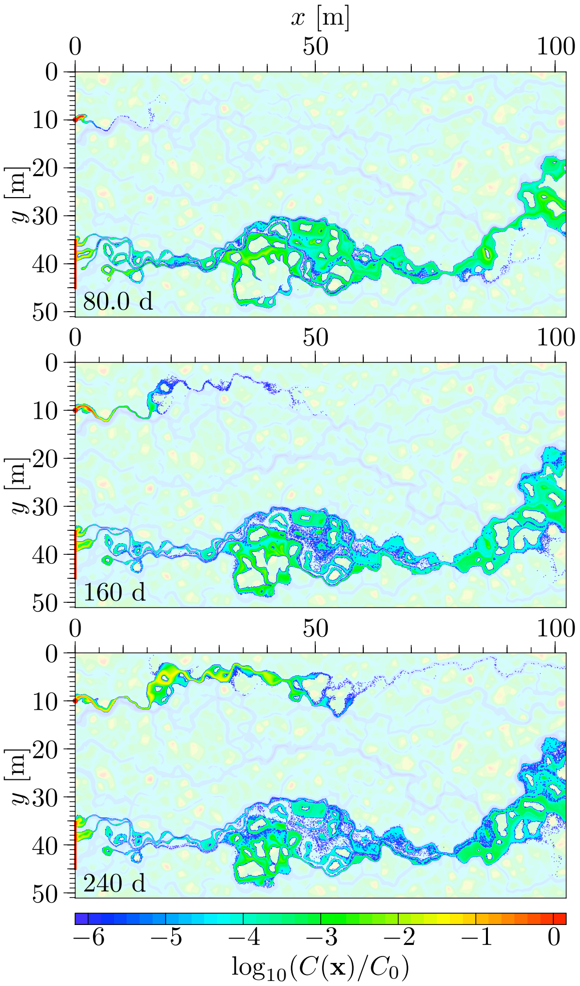


Figure 7.18. Tracer distributions evolving from point and line source in a medium whose conductivity field is identical to the one used for Figure 7.12 except for the magnitude of the heterogeneity which is $\mathrm{var}(\log_{10}(K/K_0)) = 3.77$, corresponding to $\sigma_Y^2 = 20$. The regional hydraulic gradient is again $1/102.3$ and microscopic dispersion is by molecular diffusion only.

From a more theoretical perspective we notice that the asymptotic limit of the convection-dispersion regime is reached on increasingly larger spatial scales as the magnitude of the heterogeneity increases. Hence, the REV for an effective, macroscopically uniform medium also increases with the variance σ_Y^2 and not only with the correlation length ℓ_Y.

Strongly Heterogeneous Medium We continue to consider the same conductivity field as before but with the increased variance $\mathrm{var}(\log_{10}(K/K_0)) = 3.77$ which corresponds to $\sigma_Y^2 = 20$ (Figure 7.18). Obviously, the phenomenology of transport on the scale considered here has changed dramatically when compared to the case of the weakly heterogeneous medium. The disparity of the velocities between high- and low-flow regions is so large that the latter primarily appear as stationary particle sources – if the initial distribution puts particles in such regions – while the former become high-

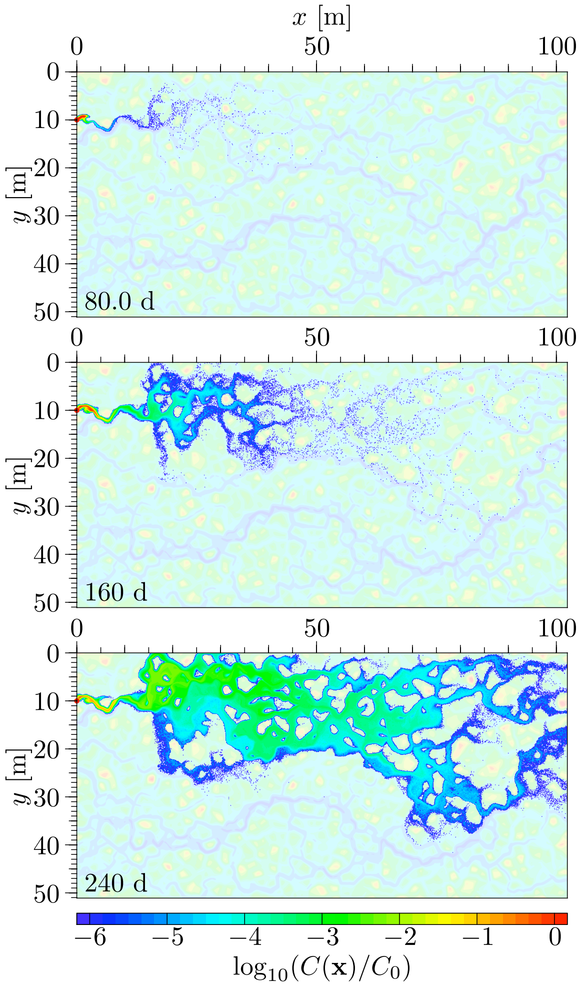


Figure 7.19. Evolution of point source from Figure 7.18 but with microscopic hydromechanic dispersion with $\lambda_\ell = 0.005$ m in addition to molecular diffusion.

speed conduits. Particles that slowly diffuse from low- to high-flow regions are entrained and form very thin and long filaments that feed from almost continuous sources.

Next, we consider the impact of microscopic dispersion for the example of the point source (Figure 7.19). While the initial evolution is very similar to the case of pure molecular diffusion the distribution virtually explodes once transverse spreading is large enough to reach neighboring flow channels. Again notice that the concentrations in the leading part of the pulse are very small because of (i) the wide spreading and (ii) the slow feed into the high-flow channels. Indeed, a single speck in Figure 7.19 corresponds to a dilution of the original concentration by a factor of some 10^7. Such a dilution renders most contamination problems irrelevant since concentrations fall below the toxicity threshold and, at least for many organic contaminants, microbes become efficient decomposers, leading to the so called *natural attenuation*. This may not be true for highly toxic and inorganic solutes, however. A notorious example is plutonium that originates from nuclear tests and from waste disposal sites [*Kersting et al.* 1999].

Finally, we turn to the statistical moments and apparent parameters derived from there. In contrast to the previous cases, no asymptotic limit is reached within the simulated domain which extents for some $10^2\ell_Y$. While this could be remedied by considering a much larger domain and we know from the CLT that such a limit exists, there is little motivation for such an approach. The main reason for this is that the resulting moments and effective parameters would be of very limited interest since the deviations between realizations are so large that the ensemble average, which is described by the moments, is no more representative for any individual realization.

7.1.7
Aquifers

Natural formations like aquifers and soils typically exhibit hierarchical architectures (Section 3.1) which, we expect, give rise to hydraulic structures whose size ℓ_Y increases with the size ℓ of the flow domain considered. This in turn would lead to an increase of apparent dispersivities λ_ℓ with scale ℓ. Indeed, *Gelhar et al.* [1992] found a significant relation between the reported values of λ_ℓ and ℓ for a number of tracer experiments, roughly $\lambda_\ell = \ell/10$ (Figure 7.20). While the scatter of the data is large, some two orders of magnitude, there is nevertheless a reasonable correlation. Interpretation of such an assembly of data just in terms of the flow domain's size ℓ is still a bit daring, however, since the experiments were performed in quite different geological environments. Their hydraulic properties, in particular σ_Y^2, probably deviated considerably. However, the scatter of the data is so large that it can easily accommodate the corresponding variations. Specifically, σ_Y^2, the variance of the hydraulic conductivity's *logarithm*, can be expected to not vary much more than the two orders of magnitude of the data. As a reference, the range covered by the simulations in Section 7.1.6 represents a factor of 25.

A further issue with the data shown in Figure 7.20 concerns the premisses for their derivation. Most of them assumed a macroscopically uniform medium, hence a uniform velocity field with all the observed spreading assigned to dispersion. This is one extreme end for a design decision in analyzing data from hierarchically structured formations. The other end would be the complete representation of the hydraulic structure with dispersion through molecular diffusion only. Clearly, the resulting values for the apparent dispersion coefficient varies greatly with such a choice, roughly between 10^{-9} m^{2}s^{-1} for a full representation of the structure and $\ell v/10$, where v is the mean velocity, for a uniform medium. Indeed, *Gelhar et al.* [1992] reanalyzed some of the data by accounting for some of the heterogeneity and found that

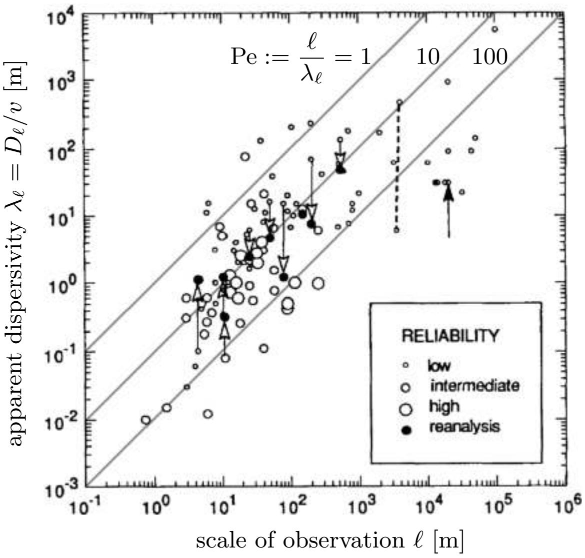


Figure 7.20. Apparent longitudinal dispersivity estimated from field experiments reported in the literature. The size of a symbol indicates the reliability of the data. It depends on experimental setup and resolution. Arrows link reported values with estimates after reanalysis. The dashed line connects two values that were estimated for two measuring locations in the same experiment. Gray lines correspond to constant values of the macroscopic Peclet numbers defined by (7.70). (Adapted from Fig. 3 of *Gelhar et al.* [1992] who first introduced this representation.)

this reduced the estimates of the effective dispersivity significantly (indicated by the downwards-pointing arrows in Figure 7.20). On the other hand, there exist circumstances where the effective dispersivity is underestimated, for instance when highly localized measurements in a flow channel are used. A few such cases have also been reanalyzed by *Gelhar et al.* [1992] and are indicated by upwards-pointing arrows. Apparently, there is a considerable methodological latitude in estimating the dispersivity for large heterogeneous formations.

The matter is further complicated by two additional issues. The first is that the asymptotic limit has not yet been reached in most of the experiments, hence small-scale details still persist in the tracer distribution. This is indicated by the rather small values of the macroscopic Peclet number, which we define in analogy to (7.25) as

$$\mathrm{Pe} := \frac{\ell v}{D_\ell^{\mathrm{eff}}} = \frac{\ell}{\lambda_\ell^{\mathrm{eff}}} \; . \tag{7.70}$$

The second issue is that measurements are obtained from typically point-like instruments, observation wells or drill cores. These do not represent the presumed uniform medium but relate to a much higher spatial resolution. This is indicated by the dashed line in Figure 7.20 which connects two measurements obtained for two observation wells with roughly the same travel distance (some 4 km) at the same site (Hanford).

A faithful representation of solute transport in natural formation still poses formidable technological challenges in that it requires an accurate description of the underlying hydraulic structure with a reasonable spatial resolution, say $\ell/10\ldots\ell/100$ where ℓ is the typical transport distance. No less important is the observation of concentrations at a scale that is commensurate with that resolution.

Exercises

7.1† Diffusion Consider quantity $f(x,t)$ whose evolution is governed by the diffusion equation $\partial_t f - D\partial_{xx} f = 0$. For simplicity assume an unbounded domain and let $f(x,0) = f_0(x)$. Discuss the evolution of $f(x,t)$ from two complementary perspectives: (i) transfer function and (ii) Fourier transform.
Applications: The function f may represent solute concentration C_w in the stagnant water phase of a porous medium (7.14), pressure p in a confined aquifer (5.2), temperature T in a soil or an aquifer (8.21), or velocity v in a viscous fluid (2.20).

7.2 Salinization Consider a uniform medium with a constant water table at depth ℓ. Let the solute concentration in the groundwater be $C_0 > 0$ and assume a constant evaporative (upwards) flux j_w^{evap}. Finally assume that the corresponding velocities are so small that hydromechanic dispersion is negligible relative to molecular diffusion. Calculate and discuss the stationary concentration profile which will develop in this system. Neglect effects resulting from density changes and possible precipitation as the concentration in the water phase increases.

7.3 Apparent Transport Parameters in MIM Model For the MIM model discussed in Section 7.1.5, sketch $v(\langle x\rangle)$ and $D(\langle x\rangle)$ as they would be deduced from the travel distance pdf $p_x(x;t)$ and compare them to the corresponding parameters obtained from the travel time pdf $p_t(t;x)$.

7.4 Modeling Transport through Glass Beads Medium as MIM Model Consider the medium illustrated in Figure 7.10 and discuss the dependence of R and ω on the mean flow velocity in an equivalent MIM model.

7.5 Probability for Remaining in Mobile/Immobile Phase in MIM Model At $t = 0$, let the concentration be $C_{\text{im}} = 0$ in the immobile phase and $C_{\text{m}} > 0$ in the mobile phase. Calculate the concentration of particles that remain in the mobile phase after time t. Do the same for the complementary situation where at time $t = 0$ $C_{\text{m}} = 0$ and $C_{\text{im}} > 0$ and discuss the results.

7.6 Subscale Dispersion: Diffusion or Hydromechanic Dispersion? We assumed for Figure 7.12 that microscopic dispersion is only due to molecular diffusion. Calculate a characteristic scale for the textural features of the medium for this to be a reasonable assumption.

7.7 Peclet Number in Heterogeneous Medium Adapt the microscopic Peclet numbers defined by (4.15) to the transport simulations considered in Section 7.1.6. Calculate its value for the three cases of microscopic dispersion considered there, i.e., $\lambda_t = \{0, 0.0005, 0.0025\}$ m.

7.8 PDF from Line and from Point Source Notice that the maximum values of the pdfs that originate from the line and from the point source in Figure 7.12 on page 223 are almost two orders of magnitudes different and explain this.

7.9 Apparent Velocity in Heterogeneous Media and Subscale Dispersion Discuss the influence of microscopic dispersion on the apparent velocity of a particle distribution evolving from a point and from a line source as illustrated in Figure 7.14 on page 228.

8

Soil Heat

The thermal dynamics of soils is important for a number of environmental processes. Most prominent, certainly at larger scales, is the conversion of incoming solar radiation into heat and the distribution of the corresponding energy flux into various channels. These include longwave radiation, latent heat flux resulting from evaporating water, sensible heat flux caused by convection of air, and heat flux into the ground, all of which depend strongly on temperature as well as on soil water content. The distribution between the different channels has a strong impact on the microclimate. Just think of an oasis in the hot desert, of a cool forest on a hot summer's day, or of a steaming hot tropical forest. Here, as well as in many other cases, the dynamics of soil heat is closely coupled with that of soil water through the large specific enthalpy of water's phase transitions which is $0.333\ \mathrm{MJ\,kg^{-1}}$ for fusion and $2.26\ \mathrm{MJ\,kg^{-1}}$ for evaporation. An evaporation of $1\ \mathrm{mm\,d^{-1}}$ thus corresponds to a cooling of the soil surface with $26.2\ \mathrm{W\,m^{-2}}$ and a corresponding warming of the atmosphere upon condensation of the vapor.

Soil temperature is also a crucial variable for a number of processes within the soil. Examples include the activity of microbes which determines for instance the decomposition of soil organic matter and with it the CO_2- and CH_4-emissions from soils, the activity of plant roots, and the stability of soils, particularly in cold regions. Along more exotic routes, temperature profiles in boreholes are employed to reconstruct the paleoclimate and temperature is used as a "tracer", e.g., for the infiltration of warm water from a river or of rain into permafrost soils.

The thermal dynamics of geologic formations is also of interest for a number of engineering applications. These include the construction of buildings, roads, or pipelines in permafrost regions, the intermediate storage of heat in groundwater or dry rock, and the design of disposal sites for nuclear or municipal waste, both of which entail a significant generation of heat.

At vertical scales of tens of meters and beyond, the thermal dynamics of the subsurface is dominated by heat conduction driven at the surface by the fluxes of radiation, latent heat, and sensible heat, and from below by the geothermal heat flow, some $0.05\ \mathrm{W\,m^{-2}}$, which originates from radioactive decay and from the cooling of the Earth's core. As we will find, the temperature fluctuations induced at the surface get damped exponentially with depth with the penetration depth increasing proportional to the square root of the fluctuation's period. This simple regime is modulated by the thermal properties of the typically heterogeneous subsurface, which in turn depend on the water content. Further modifications result from heat convection by groundwater flow and near active geothermal zone.

In soils, and at vertical scales up to a few meters, the situation is much more complicated due to phase transitions of water in the unsaturated soil profile. As a consequence, the dynamics of water and thermal energy become closely coupled. This is particularly manifest in permafrost soils but also in arid regions where water does not evaporate right at the soil surface but at greater depths. The situation becomes even more complicated in anthropogenically modified environments with their modified thermal properties and artificial heat sources, for instance near waste disposal sites.

8.1
Thermodynamics of Water

Here, we look at water *per se* and at water in a porous medium from a thermodynamic perspective. This entails the conceptual separation of the system of interest from its environment in such a way that the system is influenced by the environment but not *vice versa*. We will furthermore restrict our consideration to equilibrium states, which implies that the relaxation times of all relevant system parameters is much shorter than the time on which the corresponding environmental variables change.

Experience shows that thermodynamic equilibrium states can be described by a few conjugate macroscopic variables the product of which yields the energy associated with the corresponding aspects of the system. In the context of our interest here, these encompass the (i) thermal energy with temperature T and entropy S, (ii) the mechanical energy with pressure p and volume V, (iii) the chemical energy of the composition with chemical potential μ_j and number N_j of moles of type j particles, and (iv) the interfacial energy with interfacial energy σ_{jk} and area A_{jk} between phases j and k. An equilibrium state is prescribed by controlling one of the conjugate variables in each of the pertinent pairs. For every choice of control variables, there exists a *thermodynamic potential* which has the properties that (i) it is linear in the size of the system and (ii) it becomes extremal for the system in equilibrium.

Next, we specifically look at a porous medium with water in the vapor (v), liquid (ℓ), and solid (s) phase. The *internal energy* $U(S, V, N_j, A_{jk})$ is the appropriate thermodynamic potential for the given, externally controlled, independent variables. As is the case for potentials in general, the differential of U is of most interest, i.e.,

$$\mathrm{d}U = T\,\mathrm{d}S - p\,\mathrm{d}V + \sum_j \mu_j\,\mathrm{d}N_j + \sum_{j<k} \sigma_{jk}\,\mathrm{d}A_{jk}\ . \tag{8.1}$$

For the system at hand, water in a porous matrix in a natural setting, V and S are not the appropriate control variables, however, but rather p and T. The corresponding thermodynamic potential is the *Gibbs energy* $G(T, p, N_j, A_{jk}) := U + pV - TS$. Inserting (8.1) yields for its differential

$$\mathrm{d}G = -S\,\mathrm{d}T + V\,\mathrm{d}p + \sum_j \mu_j\,\mathrm{d}N_j + \sum_{j<k} \sigma_{jk}\,\mathrm{d}A_{jk}\ . \tag{8.2}$$

A thermodynamic system whose variables S, V, N_j, A_{jk} are controlled externally will equilibrate to a state that minimizes the internal energy U. If, on the other hand, the controlled variables are T, p, N_j, A_{jk}, then the system will relax to a state that minimizes the Gibbs energy G. We will encounter further thermodynamic potentials below, in particular the enthalpy $H(S, p, N_j, A_{jk})$ and the Helmholtz free energy $F(T, V, N_j, A_{jk})$, all with the same property that a system relaxes to a state that minimizes them under the condition that the corresponding variables are fixed. These potentials can be transformed into each other through a Legendre transform in analogy to the above transition from U to G. Hence, the appropriate description for different physical situations can be readily obtained.

8.1.1
Free Water

Before looking into porous media, we study free water where the interfacial energy, the last term in (8.2), is negligible compared to the other terms. The phase diagram is characterized by two distinguished points and by three corresponding lines (Figure 8.1). The points are (i) the *triple point* at {0.01°C, 611 Pa}, where water can exist in any of the three phases, and (ii) the *critical point* at {374°C, 22.3 MPa}, where the distinction between liquid and vapor phase disappears. The corresponding lines are the evaporation (vapor pressure), the melting, and the sublimation curve, respectively.

Evaporation (Vapor Pressure) Curve Consider a mass element in which liquid (ℓ) and vapor (v) phase coexist in thermal and mechanical equilibrium with their surroundings, i.e., temperature and pressure are constant throughout the system. Since we consider a mass element, there is no exchange of mass with the surroundings. Hence, the Gibbs energy G is the appropriate thermodynamic potential for this system and its total energy is given by $G = G_\ell + G_v$ with differential

$$\mathrm{d}G = -S\,\mathrm{d}T + V\,\mathrm{d}p + \mu\,\mathrm{d}N\ . \tag{8.3}$$

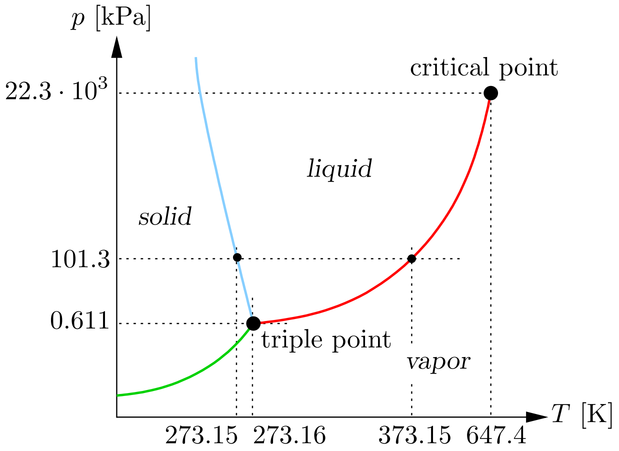


Figure 8.1. Schematic phase diagram of free water with evaporation (vapor pressure) curve (red), freezing curve (blue), and sublimation curve (green). Notice that axes are grossly out of scale and distorted.

With T, p, and N fixed externally, the system is still free to transfer mass between the two phases and to thus minimize its energy such that

$$\mathrm{d}G = \mu_\ell \, \mathrm{d}N_\ell + \mu_v \, \mathrm{d}N_v = [\mu_\ell - \mu_v]\mathrm{d}N_\ell = 0 \ , \tag{8.4}$$

where the second equality results from the total number of molecules being fixed. Hence, $\mu_\ell = \mu_v$. In equilibrium, with T and p controlled, the chemical potentials μ_j of all phases in contact with each other thus are equal.

Now consider the vapor pressure curve $p(T)$ that describes the chemical equilibrium between the liquid and the vapor phase. First notice with (8.3) that $\mu = \partial G/\partial N|_{T,p}$ does not depend on N, but will in general depend on T and p. Hence, the chemical potential $\mu = G/N$ is the Gibbs energy per mole and depends only on T and p. This may also be postulated directly from the linearity of the Gibbs energy with system size. Thus, with $N = \text{const}$,

$$\mathrm{d}\mu = \frac{1}{N}\mathrm{d}G = \frac{1}{N}\frac{\partial G}{\partial T}\,\mathrm{d}T + \frac{1}{N}\frac{\partial G}{\partial p}\,\mathrm{d}p = -s\,\mathrm{d}T + v\,\mathrm{d}p \ , \tag{8.5}$$

where $s = S/N$ and $v = V/N$ are molar entropy and volume, respectively. On the evaporation curve, T and p are related to each other, which we express by $p(T)$. In order to obtain this function, we take the derivative of the condition for chemical equilibrium, $\mu_\ell(T, p(T)) = \mu_v(T, p(T))$, which leads to

$$\frac{\partial \mu_\ell}{\partial T} + \frac{\partial \mu_\ell}{\partial p}\frac{\mathrm{d}p}{\mathrm{d}T} = \frac{\partial \mu_v}{\partial T} + \frac{\partial \mu_v}{\partial p}\frac{\mathrm{d}p}{\mathrm{d}T} \tag{8.6}$$

and further, using (8.5), to the *Clausius-Clapeyron* equation

$$\frac{\mathrm{d}p}{\mathrm{d}T} = \frac{s_\ell - s_v}{v_\ell - v_v} = \frac{\Delta s}{\Delta v} \ , \tag{8.7}$$

where Δs and Δv are the specific changes of entropy and volume, respectively, for the phase transition.

To transform (8.7) into a more convenient form, we consider the *enthalpy* $H(S, p, N_j, A_{jk}) = U + pV$ with differential

$$\mathrm{d}H = T\,\mathrm{d}S + V\,\mathrm{d}p + \sum_j \mu_j\,\mathrm{d}N_j + \sum_{j<k} \sigma_{jk}\,\mathrm{d}A_{jk}\ . \tag{8.8}$$

For the case considered here – constant pressure, equal chemical potentials and constant total mass, interfacial energy negligible – Δs may be expressed as $\Delta h/T$, where Δh is the specific change of the enthalpy. Choosing N such that a unit mass results, (8.7) becomes

$$\frac{\mathrm{d}p}{\mathrm{d}T} = \frac{H_{\ell v}}{T[\rho_v^{-1} - \rho_\ell^{-1}]}\ , \tag{8.9}$$

where $H_{\ell v}$ is the enthalpy per unit mass for the phase transition. Its value is 2.503 $\mathrm{MJ\,kg^{-1}}$ at 273 K and 2.259 $\mathrm{MJ\,kg^{-1}}$ at 373 K. Notice that $H_{\ell v} = H_v - H_\ell$, hence the change of sign in (8.9).

Finally we mention that $p(T)$ as given by (8.9) also corresponds to the partial pressure of the water vapor in a multicomponent system, i.e., water vapor in the air above a lake. It is thus also referred to as the *vapor pressure curve*.

A popular parameterization of the vapor pressure curve is the *Magnus formula*

$$p(T) = a\exp\Big(\frac{bT}{T+c}\Big)\ . \tag{8.10}$$

Buck [1981] gives the coefficients $a = 6.1121$, $b = 17.502$, and $c = 240.97$ for a maximum relative error of 0.2% for temperature T [°C] in the interval [−20, +50], and he also provides further, more accurate parameterizations.

Melting Curve In complete analogy to the evaporation curve we may also deduce the melting curve, the equilibrium between solid and liquid phase, as

$$\frac{\mathrm{d}p}{\mathrm{d}T} = \frac{H_{s\ell}}{T[\rho_\ell^{-1} - \rho_s^{-1}]}\ . \tag{8.11}$$

At 273 K, $H_{s\ell} = 0.333\ \mathrm{MJ\,kg^{-1}}$. Notice that, in contrast to most other substances, the melting curve of water has a negative slope as a consequence of the lower mass density of its solid phase.

Wagner et al. [1994] give for the parameterization of the melting curve of ice I, the one relevant for freezing in soils, in the temperature interval [251.165 K, 273.16 K]

$$\pi = 1 - a[1 - \vartheta^{-3}] + b[1 - \vartheta^{-21.2}] \tag{8.12}$$

with the normalized quantities $\pi = p/611.657$ Pa and $\vartheta = T/273.16$ K and with the coefficients $a = 626'000$ and $b = 197'135$.

8.1.2

Water in Porous Medium

In porous media, the contribution of interfaces to the internal energy U is no longer negligible. In the following, we look at three consequences of this: (i) the matric potential ψ_m already introduced with (3.19), (ii) the vapor pressure above a curved surface, and (iii) the soil freezing characteristic.

Matric Potential Consider a rigid, isothermal porous medium whose pore-space is divided between a liquid water phase (ℓ) and a gaseous phase (g) that consists of air (a) and water vapor (v). Assume the pressure in the gaseous phase to be constant, $p_g = p_v + p_a = \text{const}$, as would for instance be the case in soils under a wide range of conditions. The appropriate thermodynamic potential here is the *Helmholtz free energy* $F(T, V, N_j, A_{jk}) = U - TS$. Taking into account that $T = \text{const}$ and also $V_\ell + V_g = \text{const}$ because the medium is rigid, the differential of F becomes

$$\mathrm{d}F = -[p_\ell - p_g]\,\mathrm{d}V_\ell + [\mu_\ell - \mu_v]\mathrm{d}N_\ell + \sigma_{\ell g}\,\mathrm{d}A_{\ell g}\ , \tag{8.13}$$

where we have further used that the total mass of water is constant, hence $\mathrm{d}N_v = -\mathrm{d}N_\ell$. In thermodynamic equilibrium $dF = 0$ and further, for material equilibrium between the phases, $\mu_\ell = \mu_v$. Hence

$$\underbrace{p_\ell - p_g}_{\psi_m} = \sigma_{\ell g}\frac{\mathrm{d}A_{\ell g}}{\mathrm{d}V_\ell}\ . \tag{8.14}$$

We recognize that $\psi_m = p_\ell - p_g$ is the matric potential introduced with (3.19) when we did not yet have to distinguish between the pressure from the air and from the water vapor. As found earlier, ψ_m is the energy per unit volume of liquid water to bring it into a porous medium when only interfacial (capillary) forces are considered. Now, (8.14) shows that this energy is indeed invested in extending the interfacial area.

Vapor Pressure above Curved Interface We address this problem with the Boltzmann factor $\exp(-\Delta E/[kT])$ that gives the occupation probability for a state with energy ΔE with respect to that of a state with energy 0. In addition, we have to factor in the degeneracy of the two states, i.e., the number of states with the same energy. In this case – we are considering the transition across an infinitesimally thin interface – they are equal and proportional to the volume in the immediate neighborhood of the interface. Hence, the factors cancel out.

We obtain ΔE for the phase transition from the matric potential ψ_m, the energy required to move a unit volume of water from the vapor phase into the liquid phase. The energy per molecule of water is then $m_w\psi_m/[\rho_w N_A]$, where m_w is the molar mass of water and N_A is Avogadro's constant. For bound water $\psi_m < 0$, hence the pressure $p(\psi_m)$ of water in the vapor phase is lower

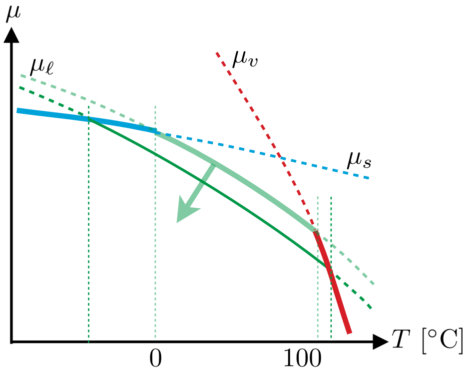


Figure 8.2. Chemical potential of solid (cyan), liquid (light green), and vapor (red) phase of free water. The dashed parts of the curves are unstable since the Gibbs energy is smaller in another phase. Interfacial forces in a porous medium reduce $\mu_\ell(T)$ as ψ_m becomes more negative when liquid water retreats into smaller pores. This leads to a freezing-point depression as well as to a boiling-point elevation (dark cyan). Notice that the axes are not to scale.

than p_0, the vapor pressure over a flat surface. Thus we obtain the so-called *Kelvin equation* [*Skinner and Sambles* 1972; *Fisher et al.* 1981]

$$\frac{p(\psi_m)}{p_0} = \frac{\rho_v(\psi_m)}{\rho_{v_0}} = \exp\left(\frac{m_w \psi_m}{\rho_w \mathrm{R} T}\right) , \tag{8.15}$$

where $\mathrm{R} = N_A k = 8.3144\,\mathrm{J\,mol^{-1}K^{-1}}$ is the universal gas constant and p_0, a function of T, is for instance described by (8.10). Notice that the mass density ρ_v of the water vapor, the concentration of water molecules, is proportional to p, hence (8.15) may also be read as a concentration ratio.

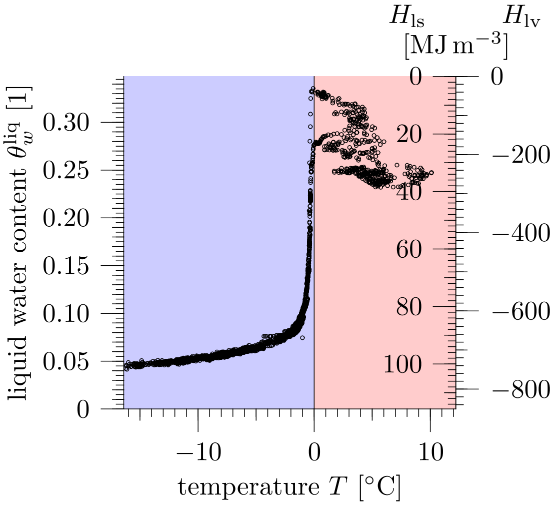
Soil Freezing Characteristic In a first step, we consider $\mu(T)$ for free water at some fixed external pressure and notice with (8.5) that $\frac{\partial \mu}{\partial T}|_p = -s$. Recalling that the entropy measures the disorder of a system, we realize that $s_{\text{solid}} < s_{\text{liquid}} < s_{\text{vapor}}$. Finally, since in a system at equilibrium with controlled T and p, the Gibbs energy is minimal and $\mu = G/N$, we obtain $\mu(T)$ for water as sketched in Figure 8.2. In such an unconstrained system and at standard conditions, water freezes at 0°C.

In analogy to the vapor pressure that gets reduced in porous media according to (8.15), the freezing point also gets depressed with decreasing radius of a pore. To understand this, we first recall that in thermodynamic equilibrium, the chemical potential μ in coexisting phases is equal. As temperature drops below 0°C in a porous medium, liquid water in the largest voids will freeze. As a consequence the liquid-ice interface will move into pores with smaller radii. Here, μ_ℓ, the Gibbs energy per mole, is reduced due to the attraction by the mineral surface and by the interfacial tension $\sigma_{\ell s}$ between the liquid and solid water phase. Recalling the matric potential ψ_m as energy per unit volume of soil produces, for a pore with radius R, the chemical potential

$$\mu_\ell(\psi_{\ell s}) = \mu_{\ell_0} + \frac{m_w}{\rho_w}\psi_{\ell s} = \mu_{\ell_0} - 2\frac{m_w}{\rho_w}\frac{\sigma_{\ell s}}{R} , \tag{8.16}$$

where $\psi_{\ell s}$ is the matric potential obtained from the water-ice interface and μ_{ℓ_0} corresponds to free water. For the second equality, the Young-Laplace

Figure 8.3. Experimentally determined soil freezing characteristic at the Bayelva field site on Svalbard. On the left, enthalpy axes are drawn for the solid-liquid and for the liquid-vapor phase transitions. They indicate the required transfer of energy to accomplish a given change in temperature or liquid water content. Instruments used are a PT-100 temperature sensor and a nearby TDR-probe for the liquid water content.



equation (3.2) with the radius of interfacial curvature given by $-R$ was used. The interfacial tension between water and ice is reported in the range between 0.027 and 0.032 J m^{-2} [*Hardy* 1977; *Huang and Bartell* 1995]. Application of the Young-Laplace equation is warranted because we are looking at equilibrium processes where the liquid-solid interface has sufficient time to relax into the energetic minimum through local melting and freezing.

For ice and liquid water in thermodynamic equilibrium we expect to find a close relation between temperature, which sets the chemical potential, and liquid water content. Clearly, this relation will depend on the material with more liquid water, at a given sub-freezing temperature, in fine-textured material than in coarse textured one. This is corroborated by Figure 8.3 which shows $(T, \theta_w^{\mathrm{liq}})$ data measured with two nearby sensors at a permafrost field site. First we notice that liquid water is indeed present at temperatures well below 0°C. Next, it is apparent that the expected relation between temperature and liquid water content holds quite precisely once temperatures are below about −0.5°C. It does not hold for higher temperatures without ice, however, despite the fact that the liquid-vapor system is also coupled by a phase transition. Why is this? In the ice-liquid system, both phases are quite immobile, the liquid water because it is located in narrow pores. Hence, the time scale of water movement is sufficiently long to allow thermal equilibrium to establish. In contrast, the liquid-vapor system has a much faster dynamics and, in addition, the external forcing through rainfall and evaporation is much stronger. Hence, an equilibrium can rarely establish.

8.2
Heat Conduction in Solids

Heat conduction is akin to solute diffusion with the random transfer of molecular kinetic energy formally corresponding to the random motion of particles. In the simplest case – heat conduction in uniform solids and diffusion in uniform fluids, respectively – both processes are described by the prototypical parabolic pde $\partial_t \diamond - D\nabla^2 \diamond = 0$, where D is a constant diffusion coefficient. Solutions of this equation are readily available for thermal [*Carslaw and Jaeger* 1990] as well as for solute problems [*Crank* 1975] and they are interchangeable through the associations

$$\mathbf{j}_h \sim \mathbf{j}_w \ , \qquad T \sim C_w \ , \qquad \rho c_h T \sim C_t \ . \tag{8.17}$$

8.2.1
Dynamics of Conductive Heat Transfer

We consider a solid with density ρ, specific heat capacity c_h, and thermal conductivity K_h. The conservation of thermal energy may then be written as

$$\partial_t[\rho c_h T] + \nabla \cdot \mathbf{j}_h = 0 \ , \tag{8.18}$$

where $\mathbf{j}_h$ is the heat flux, which may be described by the *Fourier law*

$$\mathbf{j}_h = -\mathsf{K}_h \nabla T \ . \tag{8.19}$$

Combination of the two leads to the *heat conduction equation*

$$\partial_t[\rho c_h T] - \nabla \cdot [\mathsf{K}_h \nabla T] = 0 \ , \tag{8.20}$$

which, for uniform and isotropic media with time-invariant properties, may be simplified to

$$\partial_t T - D_h \nabla^2 T = 0 \ , \quad D_h = \frac{K_h}{\rho c_h} \ , \tag{8.21}$$

where D_h is the thermal diffusivity, i.e., the diffusion coefficient for temperature. Material properties of some constituents of geologic formations are given in Table 8.1.

Solutions of the linear pde (8.21) are readily obtained for various initial and boundary conditions. For simplicity, we focus in the following on a uniform medium with a flat surface at $z = 0$, unbounded for $z > 0$, and a one-dimensional dynamics driven by temperature $T_0(t)$ at the surface. This may be formulated as

$$\begin{aligned} \partial_t T - D_h \partial_{zz} T &= 0 \ , \\ T(z;0) &= 0 \ , \\ T(0;t) &= T_0(t) \ , \\ \lim_{z\to\infty} T(z;t) &= 0 \ . \end{aligned} \tag{8.22}$$

Since this problem is linear in T, the principle of superposition may be applied and the forcing may be decomposed in a convenient way. Two contrary approaches are the decomposition into Dirac-functions $\delta(t)$ or into periodic functions $\sin(\omega t - \phi(\omega))$. The former leads to the transfer function introduced in Section 7.1.4 and is most convenient if $T_0(t)$ describes a localized event. The latter corresponds to a Fourier-decomposition and is detailed in Section A.5.2 in the appendix. It is most convenient for long periodic or quasi-periodic forcings. Since both decompositions are based on a complete set of functions, they may of course also be applied to situations for which they are not optimal, e.g., the Fourier-decomposition to a localized or a non-periodic forcing. The operational price to pay then is that a much larger of significant components has to be handled which may lead to a high computational cost.

Before embarking on solutions of (8.22), we remark that prescribing a temperature forcing at the surface is a hard thing to do physically and that it would be much easier experimentally to prescribe the heat flux. Indeed, recalling the discussion in Section 7.1.1, the natural independent variable of $T(z;t)$ is z, not t, while it would be the other way round for $j_h(t;z)$. The motivation to nevertheless choose the formulation (8.22) comes from typical observational procedures where temperature sensors are installed at fixed locations, either in air or in the ground, and T is recorded as a function of time. The more direct approach, measuring the flux which after all generally drives the system, is not possible since corresponding instruments are not available, at least not for the subsurface.

Transfer Functions The first impulse for solving (8.22) with transfer functions may be to decompose $T_0(t)$ into δ-functions and to then seek the solution for $T_0(t) = \delta(t)$. While this may be mathematically correct, though not dimensionally, and a solution would be readily obtained, the symbols would not have their intended meaning. Indeed, solving (8.22) with $T_0(t) = \delta(t)$ prescribes a flux – t is the natural variable of flux, not of temperature – and the result also corresponds to the heat flux, despite the disguising notation.

Table 8.1. Typical values for mass density ρ, specific heat capacity c_h, thermal conductivity K_h, and thermal diffusivity $D_h = K_h/[\rho c_h]$ of some constituents of soils and geologic formations.

	ρ $[10^3\ \mathrm{kg\,m^{-3}}]$	c_h $[10^3\ \mathrm{J\,kg^{-1}K^{-1}}]$	K_h $[\mathrm{J\,m^{-1}s^{-1}K^{-1}}]$	D_h $[10^{-6}\ \mathrm{m^2s^{-1}}]$
air	0.00116	1.007	0.025	21.4
quartz	2.65	0.84	8.8	3.95
humus	1.4	1.9	0.25	0.094
water	1.00	4.22	0.57	0.14
ice	0.91	2.11	2.2	1.15

As an incidental remark, j_h also satisfies the temperature diffusion equation, as differentiation of $\partial_t T - D_h \partial_{zz} T = 0$ with respect to z readily shows.

We follow a different approach here and seek a solution to a variant of the heat conduction problem (8.22), namely to

$$\begin{aligned} \partial_t T - D_h \partial_{zz} T &= 0 \ , \\ T(z;0) &= 0 \ , \\ -K_h \partial_z T(z;t)|_{z=0} &= e_0 \delta(t) \ , \\ \lim_{z\to\infty} T(z;t) &= 0 \ , \end{aligned} \tag{8.23}$$

where a heat pulse at time $t = 0$ deposits the energy density e_0 $[\mathrm{J\,m^{-2}}]$ into the ground at its surface. In this formulation, heat flux through the surface is zero, except for $t = 0$, hence the injected thermal energy will remain in the ground and just get distributed with time. Solving (8.23), most easily through a Laplace transform with respect to time, yields

$$T^\delta(z;t) = \frac{e_0}{\rho c_h} \frac{1}{\sqrt{\pi D_h t}} \exp\Big(-\frac{z^2}{4D_h t}\Big) \ , \tag{8.24}$$

where (8.21) has been used. Integrating $\rho c_h T^\delta(z;t)$ with respect to z over the entire medium for $t > 0$ yields e_0, as expected. The corresponding heat flux is obtained from Fourier's law as

$$j_h^\delta(t;z) = \frac{e_0 z}{\sqrt{4\pi D_h t^3}} \exp\Big(-\frac{z^2}{4D_h t}\Big) \ . \tag{8.25}$$

Integrating $j_h^\delta(t;z)$ with respect to t over all times $t > 0$ yields e_0, again as expected.

In complete analogy to Section 7.1.1, we may now define the transfer functions for heat conduction, namely the travel distance pdf

$$p_z(z;t) := \frac{\rho c_h}{e_0} T^\delta(z;t) = \frac{1}{\sqrt{\pi D_h t}} \exp\Big(-\frac{z^2}{4D_h t}\Big) \tag{8.26}$$

for the transfer of thermal energy in space and the travel time pdf

$$p_t(t;z) := \frac{1}{e_0} j_h^\delta(t;z) = \frac{z}{\sqrt{4\pi D_h t^3}} \exp\Big(-\frac{z^2}{4D_h t}\Big) \tag{8.27}$$

for the corresponding transfer in time.

The travel distance pdf (8.26) is readily recognized as a Gaussian, actually only one half of it due to the semi-infinite medium and the missing convection. This also explains the factor of 2 between (8.26) and the far-field solution (7.27) for solute transport. Notice that the temperature at the surface decays proportional to $1/\sqrt{\pi D_h t}$, where t is the time after the heat input.

The travel time pdf (8.27) has a more complicated form (Figure 8.4). At first glance, it could be interpreted as a heat pulse moving through the

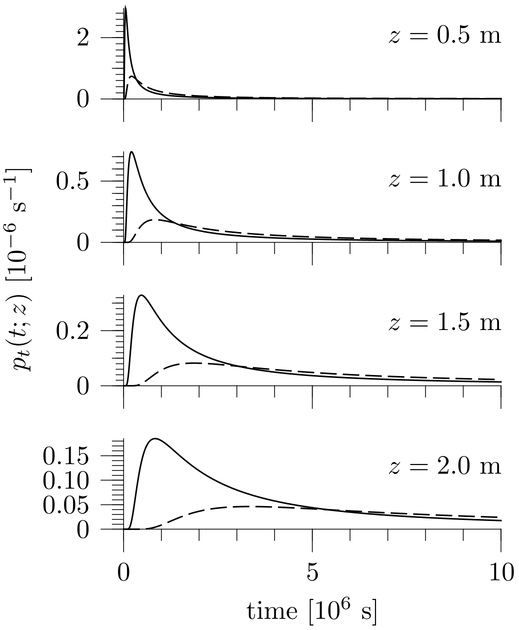


Figure 8.4. Transfer function $p_t(t;z)$, given by (8.27), for projecting time series of temperature or heat flux from the surface at $z = 0$ to depth z as described explicitly by (8.30). The lines are for $D_h = 2 \cdot 10^{-7}$ m^{2}s^{-1} (dashed) and $8 \cdot 10^{-7}$ m^{2}s^{-1} (solid). The values bracket the range typically encountered in mineral soils. The very long tails are characteristic for a pure diffusion process.

ground. However, this is not the case, since this is just a spreading Gaussian as p_z reveals. Such a spreading necessarily leads to a maximum at any depth $z > 0$ since there is not flux for very short times, before any energy has reached depth z, and again a negligible flux for very large times, after the thermal energy has spread over a large region. This maximum of the flux occurs at time

$$t_{\max} = \frac{z^2}{6D_h} \qquad \text{with} \qquad p_t(t_{\max}; z) = \sqrt{\frac{6}{\pi}} \frac{3}{\exp(3/2)} \frac{D_h}{z^2} \; . \tag{8.28}$$

Again, $t_{\max} \propto z^2/D_h$ is expected for the diffusion process.

The transfer functions p_z and p_t allow us to project "thermal information" from the surface to greater depths. Whether this information is temperature or heat flux is not relevant as long as it is decomposed into a temporal succession of "energy elements" at the ground surface. Indeed there is an equivalence between the corresponding initial and boundary conditions,

$$T(z;0) = \frac{e_0}{\rho c_h}\delta(z) \quad \sim \quad j_h(t;0) = e_0\delta(t) \; . \tag{8.29}$$

In the first one, the flux through the upper boundary is always 0 and the energy e_0 [J m^{-2}] is present initially and right at the upper boundary. In the second case, there is no initial energy present but the quantity e_0 is supplied, again right at the surface, through the boundary flux which is then again 0.

We are now finally in the position to write down the solution of (8.22) for the arbitrary surface temperature $T_0(t)$, namely

$$T(z;t) = \int_0^t T_0(\tau) p_t(t-\tau;z)\,\mathrm{d}\tau \tag{8.30}$$

for the temperature as a function of time t at a fixed depth z.

Periodic Solutions As detailed in the Appendix, pages $323f$, the solution of (8.22) for the periodic forcing $T(0;t) = \alpha\sin(\omega t)$ may be written as

$$T(z;t) = \alpha\exp(-kz)\sin(\omega t - kz)\ , \quad k = \sqrt{\frac{\omega}{2D_h}} \tag{8.31}$$

where k is the wave number and $z_0 = 1/k$ the penetration depth.

Apparently, the periodic excitation at the surface leads to a strongly damped heat wave into the ground. The wave number k, which determines the phase shift with depth and also the penetration depth, depends on the angular frequency ω of the excitation: higher frequency fluctuations penetrate faster – their phase velocity is $v_{\rm ph} = \omega/k = \sqrt{2D_h\omega}$ – but they are also damped away in more shallow layers. Some examples for the damping depth k^{-1} are given in Table 8.2 on page 256.

We only considered forcing with a single frequency here. However, (A.127) shows how an arbitrary forcing can be handled: The Fourier transform of the surface temperature – $\widetilde{T}(0;\omega)$, which gives amplitude and phase of frequency component ω – is integrated with the response to the single-frequency forcing as kernel. The complex notation used in the appendix is readily adapted to the real notation used here and (A.127) may be written as

$$T(z;t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \alpha(\omega)\exp(-kz)\sin(\omega t - kz - \varphi(\omega))\,\mathrm{d}\omega\ , \tag{8.32}$$

where $\alpha(\omega)$ and $\varphi(\omega)$ are the amplitude- and the phase-spectrum of $T(z;t)$, respectively. Notice that the dependence of k on ω has been suppressed for notational clarity.

Comments Heat conduction is a linear process, hence the principle of superposition is applicable, and has been applied above already. It is a powerful tool to build solutions for more comprehensive problems from some building blocks. Solutions of (8.22), with its particular initial and lower boundary condition which are both $T = 0$, are optimal for describing fluctuations driven at the ground surface. In the simplest case, they are just superimposed on some constant temperature $T \neq 0$.

Other building blocks may account for heat transfer from deeper layers, for the decay of some initial distribution of thermal energy, or for the generation of thermal energy within the ground. Examples for transfer from deeper layers include the geothermal flux $j_{h_{\rm geo}}$, which leads to $T(z) = T(0) - \frac{j_h^{\rm geo}}{K_h} z$, or some other heat source below the region of interest, for instance a deposit of nuclear waste.

The decay of an initial temperature distribution may be described in analogy to (8.22) by

$$\begin{aligned}\partial_t T - D_h \partial_{zz} T &= 0 \\ \rho c_h T(z;0) &= e_0 \delta(z - z_0) \\ T(0;t) &= 0 \\ \lim_{z\to\infty} T(z;t) &= 0 \ . \end{aligned} \tag{8.33}$$

Carslaw and Jaeger [1990] give as a solution for this problem

$$T^\delta(z;t;z_0) = \frac{1}{2\sqrt{\pi D_h t}} \left[\exp\left(-\frac{[z-z_0]^2}{4D_h t}\right) - \exp\left(-\frac{[z+z_0]^2}{4D_h t}\right)\right] \ , \tag{8.34}$$

where the second term represents a negative mirror source which compensates the effect of the positive source at z_0 such that the boundary condition $T(0;t) = 0$ is satisfied. Physically, this corresponds to the loss of the thermal energy that was input into the system at time 0 and depth z_0.

Finally we realize that, while heat conduction *per se* is a linear process, a heat transfer problem may become nonlinear through the boundary conditions. The most prominent example of this is a radiative boundary, where heat flux is proportional to T^4.

8.2.2
Ground Temperature as Archive for Paleoclimate

Contemplating the evolution of glaciers over the past decades and the remnants of ancient moraines, we realize that our climate is not stable and probably never was for very long periods of time. Clearly, we would want to know the dynamics of the paleoclimate in order to better understand the working of the climate system as such, to assess the environmental conditions encountered by ancient cultures and to gain an appreciation of our civilization's impact on the current and future climate.

A first insight into the paleoclimate is offered by global and regional temperature records. Unfortunately, direct measurements only started in the 19th century and we are forced to use proxy information from which temperature can be deduced. These are required to provide two pieces of information, the temperature and the corresponding time.

Environmental Archives Classical examples of environmental archives are old trees, sediments, and ice shields. The thickness of *tree rings* reflects growing conditions with an annual resolution. These rings are dated by their ^{14}C-concentration or, more directly and exactly, by correlation with tree-ring lines that have been constructed from overlapping samples and date back

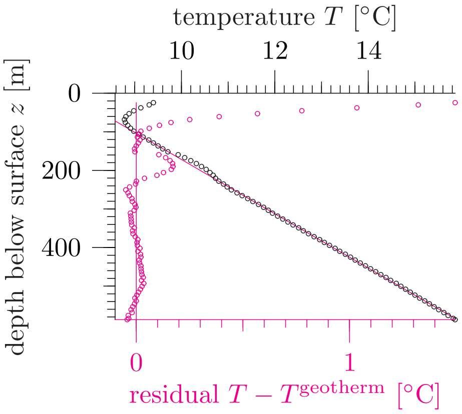


Figure 8.5. Temperatures measured in 1978 in a deep borehole at 75.7°W 45.4°N (black symbols). Estimated mean thermal conductivity at this site is 2.37 $\mathrm{W\,m^{-2}K^{-1}}$ and the mean geothermal gradient (red line) is 0.014 $\mathrm{K\,m^{-1}}$. The deviation from the mean temperature (magenta symbols) indicates a regional warming in the recent past. [*Data from* `www.geo.lsa.umich.edu/climate`.]

some 14 ky. One factor that affects the growing conditions is temperature. However, there are others like availability of water, height and exposition, pests, and competition with neighboring trees.

Sedimentary environments bury a fingerprint of the environment in dominantly chronological order. Thereby they generate very diverse and complicated archives that range from isotopic and chemical compositions to entire organisms. Retrieving for instance the abundance and distribution of pollen unfolds the vegetation cover of past times which in turn allows inferences on the climate.

Special sedimentary environments are *ice shields* which accumulate mainly water with traces of dust and preserve air samples in small bubbles. Information on past temperatures are deduced from D/H- and $^{18}\mathrm{O}/^{16}\mathrm{O}$-ratios. These reflect the fractionation between light and heavy isotopes during the evaporation of ocean water and the condensation of rain. A 420 ky record has been obtained from the Vostok ice core [*Petit et al.* 1999] and a still longer one, 800 ky, from the Dome C core [*EPICA* 2004]. All these archives yield information over long stretches of time and with a very high temporal resolution. However, the relation between the directly measured quantities and the desired temperature is rather circumstantial.

A more direct source of information are temperature logs from deep boreholes (Figure 8.5). There exist thousands of such boreholes, mainly from measurements of the global geothermal heat flux and from permafrost studies, and several hundreds of them are suitable as temperature archives of the past few millenia [*Pollack et al.* 1998].

Borehole Temperature Logs Recalling that temperature and time are required for an archive to be useful, how is paleo-temperature recorded in a borehole log? Apparently, surface ground temperature varies in a complicated way with dominant diurnal and annual periodicity. Lower frequency quasi-periodic contributions stem from decadal variations like the North Atlantic Oscillation (NOA) or the El-Niño-Southern Oscillation (ENSO). On top of this there appear temperature excursions like the *Little Ice Age* from about 1550 to 1850 AD, the preceeding *Medieval Climate Optimum* from 800 to 1300 AD, the *Younger Dryas*, a cold period of some 1.3 ky duration that ended some 11.6 ky ago, and finally the end of the *Last Ice Age*, some 14 ky ago. As a consequence of (8.31), the different components of this complicated signal are separated since they are damped away at different characteristic depths $k^{-1} = \sqrt{2D_h/\omega}$ according to their angular frequency ω (Table 8.2).

Beyond a certain depth, the fluctuations become negligible and the temperature profile is dominated by the geothermal heat flux which results from the cooling of the Earth's core and from radioactive decay. It amounts to some 0.05 $\mathrm{W\,m^{-2}}$ with a large spatial variation that depends on the thickness of the crust and on the vicinity to volcanically active zones. This flux leads to an average geothermal gradient of 0.025…0.03 $\mathrm{K\,m^{-1}}$. It is higher, up to 0.05 $\mathrm{K\,m^{-1}}$, in active regions and lower, down to 0.005 $\mathrm{K\,m^{-1}}$, in thick subduction zones.

Propagation of Temperature To study the qualitative picture described above in more detail, first recall that heat conduction is a linear process, provided that contributions from phase transitions and movement of water may be neglected. Hence, the geothermal gradient may be separated from the propagation of the temperature fluctuations at the ground surface. As a realistic example for the forcing, we choose the reconstruction of the mean paleo-temperature for the Northern hemisphere proposed by *Mann and Jones* [2003]. We refer to it as $T_{\mathrm{M-J}}(t)$ (Figure 8.6). Since this reconstruction only dates back to 200 AD, we presume that the temperature profile for earlier times corresponded to equilibrium with the constant geothermal heat flux. We finally notice that regional variations can be much larger than indicated by this global estimate.

Uniform Medium As a first approximation, we consider the medium to be thermally uniform with $D_h = 2.8 \cdot 10^{-6}\ \mathrm{m^2s^{-1}}$, corresponding to some solid rock. Then, we convince ourselves that annual and certainly diurnal fluctuations may be neglected despite their high amplitude of several hundred $\mathrm{W\,m^{-2}}$. Indeed, assuming an amplitude of 100 $\mathrm{W\,m^{-2}}$ for the annual cycle

Table 8.2. Some values for the damping depth $k^{-1} = \sqrt{2D_h/\omega}$, given by (8.31), of phenomena with different characteristic times. Values are given for two uniform materials with thermal diffusivity $D_h = 0.5 \cdot 10^{-6}\ \mathrm{m^2s^{-1}}$ and $D_h = 3 \cdot 10^{-6}\ \mathrm{m^2s^{-1}}$ which are representative for soil and solid rock, respectively.

phenomenon	time [y]	k_{soil}^{-1} [m]	k_{rock}^{-1} [m]
diurnal cycle	0.0027	0.12	0.28
annual cycle	1	2.2	5.5
little ice age	300	39	95
last glacial maximum	17'500	296	726

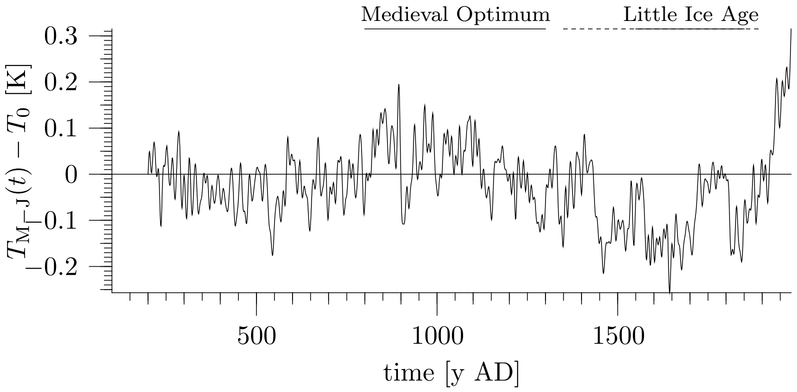


Figure 8.6. Reconstruction of mean global temperature for Northern hemisphere proposed by *Mann and Jones* [2003]. The reference temperature T_0 is arbitrarily set equal to the temperature at the beginning of the time series. [*Data from* `www.ncdc.noaa.gov/paleo/pubs/mann2003/mann2003.html`.]

at the surface, we find with (8.31) that it drops to the level of the geothermal flux at some 40 m depth. For the diurnal fluctuation, assuming an amplitude of 400 $\mathrm{W\,m^{-2}}$, the same is true at about 2.5 m depth. With these premises, the temperature anomaly is readily obtained from (8.30) as the convolution integral

$$T(z;t_1) - T^{\mathrm{geotherm}}(z) = \int_{t_0}^{t_1} T_{\mathrm{M-J}}(\tau) p_t(t_1 - \tau; z)\,\mathrm{d}\tau \ , \tag{8.35}$$

where $t_1 = 1980$ y AD, the end of the Mann-Jones reconstruction. The kernel $p_t(t;z)$ is defined by (8.27) and illustrated in Figure 8.4.

In the current contex, we may interpret the surface temperature as a stream of information – on past climate conditions – that gets filtered as it moves deeper into the ground (Figure 8.7). Apparently, the information is smeared rapidly with depth. While the temperature at $z = 50$ m reflects the average over the past few decades, the temperature at $z = 100$ m is already influenced by the history over the past few centuries. Finally, at $z = 800$ m, there only remains a very broadly averaged and quite weak signal with the amplitude decreased by some two orders of magnitude compared to the input. We also notice the phase shift that becomes apparent at the greatest depth. This is understood easily as a consequence of the shift of p_t. It could also be interpreted in terms of the Fourier representation (8.32), however. After all, (8.30) and (8.32) provide just two different descriptions of the same phenomenon.

Clearly, $T(t)$ is not observable, at least not on the long time scale of interest here. That was why we were interested in the temperature profile $T(z)$ in the first place. The anomaly calculated with (8.30) down to 1500 m below the surface is shown as thick line in Figure 8.8. We first notice that

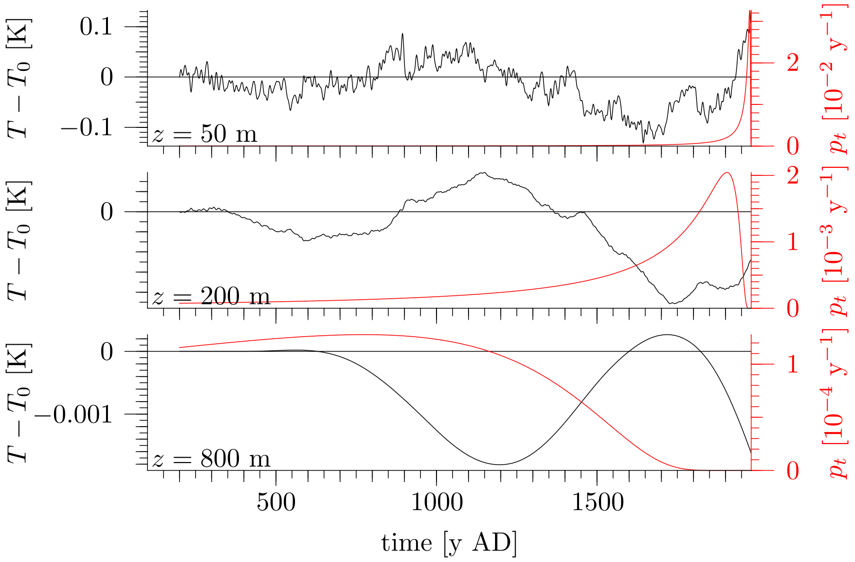


Figure 8.7. Projection of $T_{\mathrm{M-J}}(t)$ shown in Figure 8.6 to greater depths by (8.35) for a uniform medium with $D_h = 2.8 \cdot 10^{-6}$ m^{2}s^{-1}. The red lines represent the kernel $p_t(t_{\max} - t; z)$ at the particular depth z.

the strong warming since about 1850 AD is clearly reflected in the rapid drop of ground temperature with depth and also the Little Ice Age is clearly perceptible. However, already the Medieval Optimum is not very pronounced and only becomes visible after amplification. This demonstrates the exquisite measuring accuracy that is required for looking into the more distant past with this method.

Nonuniform Medium Uniform thermal properties may be a reasonable assumption for solid rock at greater depths. It is usually not correct for the topmost unconsolidated layers, however. While the thermal diffusivity of solid rocks is around $3 \cdot 10^{-6}$ m^{2}s^{-1}, it is typically much lower for soils, by a factor of about 5, and it varies with composition and water content. Adapting (8.20) to include non-uniform thermal properties but sticking to a one-dimensional and isotropic model yields

$$\partial_t T - \frac{1}{C_h} \partial_z \left[K_h(z) \partial_z T \right] = 0 \; . \tag{8.36}$$

This description is still linear and could be cast into the form (8.30). However, the kernel p_t would be much more complicated and the resulting expression would be difficult to interpret. It is thus advantageous to switch to a numerical solution at this point. Indeed, this is faster and more accurate than the evaluation of the convolution integral already for the case of the uniform medium. Simple problems like the one posed by (8.36) with a

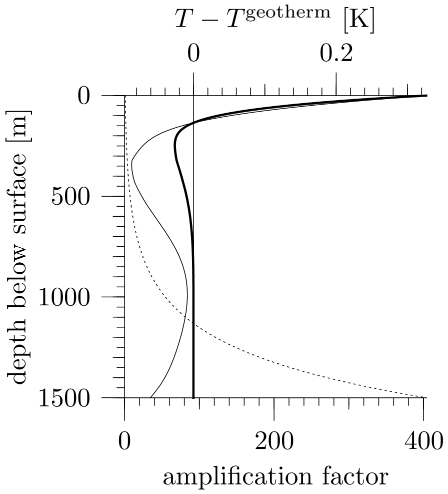


Figure 8.8. Temperature anomaly calculated for a thermally uniform medium with $D_h = 2.8 \cdot 10^{-6}$ m^{2}s^{-1} (thick line) forced by ground surface temperature of Mann-Jones reconstruction shown in Figure 8.6. The original curve amplified by $\exp(\alpha z)$ with $\alpha = 0.004$ m^{-1} (dotted line) is represented by the thin line.

moderately variable function $D_h(z)$ do not require specialized codes but can be tackled by general purpose solvers. As an example, Figure 8.9 shows a complete *Mathematica* code to read data on surface temperature, specify thermal properties, numerically solve the heat conduction equation (8.36), output the results into a datafile, and produce a corresponding graphical representation.

To see the impact of a low-diffusivity layer, consider a medium that consists of three layers – soil, sediment, solid rock – with corresponding thermal capacities of $\{2.4,\ 2.6,\ 1.7\}\cdot 10^6$ J m^{3}K^{-1} and thermal conductivities of $\{1.15,\ 3.90,\ 4.76\}$ J m^{-1}s^{-1}K^{-1} (Figure 8.10). The large difference between the temperature anomalies in the uniform and in the non-uniform medium demonstrates the importance of an accurate representation of the thermal properties for reconstructing past temperatures from borehole logs.

Reconstruction of Ground Surface Temperature Given measurements $T(z)$ of the ground temperature together with information on the thermal properties $C_h(z)$ and $K_h(z)$, reconstruction of the history $T_0(t)$ of the ground surface temperature is a typical inverse problem that may be approached with methods introduced in Section B.2. While this is straightforward in principle, the practical obstacles are significant. An obvious issue is temporal resolution which decreases rapidly as we move into the past. The reason for this is apparent from (8.30) – $T(z)$ results from the convolution of $T_0(t)$ with a kernel that rapidly widens with depth – and was already illustrated in Figure 8.7. A further consequence of the convolution is the limited scope of the method. This results from the uncertainty of the geothermal gradient G which emanates from uncertainties in the thermal conductivity $K_h(z)$. Indeed, for stationary heat flow $G = \partial_z T = -j_h/K_h$ and thus

$$\operatorname{var}\left(\frac{G'}{\overline{G}}\right) = \operatorname{var}\left(\frac{K_h'}{\overline{K}_h}\right) , \tag{8.37}$$

```
(* set working directory to "dir", where the data files are located *)
SetDirectory["dir"];

(* integration bounds *)
Tmin = 200; Tmax = 1980; Zmax = 700;

(* thermal properties in units of J, m, y, and K;
   smoothed by allowing 4 m depth for the transitions between layers *)
Kh = Interpolation[{{0,  36 10^6},  {46,  36 10^6},
                   {54, 123 10^6}, {196, 123 10^6},
                  {204, 150 10^6}, {Zmax,150 10^6}},
                  InterpolationOrder -> 1];
Ch = Interpolation[{{0, 2.4 10^6},  {46,  2.4 10^6},
                   {54, 2.6 10^6}, {196,  2.6 10^6},
                  {204, 1.7 10^6}, {Zmax, 1.7 10^6}},
                  InterpolationOrder -> 1];

(* geothermal gradient [K/m] (typical value for rock would be 0.014) *)
G = 0.0;

(* upper boundary: surface ground temperatures [Mann and Jones, 2003] *)
T0 = Interpolation[ReadList["mann-jones-03-N.dat", {Real,Real}]];

(* numerical solution *)
T = NDSolve[{
      D[u[t, z], t]==D[Kh[z] D[u[t,z],z],z]/Ch[z], (* PDE *)
      u[Tmin, z]==T0[Tmin] + G z,                  (* initial condition *)
      Derivative[0, 1][u][t, Zmax]==G,             (* lower boundary *)
      u[t, 0]==T0[t]},                             (* upper boundary *)
      u, {t, Tmin, Tmax}, {z, 0, Zmax}]            (* domain *)

(* export data *)
Export["mann-jones-03-N-z-nonunif.dat",
  Table[Evaluate[{z,u[Tmax, z]}/.First[T]] - T0[Tmin], {z,0,Zmax}],"TSV"];

(* plot data *)
Plot3D[Evaluate[u[t, z]/.First[T]], {t,Tmin,Tmax}, {z,0,Zmax/2},
  ViewPoint->{1.947,-3.374,0.943}, PlotPoints->100,
  PlotRange->All];

(* cleanup *)
Clear[Tmin,Tmax,Zmax,Dh,G,T0,T]
```

Figure 8.9. *Mathematica* code for solving heat conduction problem with spatially varying thermal diffusivity. The model is forced by the surface temperature reconstruction of *Mann and Jones* [2003] for the Northern hemisphere. The corresponding data for the years 200...1980 are read from file `mann-jones-03-N.dat` which is also displayed by Figure 8.6. Some results are shown in Figure 8.10.

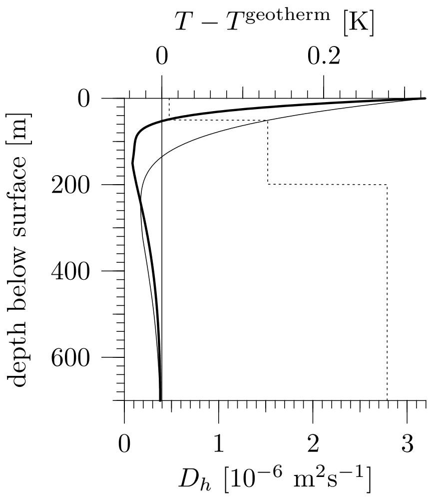


Figure 8.10.
Temperature anomaly calculated for a non-uniform medium, where D_h is piece-wise constant (dotted line). As a reference, the anomaly for the uniform medium with $D_h = 2.8 \cdot 10^{-6}$ m^{2}s^{-1} is drawn as thin line.

where $K_h = \overline{K}_h + K'_h$ is a decomposition of K_h into mean and perturbation and analogously also for $G(z)$. Hence, the relative uncertainty of G equals that of K_h, provided j_h is known accurately. Thus, traces of the ground surface temperature can only be detected as long as the resulting gradient is larger than the uncertainty of the measured thermal conductivities.

Further, and less obvious limitations are related to changing environmental conditions. As is illustrated by Figure 8.10, $T(z)$ responds rather strongly to the thermal properties of the surface layers. These depend on the water content which in turn reflects mean precipitation and mean evapotranspiration. These factors strongly change with climate, vegetation cover, and topography.

During the entire analysis so far, we presumed that heat conduction is the only relevant process. While this is in general a very good approximation, there are instances where convection processes become dominant. This is in particular the case in environments with a strong groundwater flow which in turn depends on the regional hydrology, hence again on climate and vegetation.

Finally, we go back to the example shown in Figure 8.5. At first sight, it is tempting to interpret the anomalies at 200 m and around 500 m in terms of past temperature changes. As it turns out, however, they cannot be recovered by an inversion, hence they do not result from heat conduction that is forced at the surface. Instead, they have to be ascribed to some of the effects discussed above, presumably to groundwater flow.

Relation between Ground Temperature and Air Temperature Once the history of the ground surface temperature is reconstructed from measured depth profiles, we may wish to transfer this information into other compartments for a comparison with other proxies like ^{18}O or with results from global circulation models (GCMs). As it turns out, this is no trivial task.

In particular, the naive first idea that ground surface temperature T^{ground} and near-surface air temperature T^{air} are identical is quite wrong. Reasons for a difference between the two include the latent heat of evaporating water or intervening layers like vegetation or snow. Such differences may actually be quite substantial. For instance, *Zhang et al.* [2005] found for Canada during the 20th century that $T^{\text{ground}} - T^{\text{air}}$ ranged from -2 to $+7$ K and that also the rate of change of the mean temperatures differed significantly. What makes the transfer of information particularly difficult is the probable change of the soil-atmosphere coupling with a changing climate. Warming in a cool climate will for instance reduce the mean height of a snow cover as well as the duration of the snow season. The reduced mean albedo will lead to a more efficient warming of the ground. Similarly, an increase of precipitation will lead to a higher water content in the soil, hence to a reduced thermal diffusivity, to a decreased albedo, hence a stronger warming by incoming radiation, and to an increased cooling by evapotranspiration. The net effect of these changes is difficult to assess and depends on local details. Finally, a changing climate typically entails a change of the vegetation, from steppe, to gras- and bushland, to forest or vice versa, which has a strong impact on the soil-atmosphere coupling. Similarly stronger impacts, but without associated climate changes, arise from changing land use, for instance when forests are turned into agricultural fields.

In a first round, all these effects make it difficult, maybe even impossible, to reconstruct the paleo-temperature. In a second round, however, with deeper understanding of the processes involved and probably with multiple proxies, they also offer opportunities to understand more complex aspects of the paleo-climate as a whole. This is indeed one of the motivations for further detailed studies of soil physical processes.

8.3
Heat Conduction in Porous Media

Conduction of heat in an unsaturated porous medium is a very complicated processes because (i) the constituent materials have quite different thermal properties (Table 8.1) with the geometric configuration strongly depending on the water content (Figure 8.11) and (ii) the macroscopic conduction of heat is microscopically in general a complicated multiphase process with strong convective components.

To appreciate the second aspect, consider a macroscopic volume of unsaturated soil at constant temperature, with the liquid and the vapor phase in thermodynamic equilibrium. For simplicity assume $T > 0°\text{C}$. Next, apply a macroscopic temperature gradient. Changing the temperature shifts the equilibrium between the phases, which involves the associated latent energy. The temperature gradient then leads to a corresponding gradient of the vapor

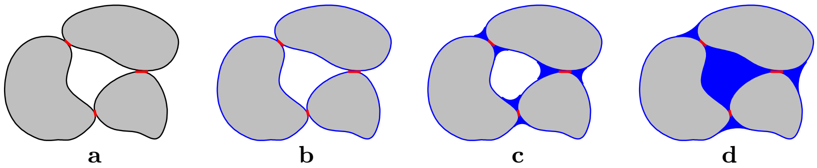


Figure 8.11. Sketch for dependence of thermal conductivity on water content in a coarse-textured porous medium. The contact between the grains is restricted to small regions (red) and the corresponding cross-sectional area is limiting for heat flow in a completely dry medium (**a**). As the water content increases, the pathways widen considerably thereby leading to a higher conductivity (**b**. . . **d**).

pressure p_v and of the matric potential ψ_m, the latter through the dependence of the interfacial tension σ_{wa} on temperature. The initial gradient of p_v, together with molecular diffusion, induces a vapor flux, from warm to cold regions, and with it a flux of latent heat. In addition, the initial gradient of ψ_m leads to a flow of liquid water which, since $\frac{\mathrm{d}\sigma_{wa}}{\mathrm{d}T} < 0$ (see Section E.2), is also from warm to cold regions. Depending on the thermal and hydraulic conductivities in the considered volume and on its coupling with the environment, this flow of water in the liquid and in the vapor phase may be sustained or, if it is hindered, water accumulates at the cold end such that a counteracting gradient develops.

From the macroscopic perspective, the various microscopic processes are no more visible individually but become part of material properties, the effective heat capacity and the effective thermal conductivity.

8.3.1
Effective Material Properties

We define the effective material properties of a porous medium in analogy to the corresponding properties of a solid. In particular, the effective heat capacity represents the change of thermal energy, both sensible and latent, in response to a change in temperature and the effective thermal conductivity relates to heat flux, again as sensible and as latent heat, to the forcing temperature gradient.

We envisage soil as consisting of the phases matrix (m), liquid water (w), possibly ice (i), and air (a). More advanced models would distinguish further soil phases, in particular quartz minerals, organic matter, or frozen water, all of which have quite different thermal properties (Table 8.1). Finally, we choose liquid water at 0°C as the reference state for thermal energy.

Effective Heat Capacity The soil heat capacity consists of two parts, (i) the inherent heat capacity of its constituents and (ii) the contribution from shifting the equilibrium between the phases in order to change the temperature. The inherent heat capacity of a composite medium may be

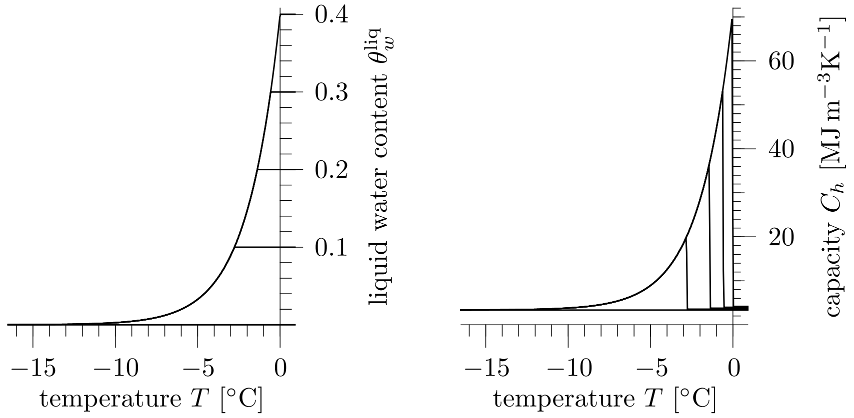


Figure 8.12. Liquid water content (left) and corresponding effective thermal capacity (right) for a soil with the idealized soil freezing characteristic (8.39) with $\phi = 0.4$ and $T_f = -2$°C for total water content $\theta_0 = \{0, 0.1, 0.2, 0.3, 0.4\}$. Notice how a lower value of θ_0 leads to a lower freezing temperature.

calculated as a weighted mean of its constituents, provided that contributions from interfaces may be neglected. Then, the effective volumetric heat capacity, in [J m^{-3}K^{-1}], of the soil becomes

$$C_h = \frac{\mathrm{d}H}{\mathrm{d}T} = \underbrace{\sum_k \theta_k \rho_k c_k}_{C_h^*} + \underbrace{\theta_a H_{\ell v} \frac{\mathrm{d}\rho_v}{\mathrm{d}T}}_{\text{evaporation}} + \underbrace{\rho_i H_{s\ell} \frac{\mathrm{d}\theta_w^{\mathrm{liq}}}{\mathrm{d}T}}_{\text{melting}} , \tag{8.38}$$

where θ_k is the volume fraction, ρ_k the mass density, and c_k the specific heat capacity [J kg^{-1}K^{-1}] of constituent $k \in \{m, w, i, a\}$. The second term is evaluated with (8.15) and the third term, which is only present for $T < 0$°C, with the soil freezing characteristic $\frac{\mathrm{d}\theta_w^{\mathrm{liq}}}{\mathrm{d}T}$, as it is shown in Figure 8.3.

As an illustration, consider the simplified model where the soil freezing characteristic is described by

$$\theta_w^{\mathrm{liq}}(T) = \phi \exp\left(-\frac{T}{T_f}\right) \tag{8.39}$$

with temperature T in °C and T_f a characteristic freezing temperature which depends on soil texture. Further assume that the actual liquid water in a soil at temperature $T < 0$°C is given by $\min(\theta_0, \theta_w^{\mathrm{liq}}(T))$, where θ_0 is the water content at $T = 0$°C (Figure 8.12). This corresponds to the understanding that liquid water preferably fills small pores as described by the soil water characteristic $\theta(\psi_m)$ introduced in Section 3.4 together with the Young-Laplace equation (3.2). Together, they yield the interfacial curvature for any given water content, with it the corresponding pore radius, and finally, with the chemical potential (8.16), the corresponding freezing point depression.

While these relations were introduced for non-frozen soils, their general form may be expected to also hold for frozen media. Indeed, *Bittelli et al.* [2003] proposed a method for determining the soil water characteristic from freezing experiments.

Notice that C_h is a macroscopically measurable quantity while the inherent heat capacity $\sum_i \theta_i \rho_i c_i$ is directly accessible. Further notice that a similar expression as (8.38) can be written for the apparent heat capacity of a single water phase, i.e., outside of a porous medium and with no air. There, $\frac{\mathrm{d}\rho_v}{\mathrm{d}T}$ and $\frac{\mathrm{d}\theta_w^{\mathrm{liq}}}{\mathrm{d}T}$ both become δ-function, however, and it is more useful to explicitly account for the corresponding phase transitions. In contrast, the transitions are smooth in multiphase porous media and the introduction of an effective heat capacity is appropriate.

Effective Thermal Conductivity In analogy to the effective heat capacity, the total heat flux in an unsaturated soil also consists of two conceptually different components, which result (i) from the inherent heat conduction by the composite material and (ii) from the transport of latent heat in the vapor phase and of sensible heat in the liquid phase. However, the situation is much more complicated than with the heat capacity because now the geometric arrangement of the various phases matters. Looking at Figure 8.11, we realize that the inherent thermal conductivity depends on three interrelated aspects: (i) the conductivity of the microscopic constituents, (ii) the network of high-flow regions and in particular the cross-sectional areas of constrictions, (iii) the microscopic temperature gradients which may strongly differ from the macroscopic gradient.

We take two steps to approach this difficult matter. In the first step, we express all contributions to the total heat flux in terms of $-\nabla T$, which entails the hypothesis of local thermodynamic equilibrium, and then interpret the respective coefficient as contribution to the total thermal conductivity. In the second step, we will consider some heuristic parameterizations.

Generic Formulations We first separate the total heat flux into three components, $\mathbf{j}_h = \mathbf{j}_h^* + \mathbf{j}_h^v + \mathbf{j}_h^w$, where $\mathbf{j}_h^*$ stems from the inherent heat conduction of the composite material, $\mathbf{j}_h^v$ from the flux of latent heat associated with vapor diffusion, and $\mathbf{j}_h^w$ from the convection of sensible heat in the liquid phase.

For calculating the macroscopic flux $\mathbf{j}_h^*$ due to the inherent heat conduction, we consider the microscopic phases as if they were solid and calculate the spatial average of the heat flux through some plane in much the same way as is illustrated in Figure 3.17 on page 54 for the water flux. With the microscopic flux given by $\langle \mathbf{j}_h^\mu = -K_h^\mu \nabla T^\mu$, where K_h^μ is the microscopic conductivity, this yields

$$\mathbf{j}_h^* = -\big\langle K_h^\mu \nabla T^\mu \big\rangle = -\underbrace{K_{h_0} \big\langle \kappa(\mathbf{x}) \boldsymbol{\vartheta}(\mathbf{x}, \theta) \big\rangle}_{=:K_h^*(\theta)} \nabla T \; , \tag{8.40}$$

where much the same argument was used as in the derivation of (4.47). In particular, $\kappa(\mathbf{x}) = K_h^\mu(\mathbf{x})/K_{h_0}$, where K_{h_0} is some normalizing constant, e.g., the geometric mean of K_h^μ, and $\boldsymbol{\vartheta}(\mathbf{x},\theta)$ relates the microscopic thermal gradient to the macroscopic one, $\nabla T^\mu = \boldsymbol{\vartheta}(\mathbf{x},\theta)\nabla T$, an equality which is warranted by the linearity of the process. Apparently, the two functions $\kappa(\mathbf{x})$ and $\boldsymbol{\vartheta}(\mathbf{x},\theta)$, the latter a tensorial function even for a microscopically isotropic medium, are of a complicated shape and vary on the pore-scale. Their product represents the correlation between the local conductivity and the local thermal gradient, with the latter tending to be smaller in highly conductive regions. The important aspect here is that they exist in the given form, in particular independent of T and ∇T, and thereby guarantee the macroscopic version of Fourier's law (8.19).

Next, we consider the contribution from the transport of latent heat in the vapor phase. As is apparent from (8.9) together with (8.15), a temperature gradient induces a gradient of the vapor density ρ_v, provided the matric potential ψ_m is the same everywhere. This in turn leads to a diffusive vapor flux $\mathbf{j}_v$ from warmer to cooler regions. In a first approximation, we model this flux as $\mathbf{j}_v = -D_{\text{eff}}^{\text{diff}}\nabla\rho_v$, where the effective diffusion coefficient $D_{\text{eff}}^{\text{diff}}$ is parameterized by one of the Millington-Quirk models (4.55), with θ replaced by the volumetric air content $\phi-\theta$. With $\rho_v(T)$, this may be written as

$$\mathbf{j}_v = -D_{\text{eff}}^{\text{diff}}\frac{\mathrm{d}\rho_v}{\mathrm{d}T}\nabla T\ . \tag{8.41}$$

Approximating water vapor as an ideal gas and recalling (8.15) for the reduction of vapor pressure above a curved interface, leads to

$$\rho_v = \frac{m_w}{\mathrm{R}T}p(T,\psi_m) = \frac{m_w}{\mathrm{R}T}\exp\Big(\frac{m_w\psi_m}{\rho_w\mathrm{R}T}\Big)p(T) \tag{8.42}$$

with molar mass m_w and with the vapor pressure curve $p(T)$ given by (8.9). This emphasizes the importance of the total derivative in (8.41), which accounts for $\rho_v(T,p(T))$.

As it turns out, (8.41) underestimates the true vapor flux for two reasons. The first one is that the Millington-Quirk parameterization is based on the assumption that diffusion occurs only in one phase, here in the vapor phase. However, water may also condense on one side of a blocking liquid region and evaporate at the other one (Figure 8.13). Transfer through the liquid is fast because it is not the molecules who travel but only the pressure wave. The second reason is that microscopic temperature gradients in the air phase tend to be considerably larger than the macroscopic gradient because the thermal conductivity of both water and solid matrix is much higher than that of air. These two effects are accounted for by a heuristic factor $F(\theta,T) > 1$, the so-called enhancement factor. With this, the flux of latent heat with the temperature-induced vapor flux becomes

$$\mathbf{j}_h^v = -H_{lv}D_{\text{eff}}^{\text{diff}}F(\theta,T)\frac{\mathrm{d}\rho_v}{\mathrm{d}T}\nabla T\ . \tag{8.43}$$

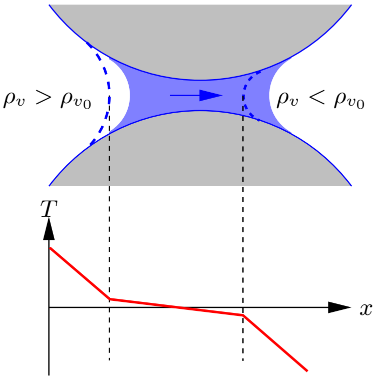


Figure 8.13. Vapor diffusion through a pendular ring between two sand grains. In static equilibrium (blue) the total curvature is the same everywhere and related to ρ_{v0}. With an imposed vapor flux, here from left to right, the interfacial curvatures change due to condensation and evaporation (blue dashed lines). This in turn creates the pressure gradient in the liquid that drives the flow there. The temperature profile through the center of the pendular ring is sketched in the lower graph.

Finally, we consider the temperature-induced convection of sensible heat in the liquid phase. As a preliminary, we recall the Young-Laplace equation (3.2), $\psi_m \propto \sigma/r$, where r is the interfacial radius, $-r$ corresponds to the radius of the largest water-filled pores, and the surface tension σ decreases with increasing temperature. Then imagine an unsaturated soil column which initially is isothermal, with the water phase in thermodynamic equilibrium. Hence, $\psi_w = \psi_g + \psi_m$ is equal throughout the column and at every point, the water phase occupies the region of the pore space that corresponds to radii smaller than $-r$. Let then the temperature increase in some region. As a consequence, σ decreases and the largest pores cannot bind the water anymore.The excess water will then redistribute and move towards the colder regions. Once that redistribution is completed, the water flow will cease again, except if the liquid water is removed at the cold end, for instance by freezing [e.g., *Cary and Mayland* 1972].

Combining (8.40) and (8.43), and interpreting the factor relating $-\nabla T$ to $\mathbf{j}_h$ as effective thermal conductivity, finally leads to

$$K_h(\theta, T) = K_h^*(\theta) + H_{lv} D_{\text{eff}}^{\text{diff}} F(\theta, T) \frac{\mathrm{d}\rho_v}{\mathrm{d}T} \ . \tag{8.44}$$

The generic formulations discussed so far are important for understanding the underlying processes. They are not very useful operationally, however, and demand appropriate parameterizations. We will in the following explore two conceptually different approaches, *heuristic parameterizations* and *process-based parameterizations*. In their respective pure form, they aggregate comprehensive datasets into a few numbers without recourse to a deeper understanding of the associated processes or, conversely, they predict observed quantities based on fundamental processes and a few natural constants. Various such descriptions have been compiled, for instance by *Farouki* [1981] with a focus on cold regions and by *Peters-Lidard et al.* [1998] with a specific interest in soil-vegetation-atmosphere transfer.

Heuristic Parameterizations The effective thermal conductivity $K_h(\Theta)$ as a function of water saturation Θ is interpolated between its dry and its saturated value as

$$K_h(\Theta) = \kappa_e(\Theta)[K_h^{\text{sat}} - K_h^{\text{dry}}] + K_h^{\text{dry}} \ , \tag{8.45}$$

where $\kappa_e(\Theta)$ is the dimensionless thermal conductivity function, also called the *Kersten number*. *Johansen* [1977] proposed the parameterization

$$\kappa_e = \max(0, 1 + \alpha \log_{10}(\Theta)) \tag{8.46}$$

where α depends on soil texture, with $\alpha = 0.7$ for coarse- and $\alpha = 1$ for fine-textured soils. For frozen soil, $\kappa_e = \Theta$. The conductivity in the dry state is described by

$$K_h^{\text{dry}} = \frac{0.135\rho_b + 64.7}{2700 - 0.947\rho_b} \ , \tag{8.47}$$

where ρ_b is the soil bulk density. This expresses the fact that the actual composition of the soil matrix has no strong effect on K_h^{dry} which is dominated by the huge contrast in conductivities between air and matrix (Table 8.1). This is different in the saturated state, where the conductivities of the phases are comparable. Consequently, K_h^{sat} is expressed in terms of the conductivities of the pure materials, K_{h_s} and K_{h_w}, and of their respective volume fraction as

$$K_h^{\text{sat}} = K_{h_s}^{1-\phi} K_{h_w}^{\phi} \ , \tag{8.48}$$

with porosity ϕ and $K_{h_w} = 0.57\ \mathrm{W\,m^{-1}K^{-1}}$. The conductivity of the soil matrix is sometimes further resolved into $K_{h_s} = K_{h_q}^{q} K_{h_o}^{1-q}$, where q is the quartz volume fraction within the matrix, with $K_{h_q} = 7.7\ \mathrm{W\,m^{-1}K^{-1}}$, and correspondingly $1 - q$ the fraction of other minerals, with $K_{h_o} = 3.0\ \mathrm{W\,m^{-1}K^{-1}}$ for $q < 0.2$ and $K_{h_o} = 2.0\ \mathrm{W\,m^{-1}K^{-1}}$ otherwise.

Apparently, this parameterization requires a minimal data input, in the strict sense just the porosity ϕ and the quartz fraction q, and it could easily be extended to include other phases along the same lines. In many applications, the parameter α and possibly also K_{h_o} would be fitted to some data. Despite its simplicity, *Farouki* [1981] finds that this method yields equally satisfactory results as the much more complicated process-based de Vries parameterization described below.

Johansen's parameterization (8.46) is not smooth, which may lead to difficulties in some models. This is remedied by the smooth representation

$$\kappa_e = \exp(\alpha[1 - \Theta^{\alpha - 1.33}]) \tag{8.49}$$

proposed by *Lu et al.* [2007]. For the texture-dependent parameter α they report $\alpha = 0.96$ for coarse-textured soils and $\alpha = 0.27$ for fine-textured soils. They furthermore replace (8.47) by

$$K_h^{\text{dry}} = a\phi + b \ , \tag{8.50}$$

where $a = -0.56$ and $b = 0.51$ were fitted to a set of soils. Comparing predicted and measured thermal conductivities for a number of soil samples, *Lu et al.* [2007] found their parameterization to perform better than *Johansen*'s with the root mean square error of K_h in the range of 0.04…0.14 $\mathrm{W\,m^{-1}K^{-1}}$, down from 0.07…0.20 $\mathrm{W\,m^{-1}K^{-1}}$ for *Johansen*'s formulation.

We comment that a comparison with the de Vries parameterization in Figure 8.14 below demonstrates that (8.49) is indeed a rather good interpolator between the extremes K_h^{dry} and K_h^{sat}, but that the values of α will in general have to be fitted. Values required for Figure 8.14 are 0.08 for the sandy material and 0.18 for the silty material. Despite the very limited predictive power of this parameterization, its simplicity and flexibility makes it a prime choice for inversion and data assimilation procedures, where optimal parameters are estimated from measured temperatures.

Process-Based Parameterizations As a preliminary notice that, so far, the processes that comprise heat transfer in wet soils proved to be too complicated to forsake heuristics completely, even on the "process-based" branch. Indeed, if just an operational formulation is sought, the parameterization (8.49) of *Lu et al.* [2007] with parameter set $\{K_h^{\mathrm{dry}}, K_h^{\mathrm{sat}}, \alpha\}$, all depending weakly on temperature and all to be determined empirically, currently appears as the best bet for moderate temperatures, say for $T < 40°\mathrm{C}$. Under that premise, the following elaborations may be safely skipped even though they do provide deeper insight into the physics of the complicated matter.

Current models start from the mean-field representation given by *de Vries* [1963, 1975]. It is based on the fiction that liquid water is the continuous phase, with the other components being isolated small inclusions. Such models are typically formulated for an arbitrary number of phases. We will stay in the following with the simple three-phase model that consists of matrix, liquid water, and air. The effective thermal conductivity of the composite medium may then be written as

$$K_h(\theta, T) = \frac{\sum_i \theta_i w_i K_{h_i}}{\sum_i \theta_i w_i} \ , \tag{8.51}$$

with phases $i \in \{m, w, a\}$ with volume fractions θ_i and thermal conductivities $K_{h_i}(T)$. The coefficients $w_i(\theta, T)$ are weight functions which represent the average temperature gradient in phase i with respect to the average temperature gradient in the embedding medium. Calculating these coefficients is the main goal in a mean-field theory. In a typical approach, small and widely separated inclusions, usually with ellipsoidal or even spherical shapes, are assumed to be embedded in a uniform medium. *Campbell et al.* [1994] pushed this concept and envisaged some mean embedding medium with thermal conductivity

$$K_{h_f}(\theta, T) = K_{h_a} + \alpha(\theta, T)[K_{h_w} - K_{h_a}] \ , \tag{8.52}$$

where

$$\alpha(\theta, T) = \left[1 + [\theta/\theta_0]^{-q(T)}\right]^{-1} \ , \quad q(T) = q_0[T/T_0]^2 \tag{8.53}$$

is an empirical function which describes the transition between air and water as embedding phase. This transition occurs at the volumetric water content θ_0 with the width of the transition zone determined by parameter q_0 and modified by temperature, with $T_0 = 273.15$ K chosen as reference temperature. For $\theta \ll \theta_0$, water is only present as adsorbed films and does not affect the overall thermal conductivity K_h significantly. For $\theta \approx \theta_0$, pendular rings are abundant and K_h depends strongly on θ. Finally, for $\theta \gg \theta_0$, K_h still increases with θ since less conductive air is replaced by more conductive water. However, the geometry of the network of high-flow regions is no more affected strongly, such that K_h increases less rapidly. Assuming ellipsoidal inclusions, *Campbell et al.* [1994] use for the empirical weight functions

$$w_i = \frac{1}{3}\left[\frac{2}{1+\left[\frac{K_{h_i}}{K_{h_f}}-1\right]g_a} + \frac{1}{1+\left[\frac{K_{h_i}}{K_{h_f}}-1\right][1-2g_a]}\right] , \tag{8.54}$$

again for phases $i \in \{m, w, a\}$ and with the shape parameter g_a determining the form of the inclusions. The dependence of K_{h_f} on θ and T, and of K_{h_i} on T, has been suppressed for clarity. The former, $K_{h_f}(\theta, T)$, has been formulated already in (8.52), now we consider $K_{h_i}(T)$. For the temperatures of interest here, K_{h_m} is constant. The conductivities of water, K_{h_w}, and of dry air, $K_{h_{da}}$, may be approximated as

$$\begin{aligned} K_{h_w}(T) &= -0.554 + 2.24 \cdot 10^{-3}[T - T_0] - 9.87 \cdot 10^{-6}[T - T_0]^2 , \\ K_{h_{da}}(T) &= 0.024 + 7.73 \cdot 10^{-5}[T - T_0] - 2.6 \cdot 10^{-8}[T - T_0]^2 , \end{aligned} \tag{8.55}$$

where $T_0 = 273.15$ K. The situation is much more complicated for wet air, where the diffusion of water vapor leads to the flux of latent heat. A number of variously elaborated schemes exist for parameterizing this contribution. *Campbell et al.* [1994] follow the original work of *de Vries* [1963] most closely and extend it to account for temperature-dependence, in particular of the vapor flux.

We take a slightly different approach and start from the formulation (8.43). Looking at the individual terms, we first insert (8.42), together with parameterization (8.9) of the vapor pressure curve $p(T)$. Next, we focus on the diffusion coefficient of water vapor in air and are going to use the second Millington-Quirk model in its original formulation for the diffusion of an insoluble gas and write it as $D_m = D_{m_0}[\phi - \theta]^2/\phi^{2/3}$. As a passing comment when comparing this with the formulation used in (4.55) on page 95, notice that (i) the volume fraction, of the phase where diffusion is considered, is now $\phi - \theta$ and (ii) the additional factor θ, which was split off in (4.55), has now to be included. Finally, about the diffusion coefficient, we use the approximation given by *Massman* [1998] to account for the temperature dependence. With this

$$D_{\text{eff}}^{\text{diff}}(\theta, T) = D_m \frac{[\phi - \theta]^2}{\phi^{2/3}} \left[\frac{T}{T_0}\right]^{1.81} \tag{8.56}$$

with $D_m = 21.78 \cdot 10^{-6}$ m^{2}s^{-1} and $T_0 = 273.15$ K. Notice that invoking the Millington-Quirk model first appears daring since it is only applicable for insoluble gases. However, here we use it in exactly this capacity, as a representation of transport in the vapor phase, which may have to be augmented with an enhancement factor. Next, we approximate the latent heat of vaporization as

$$H_{lv}(T) = \big[2.503 - 2.44 \cdot 10^{-3}[T - 273.15]\big] \ \mathrm{MJ\,kg}^{-1} \ . \tag{8.57}$$

The last term to consider is the enhancement factor $F(\theta, T)$, which was introduced heuristically in (8.43) to account for observed water fluxes that were much larger than predicted. This is a recurrent and still a largely unresolved issue in all parameterizations. While some argue that once capillary flow is represented correctly, this factor is needed no more [*Shokri et al.* 2009], or is at least greatly reduced [*Ho and Webb* 1996], others specifically set out to estimate it. For instance, *Lu et al.* [2011], based on earlier work of *Cass et al.* [1984], find F to monotonically increase from 1 to about 15 with increasing values of θ and propose

$$F(\Theta) = \alpha_1 + \alpha_2\Theta - [\alpha_1 - \alpha_4]\exp\big(-[\alpha_3\Theta]^{\alpha_5}\big) \tag{8.58}$$

with $\Theta = \theta/\phi$, constant parameters $\{\alpha_1, \ldots, \alpha_5\}$, and independent of temperature as a useful description. We take a pragmatic approach and include this factor, most importantly because it is very probable that some sort of enhancement is required with the above Millington-Quirk formulation.

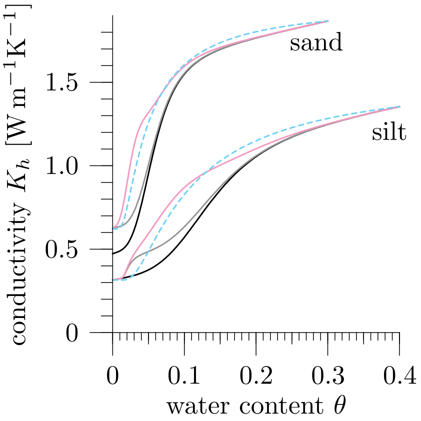
Summarizing, the thermal conductivity of the wet air phase, which is to be used in (8.52) and (8.54), becomes

$$\begin{aligned} K_{h_a}(\theta, T) \ &= \ K_{h_{da}}(T) \qquad\qquad (8.59)\\ &+ \frac{H_{lv}(T) m_w}{R} D_m \frac{[\phi - \theta]^2}{\phi^{2/3}} \left[\frac{T}{T_0}\right]^{1.75} F(\theta, T) \frac{\partial \log\big(p(T)\big)}{\partial T} \ . \end{aligned}$$

With this, we are in the position to assemble the *modified de Vries* parameterization for the effective thermal conductivity K_h: (8.59) and (8.55) describe the thermal conductivities of the individual phase and (8.54) gives the relative weights to add them up with (8.51).

A graphical representation of the effect of the heuristic enhancement factor is given in Figure 8.14. We first look at the curves for silt and notice that vapor diffusion alone, the gray curves, has no effect at the very dry end, for $\theta < 0.02$. Here, the reduction of the vapor pressure due to the correspondingly very low matric potential ψ_m is so large that the amount of water in the vapor phase is negligible. As θ increases, this reduction decreases rapidly and leads to corresponding growth of K_h. For still larger values of θ, this effect decreases again and has all but vanished for $\theta > 0.15$. This is a consequence of the Millington-Quirk factor which is proportional $[\phi - \theta]^2$. The effect of the enhancement factor $F(\theta)$ is strongest for intermediate values

Figure 8.14.
Effective thermal conductivity in modified de Vries parameterization for a sandy and a silty material without vapor diffusion (black), with pure vapor diffusion given by (8.59) with $F(\theta, t) = 1$ (gray) and with the heuristic enhancement factor $F(\theta)$ given by (8.58) (light magenta). All curves are for 20°C. The heuristic parameterization (8.45) with normalized conductivity function (8.49) and $K_h(0)$ and $K_h(\phi)$ taken from the de Vries approximates the latter rather well (dashed cyan). This requires to fit the parameter α, however.



of θ and thus represents the expectation that it extends the effect of vapor transport towards larger values of θ but that it also fades away as the volume fraction of the air phase decreases. The overall effect of the transport of latent heat through vapor diffusion is to shift the $K_h(\theta)$ towards lower values of θ but without affecting the extreme values of a completely dry or completely saturated medium.

For the sand, we observe qualitatively the same features with the difference that the effect of vapor diffusion sets in at very low values of θ, too low indeed to be represented in the graphics. The reason for this is that for the coarse-textured medium, the soil water characteristic at the dry end is so steep that already a minute increase of θ brings ψ_m to values that are no more limiting for the vapor pressure curve.

Next, we focus on the temperature dependence of K_h (Figure 8.15) and notice a strong dependence for intermediary values, except for the coarse-textured sand where vapor diffusion is already strong for dry states. Indeed, $K_h(\theta)$, which is a monotonic function at low temperatures, develops a maximum at some intermediate value of θ. This is also expected from the physical concept which has the mass of water vapor increase rapidly with temperature and transport will be most efficient as long as a large fraction of the pore space is air-filled, hence the local temperature gradients are large. Corresponding data are for instance reported by *Campbell et al.* [1994].

Thermal Diffusivity The thermal diffusivity D_h was introduced with (8.21) as the ratio between conductivity K_h and capacity C_h. In the strict sense, D_h is only useful for uniform media with stationary thermal properties, only then can K_h and C_h be extracted from the respective differential operators. Still, we may use $D_h := K_h/C_h$ for a rough characterization of the thermal dynamics (Figure 8.15).

Comments Apparently, the processes that lead to the macroscopically observable thermal material properties are very complicated at the microscopic

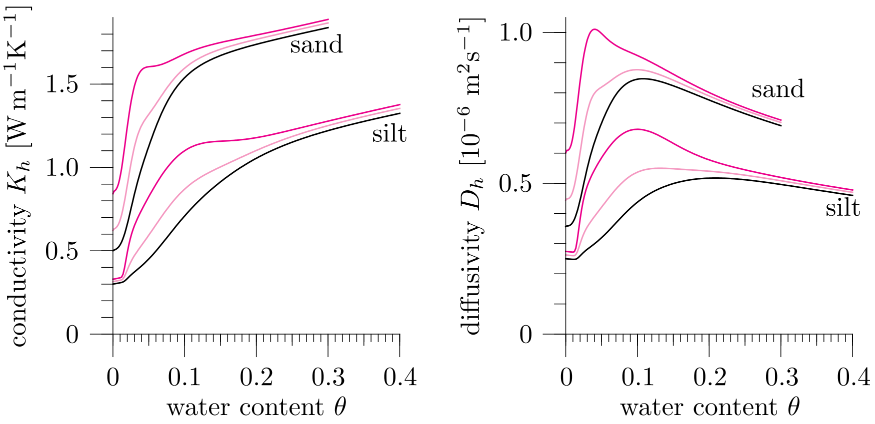


Figure 8.15. Effective thermal conductivity (left) and thermal diffusivity (right) for the materials already used for Figure 8.14, now for temperatures of 0°C (black), 20°C (light magenta), and 40°C (dark magenta). For all curves, the parameterization (8.59) with enhancement factor (8.58) was used.

scale and highly dependent on pore-scale details. This is in particular the case for the thermal conductivity. While we appear to understand the physics of these microscopic processes sufficiently well, this is not true for their quantitative description in a typical natural pore space. As a consequence, the quantitative upscaling, the prediction of macroscopic properties remains a rather heuristic undertaking. This general situation is reminiscent of the discussion of hydraulic material properties in Section 3.4. The situation with thermal properties for a wide range of soils and naturally occurring conditions is much more benign than for hydraulic properties since the range of values is rather small, seldom spanning a factor of more than 5. Compare this to the many orders of magnitude for hydraulic conductivity, for instance. In addition, the underlying dynamics is a diffusion process which by its nature is not very sensitive to the parameters. Still, the dependence of K_h and C_h on temperature turns the process nonlinear, if only weakly.

From an operational perspective, i.e., when heat transfer is not so much of interest *per se* but as part of a larger description of some environmental system, the current best approach appears to be to (i) obtain time series of temperature at relevant locations, (ii) choose a simplified parameterization, probably (8.45) together with (8.46) or (8.49), and (iii) adjust the parameters such that an optimal agreement between data and model is obtained. One mandatory prerequisite for this is that the chosen parameterization has the correct structural flexibility with a minimal set of parameters. In addition, these parameters must be mutually independent such that different features of a curve, $K_h(\theta, T)$ for instance, are affected by only one parameter.

8.3.2
Effective Dynamics

Inserting the effective material properties $C_h(T,\theta)$ (8.38) and $K_h(T,\theta)$ (8.44) into the heat conduction equation (8.20) yields

$$\partial_t[C_h T] - \nabla\cdot[K_h \nabla T] = r_h \ , \tag{8.60}$$

where C_h corresponds to ρc_h, the possible tensorial character of K_h, which would originate in an anisotropic microscopic structure, has been neglected, and the source r_h of thermal energy has been added. Concerning the last term, notice that the thermal energy associated with all equilibrium phase transitions is already contained in the effective parameters, hence must be added again through r_h. However, this term is useful to account for additional thermal input, for instance through infiltrating warm rainwater, or for non-equilibrium phase transitions.

The material properties depend on T and θ, hence (8.60) in general demand a numerical solution. For many situations, reasonable approximations are feasible, however. In order to explore this, we focus on heat transport only, neglecting the dependence on θ, and expand (8.60) as

$$C_h\Big[1+\underbrace{\frac{T}{C_h}\frac{\mathrm{d}C_h}{\mathrm{d}T}}_{\alpha}\Big]\partial_t T - K_h\nabla^2 T - \underbrace{\frac{\mathrm{d}K_h}{\mathrm{d}T}\nabla T\cdot\nabla T}_{\beta} = r_h \ . \tag{8.61}$$

Apparently, the uniform heat conduction equation (8.21) may be used to approximate (8.60) to the extent to which the terms α and β are negligible. First consider α and notice that it will only be large phase transitions of the entire system, i.e., for water under standard condition near 0°C (Figure 8.12) and near 100°C. With (8.38) and focussing at low temperatures, $\alpha \approx \rho_i H_{sl}\frac{\mathrm{d}^2\theta_w^{\mathrm{liq}}(T)}{\mathrm{d}T^2}$, where the approximation is very good due to the low temperature-dependence of the individual materials. Inserting the values used for Figure 8.12 shows that $\alpha < 0.2$ for $T < -3$°C and that the situation would be much more favorable for the measured soil freezing characteristic of Figure 8.3.

Next, consider β, for which a general estimation is more difficult. We have to compare $K_h\nabla^2 T$ with β and first notice that $K_h = \mathcal{O}(1)$ and, from Figure 8.15, that for low temperatures $\frac{\mathrm{d}K_h}{\mathrm{d}T} = \mathcal{O}(10^{-2})$, often very much smaller, in particular for high water saturations. Further, $|\nabla T|^2 = \mathcal{O}(|j_h|^2)$ since $K_h = \mathcal{O}(1)$, and of course $K_h\nabla^2 T = -\nabla\cdot\mathbf{j}_h$. Summarizing, and considering the vertical coordinate only, we look for conditions for which $|\partial_z j_h| \gg 10^{-2}|j_h|^2$. For soils, this can be satisfied at a certain distance from the surface, j_h decreases roughly exponentially with depth, and with at high water saturations.

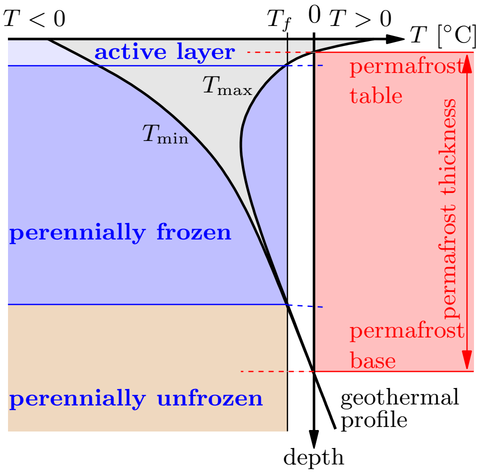


Figure 8.16. Notations for permafrost soils according to *van Everdingen* [1998]. Permafrost is defined with respect to the temperature $T = 0$°C. However, due to the freezing point depression in porous media, water is only frozen for $T \leq T_f < 0$°C. Hence, the definitions of frozen ground and of permafrost do not coincide. The gray area represents the range of temperatures in the profile.

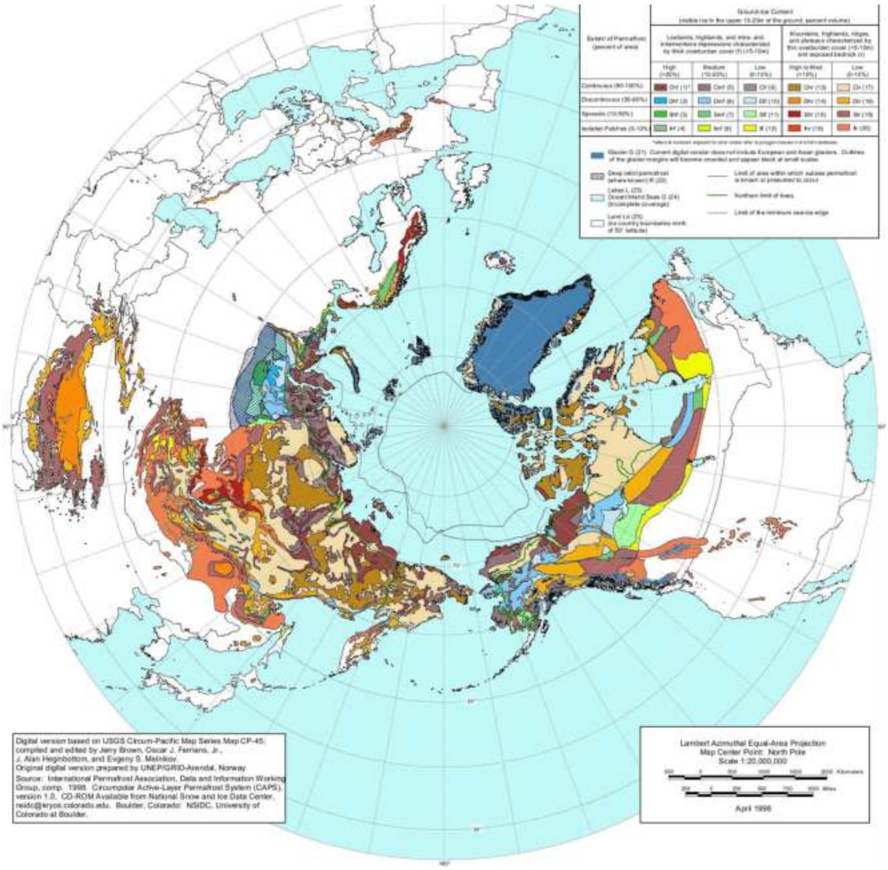
8.4
Permafrost Soil

By definition, permafrost encompasses terrestrial and submarine soils and sediments that are not permanently covered by ice and whose mean annual ground temperature is at or below 0°C for at least two consecutive years (Figure 8.16). They are often, but not always, associated with the presence of ground ice.

The top layer of a permafrost site is the seasonally thawing and freezing *active layer* which extends from a few centimeters depths to several meters, depending on the site. This is the layer where most of the action takes place and where most of the defining features of permafrost landscapes originate. Important aspects include (i) the coupling between mechanical, thermal, and hydraulic processes which often leads to self-organized ground patterns [*Kessler and Werner* 2003], (ii) a ground that is turned into a non-trafficable mass by the seasonal thawing and poses a severe and expensive challenge to engineering structures [*Nelson et al.* 2001], and (iii) a closely coupled thermal and hydraulic dynamics, a consequence of the soil freezing characteristic, that leads to the stabilization of various characteristic states (Figure 8.20 on page 280 below).

The active layer is typically followed by the perennially frozen layer that may be as deep as 1 km. It typically contains layers of solid ice with thicknesses up to several meters. Deeper down extends the perennially unfrozen layer where the geothermal heat flow prevents temperatures to fall below the freezing point.

Some 24% of the non-glaciated land surface on the Northern Hemisphere belong to the permafrost region. About 60% of this area are less than 500 m above sea level (asl), and at correspondingly high latitudes, while some 10% belong to the high-altitude, low-latitude permafrost. The latter is



Digital version based on USGS Circum-Pacific Map Series Map CP-45; compiled and edited by Jerry Brown, Oscar J. Ferrians, Jr., J. Alan Heginbottom, and Evgeny S. Melnikov. Original digital version prepared by UNEP/GRID-Arendal, Norway. Source: International Permafrost Association, Data and Information Working Group, comp. 1998. Circumpolar Active-Layer Permafrost System (CAPS), version 1.0. CD-ROM Available from National Snow and Ice Data Center, nsidc@kryos.colorado.edu. Boulder, Colorado: NSIDC, University of Colorado at Boulder.

Figure 8.17. Distribution of permafrost and ground ice north of 30°N. [Map prepared by *International Permafrost Association* and *National Snow and Ice Data Center*. Map and data available from `nsidc.org/fgdc/maps`.]

mostly located on the Qinghai-Xizang (Tibet) plateau, but smaller patches are found in all major mountain ranges like the Alps and the Rocky Mountains (Figure 8.17). The Southern Hemisphere contains only small permafrost regions because land masses are either covered by ice (Antarctica) or not at sufficiently high latitudes.

In cold regions, the thermal and hydraulic dynamics of permafrost soils controls the fluxes of energy, water, and carbon between land surfaces and the atmosphere. This is of particular interest on the background of the current climate warming which is already very pronounced in the Arctic [*Johannessen et al.* 2004] and is anticipated to increase dramatically in the coming decades. This warming leads to a decay of the permafrost, thereby exposing large quantities of organic carbon in thawing soils, particularly in Siberia and to a lesser degree in Canada. Decomposition of the organic carbon is expected to contribute to a further increase of atmospheric CO_2 concentrations, fueling one of the positive feedback mechanisms of the Earth's climate.

8.4.1
High-Latitude Permafrost

Solar radiation at high latitudes is characterized by a strong annual cycle, the polar day and night, that is of comparable magnitude as the diurnal cycle, sometimes even stronger. Net radiation – the difference between incoming and outgoing radiation – is always negative during the polar night. During the polar day, it is generally positive although the low angle of incidence makes it strongly dependent on surface exposition.

As a typical example, we study the thermal and hydraulic dynamics at the Bayelva site on Svalbard (Spitzbergen), an island in the North Atlantic at the latitude of Northern Greenland. Despite the high latitude, the climate is rather warm due to the Gulf Stream, and even during the winter months sporadic rain events occur.

The Bayelva Site The site is located at 78°55'N 11°50'E on a small hill some 30 m asl and 2 km downstream of a glacier [*Roth and Boike* 2001]. Continuous permafrost reaches some 100 m deep and the active layer is about 1.4 m thick. The surface is covered by *mud boils*, also called *unsorted circles* (Figure 8.18). In order to illuminate the function of this site we focus on a single, fully developed mud boil and envisage a simple model for its working. We start with the frozen state at the end of the polar night. As the net radiation increases with the lengthening days, snow starts to melt and first exposes the top of the mound. Water from the thawing snow will flow downhill on the still frozen surface and accumulate in the ring of organic matter that girdles the mound. Here it will infiltrate eventually. The thawing front may be expected to start at the mound's top and penetrate quickly since the water content tends to be minimal there. In contrast, the front will advance more slowly beneath the organic matter because (i) the snow melts later, (ii) organic matter is a good thermal isolator, and (iii) the infiltrating snowmelt increases the water content of the soil and thus its thermal capacity. The timing and the different velocities of the thawing front beneath the center of the mud boil and beneath its fringe lead to its concave shape, which in turn causes the infiltrating melt water to accumulate under the mound's center. While this increases the thermal capacity somewhat, the heat input modulation at the soil surface, mainly through the thermal isolation by the plants and the organic matter, dominates the heat fluxes also at greater depths. Hence the depression persists for some time while the thawing front progresses. Before it becomes leveled out, however, the freeze-back of the following winter sets in. Since soil with a higher water content expands more, the mound is lifted more than the fringe region. This is the engine behind the flow of solid matter. Surface erosion counteracts this

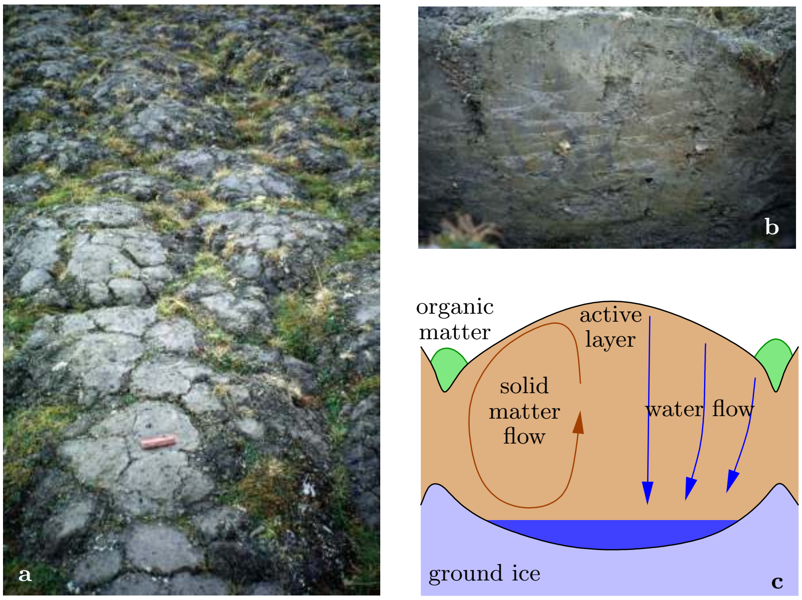


Figure 8.18. Mud boils, also called unsorted circles, at Bayelva site (**a**), cross-section through one of them (**b**), and conceptual sketch of its dynamics (**c**) where the average flow of solid matter is indicated in the left part and an episodic seepage of water in the right part. In reality, the flow of water is much more complicated frequently changing direction and also switching between dominating liquid and vapor flow. The diameter of a typical mud boil at this site is some 1.5 m.

process. The long-term balance between the two determines the height of the mounds. Furthermore, erosion is the engine that drives lateral transport away from the mounds. Some aspects of this simple model have indeed been verified by direct observation. These are in particular the concave shape of the thawing front by the time it reaches its maximum depth and the high water content above the ground ice.

We notice that the flows of energy and water on the one hand and of solid matter on the other occur at greatly different time scales. The two processes may thus be separated and water flow may be understood, at least qualitatively, in the context of a rigid porous medium.

Observations Figure 8.19 shows temperatures and liquid water contents measured in a profile at the center of a mud boil for a period of about two years. In addition, some meteorological variables and the snow height at the site are plotted.

Net Radiation As expected for a high-latitude site, net radiation is negative during the winter months, here roughly between September and April, and the surface cools radiatively. During the summer months, June to August, net radiation is positive, sometimes exceeding 200 $\mathrm{W\,m^{-2}}$. The annual average

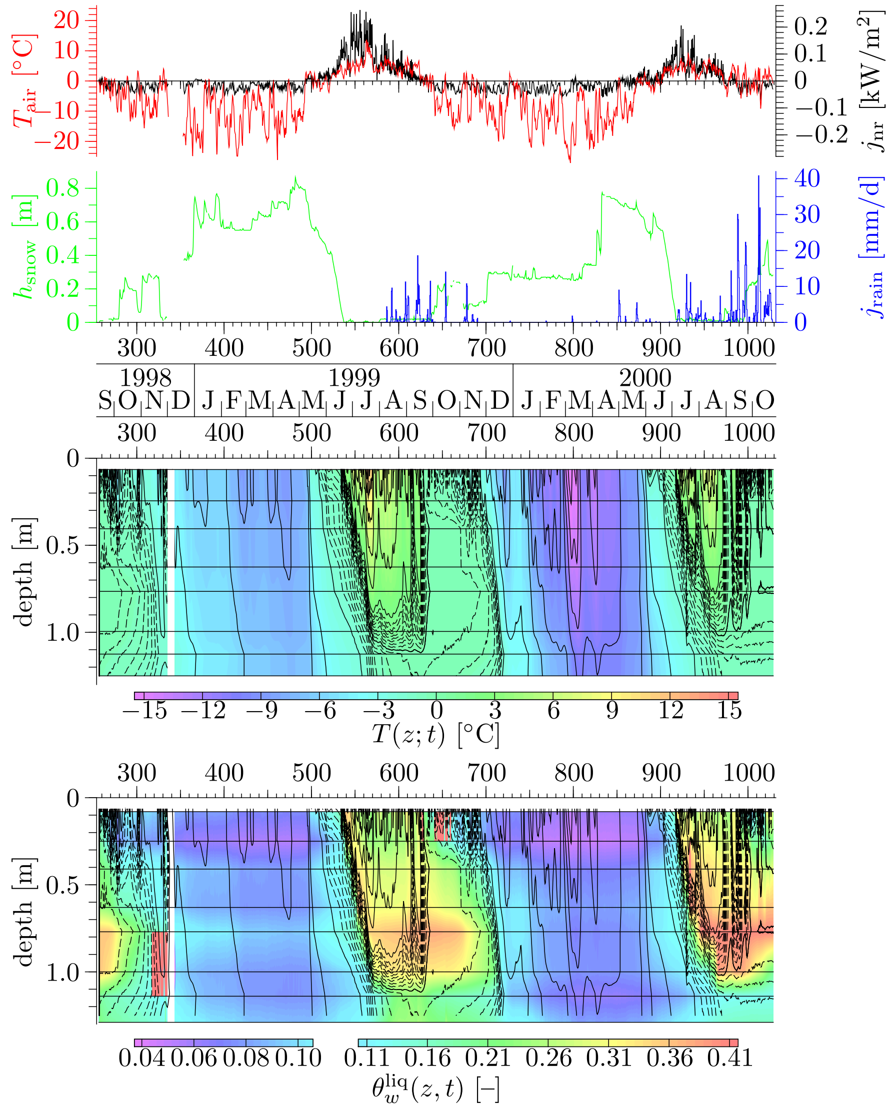


Figure 8.19. Thermal and hydraulic dynamics observed at the Bayelva site, averaged over 6 h periods. The line graphs show air temperature T_{air}, net radiative flux j_{nr}, snow height h_{snow}, and rainfall j_{rain}. The upper contour plot shows soil temperatures measured in a vertical profile through the center of a mud boil. Contour lines are spaced at 2°C (solid) and 0.2°C (dashed), respectively. The lower contour plot represents measured liquid water contents. Contour lines belong to temperature, for easier comparison, and the spacing of the colors is broken to increase the resolution for low values. White areas indicate missing data, horizontal black lines locations of sensors.

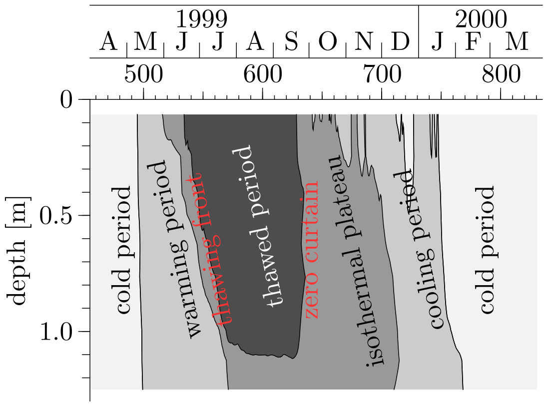


Figure 8.20. Typical thermal periods at a high-latitude permafrost site illustrated for an excerpt of Figure 8.19.

was +13.4 $\mathrm{W\,m^{-2}}$ for the first year, ending in September. During the second year, average net radiation was negative, −2.4 $\mathrm{W\,m^{-2}}$.

Air Temperature During the winter months, air temperatures are quite variable at this site. They fall well below −20°C when arctic air masses prevail but also raise above 0°C when warm air breaks in from the south. Mean annual air temperatures were −4.4°C and −5.6°C for the first and second year, respectively.

Ground Temperature During the winter, ground temperatures are influenced by the air temperature and by the height of the isolating snow. This is evident from comparing the month of March in 1999 and 2000. While air temperatures were comparable, snow heights differed by a factor of about 2 leading to ground temperatures around −6°C and −15°C, respectively.

An annual cycle consists of different characteristic periods that are indicated in Figure 8.20. We start with the *cold period* where temperatures are well below 0°C. Most of the soil water is frozen and the entire ground may be considered as a solid medium. With the sun rising in late spring and the onset of snow melt, ground temperatures rise quickly and "move up the soil freezing characteristic" (Figure 8.3 on page 248). This *warming period* is terminated by the *thawing front*, a macroscopic phase boundary that separates partly frozen ground from the completely unfrozen soil during the *thawed period*. The propagation of the thawing front consumes practically all the energy input from the positive net radiation. This energy is predominantly transported by heat conduction, an assumption that will be substantiated in the analysis below (Figure 8.21). A smaller fraction may be transported by heat convection, either as sensible heat with infiltrating warm rain and surface water or as latent heat with diffusing water vapor that condenses at the cold end. As net radiation decreases again in late summer, the thawed soil cools down to 0°C starting from the surface. Upon further cooling the reverse phase transition sets in, from liquid to solid. It releases large quantities of latent heat and thereby opposes the cooling. As a result, the entire thawed zone equilibrates towards a temperature of 0°C where freezing sets in. This is indicated by the practically vertical isotherms in Figure 8.19 and is often referred to as the closing of the *zero-curtain*. With the active layer turning isothermal, heat conduction gets suppressed and the large quantities of latent heat that need to be removed to allow further cooling can only be transported by rather inefficient processes like diffusing water vapor. This leads to the *isothermal plateau* where soil temperatures stay near 0°C for an extended time during which the domain "descends down the soil freezing characteristic". At the Bayelva site this is for up to two months. The analogous period during the warming phase is discussed in Exercise 8.3. Obviously, the lifetime of this plateau depends on the amount of liquid water and on the depth of the original thawed zone. It is eroded from above and from below, where temperatures in the air and in the underlying perennial ice are well below the freezing point. Eventually, the isothermal plateau disappears, giving way to the *cooling period* where regular conduction removes the remaining surplus heat from the active layer. It is transferred through the snow layer, which is a major modulating factor, to the atmosphere.

Liquid Water Content The measurements were done with time-domain reflectometry (TDR) where an electromagnetic pulse travels down a wave-guide in the soil and back again. From the travel time of this pulse, the speed of light is calculated which in turn yields the composite dielectric number ε_c of the soil. Its value is strongly influenced by the volumetric content θ_w^{liq} of liquid water, for which $\varepsilon_w \approx 83$. In contrast, ice has $\varepsilon_i \approx 3$ and for other soil constituents $\varepsilon < 6$. *Roth et al.* [1990] proposed a semi-empirical model for calculating θ_w^{liq} from a measurement of ε_c.

For temperatures below about -1°C, the liquid water content is strongly linked with temperature through the soil freezing characteristic. An example obtained from the Bayelva site is shown in Figure 8.3 on page 248. Changes of θ_w^{liq} in this cold range primarily stem from phase changes and, presumably to a lesser degree, from water movement through the vapor phase. Movement in the liquid phase may be expected to be small since it is restricted to ever smaller pores as temperatures decreases.

With temperatures rising beyond about -1°C, θ_w^{liq} and T become decoupled and movement in the liquid phase begins to dominate the dynamics of the water phase, at least in wet locations like Bayelva. The thawing front releases large amounts of liquid water that accumulates above the rather impermeable front. Since the frozen ground is impermeable to liquid water, θ_w^{liq} is highest at the bottom of the active layer. Consequently, the isothermal plateau lasts for a longer time at this depth than at others.

Analysis Simulation of the dynamics of a permafrost site is a daunting task since a number of highly nonlinear and coupled processes are involved. *Water* moves in the liquid and in the vapor phase primarily driven at the soil surface by rainfall, snowmelt, and evaporation. Secondary drivers within the soil are the potential gradients that result from freezing and the water supplied from thawing. The changing water content strongly modifies the hydraulic properties and, to a lesser degree, also the thermal ones. *Thermal energy* is transported as sensible and latent heat with the water phase and in addition moves by conduction. Temperature directly affects the hydraulic properties through surface tension and viscosity and indirectly through vapor movement. Whenever *ice* is present in the ground, there develops a strong coupling between the thermal and the hydraulic dynamics through the soil freezing characteristic. Ice also modifies the soil's transport properties, for water in the liquid and vapor phase by blocking parts of the pore space and for heat through the latent heat of phase change. In addition, ice strongly affects the mechanical properties of many soils. Finally, *solutes* transported in the liquid water are expelled upon freezing or evaporation and they influence the freezing point. While all these processes may in principle be incorporated into a numerical solver, the faithful estimation of the required large number of effective material properties is a problem that has yet to be addressed.

Instead of attempting a detailed numerical simulation, we follow *Roth and Boike* [2001] and consider a highly simplified model which provides qualitative insight, at the cost of quantitative detail, however. Basic assumptions are: (i) one-dimensional (vertical) conduction of heat is the dominant process, (ii) production of latent heat is instantaneous and may be calculated from the soil freezing characteristic and from the vapor pressure curve, respectively, and (iii) the inherent thermal properties of the soil are uniform and only depend on the mass fractions of its constituents (matrix, liquid water, ice, air). Obviously, these assumptions do not cover convective processes like rainfall events where warm water infiltrates and freezes or water which evaporates diffuses and condenses. Indeed, water movement is neglected altogether.

We start from the decompositions $C_h(T) = C_h^* + C_h^{\mathrm{lat}}(T)$ and $K_h(T) = K_h^* + K_h^{\mathrm{lat}}(T)$ of the thermal material properties into an inherent part, superscript $*$, and one that depends on phase transitions, superscript lat, and decompose (8.60) into

$$\begin{aligned} C_h^* \partial_t T - K_h^* \partial_{zz} T &= 0 \\ \partial_t [C_h^{\mathrm{lat}} T] - \partial_z [K_h^{\mathrm{lat}} \partial_z T] &= r_h^{\mathrm{ext}} \, . \end{aligned} \tag{8.62}$$

Notice that in this formulation, the source term r_h^{ext} only accounts for processes that are not yet included in the terms on the left hand side. These are in particular convective transport, which is most of the time negligible, and non-equilibrium phase changes with characteristic times that are small on the internal time scale $\ell^2 C_h / K_h$, where ℓ is a characteristic length of the observed system. With $\ell = \mathcal{O}(0.1 \text{ m})$ and $C_h/K_h = D_d^{-1} = \mathcal{O}(10^6 \text{ m}^{-2}\text{s})$ we obtain $\tau = \mathcal{O}(10^4 \text{ s})$, some 3 h.

Further notice that, strictly, such a decomposition is only useful for linear processes for which the principle of superposition allows to add the individual

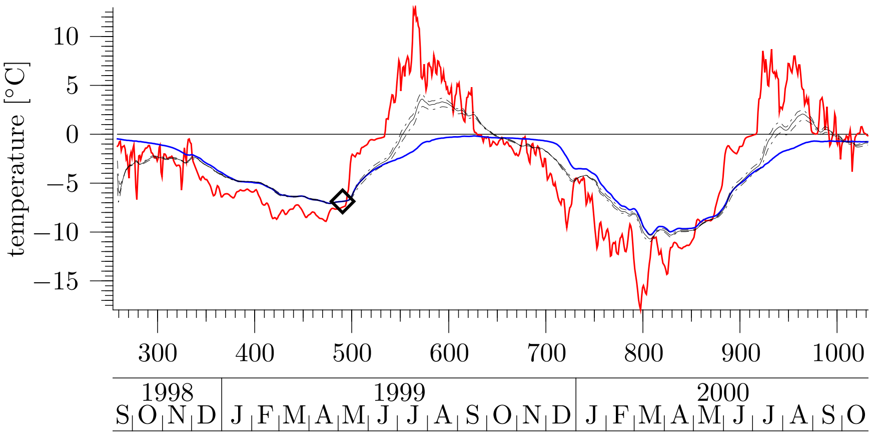


Figure 8.21. Temperature measured near the surface at 0.065 m depth (red) and projected to the bottom of the active layer at 1.25 m depth (black) to compare with measured temperature there (blue). The optimal value for the thermal diffusivity was found to be $D_h^* = 8 \cdot 10^{-7}$ m^2 s^{-1}. Projected and measured temperatures were fitted at the time indicated by the diamond. The dashed lines represent projections for a variation of D_h^* by ±20%. The deviation between projection and data for the first about 100 days results in part from the neglected initial condition and reflects the relaxation of the system.

solutions. However, Section 8.3.2 ascertains that, for most times at this wet permafrost site, the first equation of (8.62) is the dominant part with the second one being a small modification. We will thus first analyze the data based on the first equation only, and in particular estimate the value of the thermal diffusivity $D_h^* = K_h^*/C_h^*$. In the second step, the parameters will be used in the full equation to obtain a rough estimate of r_h.

Temperature Projection We calculate the temperature at depth z from the time series of the surface temperature $T(0;t)$ with the convolution integral (8.30). Neglecting a possible non-uniform initial temperature distribution, this has two parameters, the temperature T_0 at time 0 and the value of the thermal diffusivity $D_h = K_h/C_h$ which determines the transfer function (8.27). Figure 8.21 shows the result of such a projection, where the surface temperature was approximated by the measured values at $z = 0.065$ m and the value of D_h^* was chosen such that an optimal agreement between projection and measurement for $z = 1.25$ m was achieved for the first cold period roughly between days 350 and 530. The optimal value of D_h^* was found to be $8 \cdot 10^{-7}$ m^2 s^{-1}.

The agreement between projection and measurement is excellent for the cold period which demonstrates that during this time, heat transport at the Bayelva site can indeed be described by an effective conduction. Projections to more shallow depths are not shown here but they yield comparably good results and thereby corroborate the uniformity assumption.

As expected, the projection fails during the thawing period where a macroscopic phase boundary separates the two depths. Projected temperatures are much too high because the latent heat consumed for the thawing is not accounted for. Notably, however, projection and measurement again agree quite reasonably during the next cold period without adjusting any of the parameters. This indicates that (i) the thermal properties of this mud boil are constant on the time scales considered here and (ii) the net change of internal energy through the thawed period is negligible.

Heat Flux Given the value of D_h^* determined above, the thermal conductivity K_h^* can be calculated once the thermal capacity C_h^* is known. Using the weighted mean from (8.38) and inserting typical values for the volume fractions of the soil's constituents – mineral matrix, water, and possibly ice – yields $C_h^* \approx 2.4 \cdot 10^6\ \mathrm{J\,m^{-3}K^{-1}}$ for the thawed and $C_h^* \approx 2.0 \cdot 10^6\ \mathrm{J\,m^{-3}K^{-1}}$ for the frozen state. Of course, these values change with ice and water content so we will in the following use $C_h^* = [2.2 \pm 0.2] \cdot 10^6\ \mathrm{J\,m^{-3}K^{-1}}$ as a rough estimate. It leads to $K_h^* = [1.7 \pm 0.5]\ \mathrm{W\,m^{-1}K^{-1}}$ for the thermal conductivity. With this, the conductive heat flux is obtained from Fourier's law (8.19) with the finite differences approximation

$$j_h(t;z) = -K_h^* \frac{T(z_+,t) - T(z_-,t)}{z_+ - z_-} \ , \qquad (8.63)$$

where z_+ and z_- are the depths of two adjacent temperature sensors and z is chosen as $[z_+ + z_-]/2$ (Figure 8.22). Notice that this approximation only yields estimates of j_h between the temperature sensors. Hence, the field of view shrinks.

As expected, we find negative fluxes during the cold periods, an average of $-2.6\ \mathrm{W\,m^{-2}}$ for the first and $-5.5\ \mathrm{W\,m^{-2}}$ for the second one. The higher loss of heat during the second cold period again reflects the snow cover that is thinner by about a factor of 2 for much of the winter. During the warming periods, the heat flux is positive but quite low. For the first period between days 500 and 530, its mean is $4.4\ \mathrm{W\,m^{-2}}$. This contrasts with an energy input of $16.5\ \mathrm{W\,m^{-2}}$ from net radiation. The difference is consumed by the latent heat of the melting snow cover and by the evaporating water. The corresponding numbers for the second warming period between days 880 and 915 are $5.7\ \mathrm{W\,m^{-2}}$ for the mean heat flux and $26.6\ \mathrm{W\,m^{-2}}$ for the net radiation.

Latent Heat With C_h^* and K_h^* estimated, we may now proceed to roughly estimate the production r_h of latent heat for the various thermal regimes. The concept is to formulate a finite differences approximation of the first equation of (8.62), to insert the measured temperatures together with the estimated parameters C_h^* and K_h^*, and to interpret the residual, the deviation from 0, as $r_h = r_h^* + r_h^{\mathrm{ext}}$, where r_h^* represents the production of latent heat

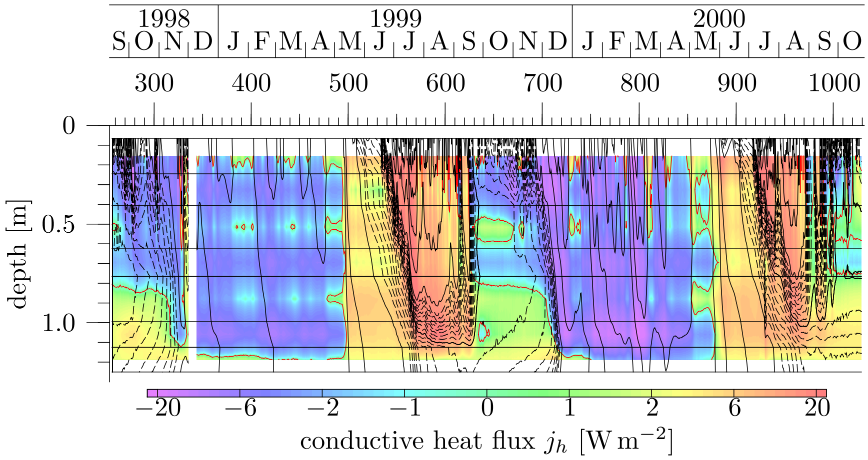


Figure 8.22. Effective conductive heat flux calculated from measured temperatures, positive in the direction of increasing depth. Black contours are isotherms, red lines show $j_h = 0$. The uncertainty of absolute values is about 30% due to the uncertainty of K_h^*. This does not affect the sign of j_h, however. The sign is sensitive to the bias of the temperature sensors, though, and corresponding artifacts may be suspected in the isothermal plateau and at the onset of the warming period.

from the internal processes as described by the second equation of (8.62) (Figure 8.23). Again, the field of view for r_h is smaller than that for T. Indeed, the effect is stronger than it was with j_h because now a second derivative is required.

We first consider the cold period and keep the soil freezing characteristic in mind which links temperature and liquid water content. Hence, as temperatures decrease, some liquid water must disappear. This may be accomplished by freezing or by evaporation and diffusive transport of the vapor. These two pathways can be distinguished based on the production of latent heat: it is positive for the freezing and negative for evaporation. Figure 8.23 indicates that during the cold period, excess water freezes below about 1 m and evaporates between 0.5 and 1 m. This vapor may be expected to diffuse upwards, towards lower temperatures. There it will condenses again and release the corresponding latent heat. The data indeed hint at a heat production above some 0.5 m. Why is there no evaporation and diffusion below 1 m but just freezing? This is the depth of the active layer. Below it, the ice is quite impermeable with the liquid and vapor phase hardly connected. Hence the only available pathway for reducing the liquid water content is freezing. We finally notice that these phenomena are more pronounced in the second cold period where temperature gradients are considerably higher.

With the beginning of the warming period, the temperature gradient reverses and with it the vapor movement. Condensation of water vapor

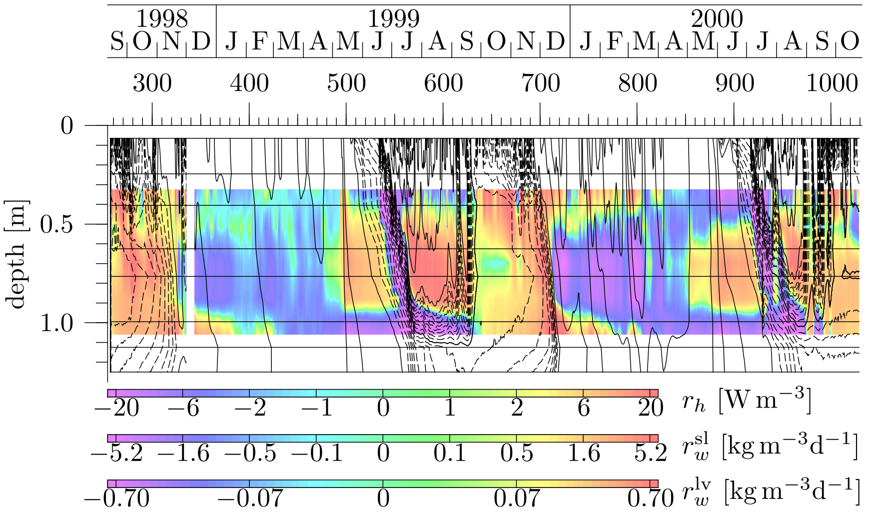


Figure 8.23. Production rate r_h of latent heat estimated from measured temperatures. The color bars also indicate the equivalent mass for the transition from the solid to the liquid phase, r_w^{sl}, and from the liquid to the vapor phase, r_w^{lv}.

and melting of ice both contribute to the required increase of the liquid water content but they lead to opposite signs of r_h. The data indicate that production from condensation dominates until the thawing front sets in. Keep in mind, though, that this does not mean that more mass of liquid water is provided by condensation than by melting since the latent heats for these two processes differ by a factor of about 7.5.

In the thawed zone, the data hint at an increasing production of latent heat towards the ice front and a very strong consumption below about 1 m. The former may again be explained by water vapor migration towards colder regions and condensation whereas the latter results from melting.

With the onset of the freeze-back and the formation of the isothermal plateau, the process direction changes one more time. Water freezes and releases the corresponding latent heat, first at a rather moderate rate – the heat cannot be transport away and thus limits the rate of freezing – towards the end of the plateau when conduction becomes more effective again, with a rapidly increasing rate.

Exercises

8.1 Enthalpy of Phase Transitions of Water To appreciate the energies associated with phase changes of water, calculate (i) ΔT if the respective enthalpies were invested in heating of the same mass of liquid water and (ii) Δz if they were invested into lifting the mass in the gravitational field of the Earth.

8.2 Humidity and Matric Potential Calculate the relative humidity for a soil at 10°C and matric heads of 1 and 10 m, or a potential of -1.5 MPa (permanent wilting point).

Conversely, calculate the matric potential for air at 20°C relative humidities of 90% (damping wet), 50%, and 10% (desert).

8.3 Zero-Curtain during Warming The zero-curtain closes upon cooling of permafrost soils and leads to the isothermal plateau where temperatures remain near 0°C for an extended time. Why is the zero-curtain not also a prominent feature during the warming period?

8.4 Permafrost Penetration Assume a uniform soil with water content $\theta = 0.3$ and mean annual ground surface temperature (MAGT) of $T_0 = -10°\mathrm{C}$. Let the average geothermal flux be $0.05\ \mathrm{W\,m^{-2}}$. Presuming for simplicity that the soil freezes at 0°C, what is the depth of the permafrost bottom in the stationary state? What is the thermal gradient above and below the permafrost bottom? Assume reasonable values wherever required.

8.5† A Simple Model for Permafrost Dynamics Assume a uniform soil with water content $\theta = 0.3$ and an average geothermal flux of $0.05\ \mathrm{W\,m^{-2}}$ as in Exercise 8.4. Again assume all the water to freeze at 0°C. Consider two cases: (i) The soil is initially in a stationary state with the surface at 0°C. For $t > 0$, the surface temperature drops to $T_0 < 0°\mathrm{C}$ and remains constant afterwards. (ii) The soil is initially in a stationary state with the surface at $T_0 < 0°\mathrm{C}$. For $t > 0$, the surface temperature raises to $T_1 > 0°\mathrm{C}$ and remains constant there. Then:

1. Calculate the temporal evolution of the freezing and of the thawing front, respectively. Neglect the heat capacity for simplicity.
2. Estimate the ice distribution at $t = 0$ for the following scenario (relict permafrost from last ice age): At $t = -30$ ky, the soil is in a stationary state with MAGT at 0°C. MAGT then drops abruptly to $T_0 = -10°\mathrm{C}$ and remains there until $t = -12$ ky. Then, MAGT jumps to $+1°\mathrm{C}$ and remains there until $t = 0$. Again neglect the heat capacity and make reasonable assumptions whenever required.

Appendix

A
Mathematical Tools

This appendix provides an introduction to mathematical tools that are employed in the main text. Naturally, this constitutes a very limited selection and the aim is intuitive understanding rather than mathematical rigor.

A.1
Vector Analysis

A.1.1
Gauss Theorem

Let V be an arbitrary volume bounded by ∂V and let $\mathbf{v}$ be a vector field that has no singularity in V. Then

$$\int_{\partial V} \mathbf{v} \cdot \mathrm{d}\mathbf{A} = \int_V \nabla \cdot \mathbf{v}\, \mathrm{d}V\ . \tag{A.1}$$

This theorem is also referred to as *Ostrogradski* theorem.

A.1.2
Total Derivative

Let $f(\mathbf{x}, t)$ be a scalar. The temporal rate of change of f at fixed location $\mathbf{x}$ is given by the partial derivative $\partial_t f(\mathbf{x}, t)$. The temporal rate of change experienced by an observer who moves on the trajectory $\xi(t)$ then is

$$\mathrm{d}_t f = \partial_t f + \mathbf{v} \cdot \nabla f\ , \tag{A.2}$$

where $\mathbf{v} = \partial_t \xi(t)$ is the observer's velocity. The derivative d_t is also referred to as *substantial* or as *material* derivative.

Let $\mathbf{w}$ be a vector with Cartesian components w_i. The total derivative of $\mathbf{w}$ is then defined by applying d_t to each component individually, hence

$$[\mathrm{d}_t \mathbf{w}]_i = \partial_t w_i + \mathbf{v} \cdot \nabla w_i =: \big[\partial_t \mathbf{w} + [\mathbf{v} \cdot \nabla]\mathbf{w}\big]_i\ . \tag{A.3}$$

A.1.3
Conservation Laws

We consider a quantity α which is conserved, i.e., neither produced nor consumed by any process of interest. This quantity may flow by itself or it may be transported by some other flow. Consider some control element in space and notice that the rate of change of the amount of α in this element must equal the net flow of the quantity through the boundary of the element. In order to formulate this statement, we introduce the density ρ_α (amount per unit volume) and flux $\mathbf{j}_\alpha$ (amount per unit area and unit time) of α. For a control element, there are two obvious choices: a volume that is fixed in space, the *Eulerian perspective*, and a fluid element that follows the movement, the *Lagrangian perspective*.

Eulerian Perspective We consider an arbitrary volume element V that is fixed in space. Let ∂V denote the boundary of V. The amount of α in this volume is $\int_V \rho_\alpha \, \mathrm{d}V$ and the outflow of α through the boundary is $\int_{\partial V} \mathbf{j}_\alpha \cdot \mathrm{d}\mathbf{A}$, where the area element $\mathrm{d}\mathbf{A}$ is pointing outwards. Conservation of α demands

$$\partial_t \int_V \rho_\alpha \, \mathrm{d}V + \int_{\partial V} \mathbf{j}_\alpha \cdot \mathrm{d}\mathbf{A} = 0 \ . \tag{A.4}$$

Exchanging differentiation and integration in the first term and applying Gauss' theorem to the second one yields

$$\int_V \left[\partial_t \rho_\alpha + \nabla \cdot \mathbf{j}_\alpha\right] \mathrm{d}V = 0 \ . \tag{A.5}$$

Since V is arbitrary, the integrand must vanish, hence

$$\partial_t \rho_\alpha + \nabla \cdot \mathbf{j}_\alpha = 0 \ . \tag{A.6}$$

Not all quantities are conserved in the above sense, however. Examples are water taken up by plant roots, momentum dissipated into through friction, or solutes that appear and disappear due to dissolution/precipitation or chemical reactions. It is convenient to account for such processes by introducing a "source-term" into (A.6) by writing

$$\partial_t \rho_\alpha + \nabla \cdot \mathbf{j}_\alpha = r_\alpha \ , \tag{A.7}$$

where r_α is the "production" rate of quantity α (amount per unit volume and unit time). A negative value of r_α corresponds to a "consumption" of α. Since (A.7) no more formulates the conservation of α, it is referred to as a *continuity* or *balance* equation which just describes the whereabouts of quantity α.

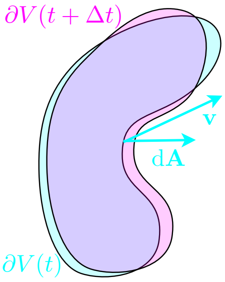


Figure A.1.
Conservation from Lagrangian perspective. The fluid element with boundary $\partial V(t)$ and area element $\mathrm{d}\mathbf{A}$ moves a small distance in the velocity field $\mathbf{v}$ during time Δt and becomes deformed with new boundary $\partial V(t+\Delta t)$. By definition, the mass of the element is conserved while the volume is not. Quantities other than the fluid mass may cross the elements boundary

Lagrangian Perspective An important tool for working in the Lagrangian context – following an arbitrary fluid element as it moves with the flow – is the total derivative $\mathrm{d}_t = \partial_t + \mathbf{v}\cdot\nabla$, (A.2), where $\mathbf{v}$ is the velocity of the flow field. Relations deduced from the Lagrangian perspective can obviously be transformed into the equivalent relations in the Eulerian view, and vice versa, since the physics must not depend on our point of view.

First consider the conservation of mass in a fluid element with volume $V(t)$ and boundary $\partial V(t)$ with outward pointing area element $\mathrm{d}\mathbf{A}$ (Figure A.1).

Denoting the mass density of the fluid by $\rho(\mathbf{x},t)$ the rate of change of the mass in the fluid element may be calculated as

$$\mathrm{d}_t \int_{V(t)} \rho\,\mathrm{d}V = \int_{V(t)} \partial_t\rho\,\mathrm{d}V + \int_{\partial V(t)} \rho\mathbf{v}\cdot\mathrm{d}\mathbf{A} = \int_{V(t)} \partial_t\rho + \nabla\cdot[\rho\mathbf{v}]\,\mathrm{d}V\ , \quad \text{(A.8)}$$

where Gauss' theorem was used to obtain the last equality. Since fluid mass is conserved $\mathrm{d}_t \int_{V(t)} \rho\,\mathrm{d}V = 0$ and further, since $V(t)$ can be chosen arbitrarily, $\partial_t\rho + \nabla\cdot[\rho\mathbf{v}] = 0$. Recognizing $\rho\mathbf{v}$ as the mass flux of the fluid, (A.6) is recovered.

Next consider some quantity α other than the fluid mass, e.g., heat or dissolved chemicals, and denote its density by $\rho_\alpha(\mathbf{x},t)$. In general, the flux of α across the boundary of a fluid element does not vanish because only its *convective* component $\rho_\alpha\mathbf{v}$ is bound to the motion of the fluid. Non-convective components arise for instance from heat conduction or from molecular diffusion. We thus decompose the flux into these two components, $\mathbf{j}_\alpha = \rho_\alpha\mathbf{v} + \mathbf{j}_\alpha^{\text{nonconv}}$, and notice that the rate of change of α in the fluid element is

$$\mathrm{d}_t \int_{V(t)} \rho_\alpha\,\mathrm{d}V = -\int_{\partial V(t)} \mathbf{j}_\alpha^{\text{nonconv}}\cdot\mathrm{d}\mathbf{A} = -\int_{V(t)} \nabla\cdot\mathbf{j}_\alpha^{\text{nonconv}}\,\mathrm{d}V\ . \quad \text{(A.9)}$$

Inserting this in (A.8), we again recover (A.6) by arguing that the resulting equation must be true for every choice of $V(t)$.

We remark that the Lagrangian perspective is often conceptually simpler since it separates processes related to the fluid flow, i.e., the convection of quantity α, from processes that directly affect α, e.g., diffusion or chemical reactions between different constituents. However, the resulting formulations are usually technically more difficult.

A.1.4
Stream Function

For incompressible flow $\nabla \cdot \mathbf{v} = 0$. Hence one may introduce the vector potential $\mathbf{\Phi}$ such that $\mathbf{v} = \nabla \times \mathbf{\Phi}$. This is particularly useful for two-dimensional flow fields, where only one component of $\mathbf{\Phi}$ is non-zero. Call this component φ and define it in Cartesian coordinates through

$$v_x = \frac{\partial \varphi}{\partial y} \ , \quad v_y = -\frac{\partial \varphi}{\partial x} \ . \tag{A.10}$$

Clearly, this velocity field satisfies $\nabla \cdot \mathbf{v} = 0$. The function $\varphi(x, y)$ is called the *stream function* because it is constant along a stream line. Indeed,

$$\mathrm{d}\varphi = \frac{\partial \varphi}{\partial x}\mathrm{d}x + \frac{\partial \varphi}{\partial y}\mathrm{d}y = -v_y \mathrm{d}x + v_x \mathrm{d}y \tag{A.11}$$

and, since by definition $\mathrm{d}y/\mathrm{d}x = v_y/v_x$ along a stream line, $\mathrm{d}\varphi = 0$.

Given a two-dimensional velocity field, the corresponding stream function is obtained by integrating (A.11) as

$$\varphi(\mathbf{x}) = \varphi(\mathbf{x}_0) + \int_{\mathbf{x}_0}^{\mathbf{x}} \mathbf{v}(\mathbf{x}') \times \mathrm{d}\mathbf{x}' \ , \tag{A.12}$$

or, in Cartesian coordinates and using that φ is a potential, as

$$\varphi(x, y) = \varphi(x_0, y_0) - \int_{x_0}^{x} v_y(x', y) \, \mathrm{d}x' + \int_{y_0}^{y} v_x(x, y') \, \mathrm{d}y' \ , \tag{A.13}$$

which clearly satisfies (A.10).

A.1.5
Lagrangian Multipliers

Let $g(\mathbf{x})$ and $\phi(\mathbf{x})$ be two differentiable functions. We wish to obtain extremal values of g under the constraint that ϕ has some fixed constant value (Figure A.2). We first notice that $\nabla \phi$ is by definition orthogonal to the subspace $\phi(\mathbf{x}) = \text{const}$. Hence, for $g(\mathbf{x})$ restricted to this subspace to be extremal, ∇g must be parallel to $\nabla \phi$. If this were not the case, ∇g would have a non-vanishing component in the subspace and g could be made more extreme by moving in the corresponding direction. A necessary condition for the constrained extremum is thus

$$\nabla g + \mu \nabla \phi = 0 \ , \tag{A.14}$$

where μ is called the *Lagrangian multiplier* [e.g., *Arfken and Weber* 1995].

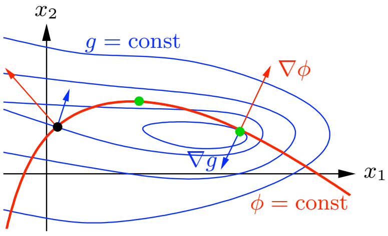


Figure A.2. Determination of extremal values of $g(\mathbf{x})$ under the constraint that $\phi(\mathbf{x})$ has some fixed value – imagine moving through some hilly terrain along the given path $\phi(\mathbf{x}) = \phi_0$ and finding local high and low points. Notice that ∇g and $\nabla\phi$ are parallel at extremal points (green dots) but not at others (black dot).

A.2
Statistical Descriptions

Only a rough outline of concepts is given in this section. For a more profound treatment refer to an appropriate textbook, e.g., *van Kampen* [1981] or *Papoulis* [1984]. Only one-dimensional quantities are considered here but the extension to more dimensions is straightforward.

A.2.1
Random Variables

A *random variable* ω is defined by (i) the set of its possible values, sometimes called the “phase space” and (ii) the distribution of probability over this set. For us it suffices to consider the singly connected interval $\Omega \subset \mathbb{R}$ as the set of possible values of ω.

Probability Distribution Functions The distribution of probability may be described by the cumulative distribution function (*cdf*)

$$P_\omega(y) := \text{Prob}\{\omega \le y\} \tag{A.15}$$

which is a non-negative and non-decreasing function of y with range $[0, 1]$. Alternatively, and more common in physics, we may use the probability density function (*pdf*)

$$p_\omega(y) := \frac{\mathrm{d}P_\omega(y)}{\mathrm{d}y} \ , \tag{A.16}$$

which is non-negative and normalized, that is $p_\omega(y) \ge 0$ and $\int_\Omega p_\omega(y)\,\mathrm{d}y = 1$.

The probability of finding ω in the interval $]y_0, y_1]$ may thus be calculated as $P_\omega(y_1) - P_\omega(y_0)$ or, for a sufficiently small interval Δy centered at y, approximated by $p_\omega(y)\Delta y$.

Expectation Values The average, or expectation value, of the random variable ω is defined as

$$\langle\omega\rangle := \int_\Omega \omega p(\omega)\,\mathrm{d}\omega \;, \tag{A.17}$$

that is, as the average of all possible values weighted with the probability $p(\omega)\,\mathrm{d}\omega$ of their occurrence. In analogy, the expectation value of some ordinary function $f(x)$ applied to ω – this maps the random variable ω to the new random variable $f(\omega)$ – is

$$\langle f(\omega)\rangle := \int_\Omega f(\omega)p(\omega)\,\mathrm{d}\omega \;. \tag{A.18}$$

Of special significance are the expectation values of powers of ω, the so-called *moments*

$$M_n(\omega) := \langle\omega^n\rangle = \int_\Omega \omega^n p(\omega)\,\mathrm{d}\omega \tag{A.19}$$

and the *central moments*

$$m_n(\omega) := \langle[\omega - \langle\omega\rangle]^n\rangle = \int_\Omega [\omega - \langle\omega\rangle]^n p(\omega)\,\mathrm{d}\omega \;. \tag{A.20}$$

The first few moments are often used to roughly describe the pdf of a random variable. Of particular interest is the *variance*, the second central moment

$$\mathrm{var}(\omega) := \langle[\omega - \langle\omega\rangle]^2\rangle = \langle\omega^2\rangle - \langle\omega\rangle^2 \tag{A.21}$$

which in the study of transport processes is related to dispersion.

A.2.2
Central Limit Theorem

Let ω_i be random variables with expectation $\langle\omega_i\rangle = \mu_i < C_\mu < \infty$ and variance $\mathrm{var}(\omega_i) = \sigma_i^2 < C_\sigma < \infty$ and define the new random variable

$$\Omega_n := \sum_{i=1}^{n} \omega_i \;. \tag{A.22}$$

If ω_i and ω_j are statistically independent for $i \neq j$, that is if for all positive integers n and m

$$\langle\omega_i^n \omega_j^m\rangle = \langle\omega_i^n\rangle\langle\omega_j^m\rangle \;, \quad i \neq j \tag{A.23}$$

then the *central limit theorem* (CLT) ascertains that in the limit $n \to \infty$ the probability density function of Ω_n approaches a Gaussian, hence

$$p_{\Omega_n}(y) \overset{n\to\infty}{\longrightarrow} \mathbb{G}(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Bigl(-\frac{1}{2}\frac{[y-\mu]^2}{\sigma^2}\Bigr) \;, \tag{A.24}$$

where $\mathbb{G}(y;\mu,\sigma^2)$ is shorthand for the Gaussian distribution with expectation $\mu = \sum_{i=1}^{n} \mu_i$ and variance $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$.

The remarkable thing about this theorem is that it does not require any assumptions about the distribution functions of the ω_i other than that expectation and variance must be finite.[1]

In a typical application of the central limit theorem, n statistically independent samples x_i, notice that these are regular numbers, of the single random variable ω are considered. For sufficiently large values of n, the pdf of $z_n = \sum_{i=1}^{n} x_i$ may be approximated by the Gaussian $\mathbb{G}(z;n\mu,n\sigma^2)$, where μ is the expectation of ω and σ^2 is its variance. As an example, consider the random variable that is obtained from summing the results of ten throws of a dice. It is instructive to run this example on a computer, try different values of n, and to compare the resulting pdf with the Gaussian. Where is the deviation smallest, where is it largest?

A.2.3
Random Functions

A *random function* $f(x,\omega)$ depends on an ordinary argument x, e.g., space or time, and on a random variable ω. It is sometimes also called a stochastic process, particularly if x refers to time. For $x = x_0$ constant, $f(x_0,\omega)$ is a random variable. For $\omega = \omega_0$ constant, $f(x,\omega_0)$ is an ordinary function, called a *realization* of the random function. A set of realizations is an *ensemble* (Figure A.3).

Probability Distribution Functions For $x = x_0$ constant, the random function simply maps the random variable ω to the new random variable $f(x_0,\omega)$ which is completely described by the corresponding cdf $P_f(y;x_0) = \text{Prob}\{f(x_0,\omega) \le y\}$ or by the pdf $p_f(y;x_0) = \mathrm{d}_y P_f(y;x_0)$. The pdf $p_f(y;x_0)$ may be expressed in terms of the pdf p_ω as

$$p_f(y;x_0) = \int_\Omega \delta\big(y - f(x_0,\omega)\big) p_\omega(\omega)\,\mathrm{d}\omega \ , \tag{A.25}$$

where Ω is the phase space of ω. This is referred to as the *1-point pdf* of the random function f. It is illustrated by the the green curve in the upper right graph of Figure A.3. In analogy, the *2-point cdf* is defined as the probability that the random function f takes on values that are simultaneously smaller than or equal to y_1 and y_2 at x_1 and x_2, respectively. Formally,

$$P_{ff}(y_1,y_2;x_1,x_2) = \text{Prob}\{f(x_1) \le y_1, f(x_2) \le y_2\} \ . \tag{A.26}$$

---

[1] The formulation $\mu_i < C_\mu < \infty$ and similarly for σ^2 in the definition of ω_i has been chosen to exclude pathological cases where μ_i is finite for all i but increases without bounds. For such a situation, the central limit theorem would not hold, that is the corresponding sum would not be described by a Gaussian distribution function. Such a situation may occur in certain models of solute transport in hierarchical media. It leads to the so-called *anomalous dispersion*.

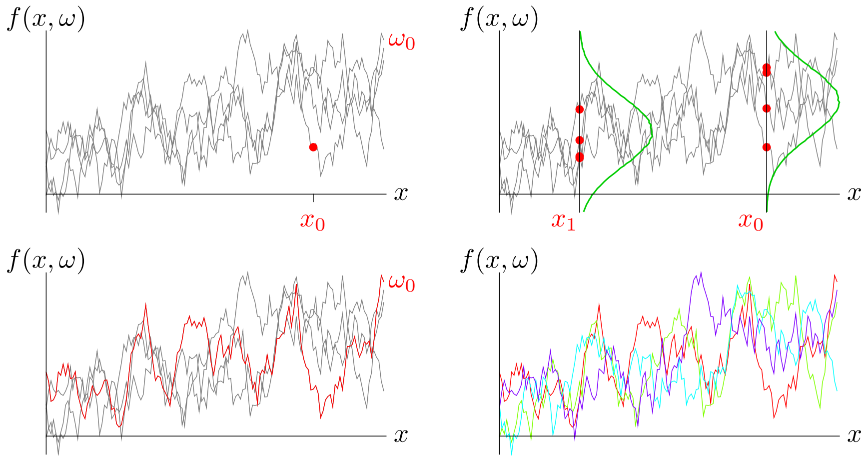


Figure A.3. Four perspectives on the random function $f(x,\omega)$: For fixed x and fixed ω, it is an ordinary number while for fixed x alone it is a random variable. For fixed ω it is an ordinary function, called a *realization* (lower left) and for both arguments free it is the *ensemble* of realizations (lower right). The green curves in the upper right graph show the pdfs of the random variable $f(x_i,\omega)$ at two locations x_0 and x_1 as estimated from some 10^5 realizations, only four of which are represented explicitly.

The pdf is again defined as the partial derivative, now with respect to both y_i,

$$p_{ff}(y_1,y_2;x_1,x_2) = \frac{\partial^2 P_{ff}(y_1,y_2;x_1,x_2)}{\partial y_1 \partial y_2} \;, \tag{A.27}$$

and it may again be expressed in terms of the pdf p_ω as

$$p_{ff}(y_1,y_2;x_1,x_2) = \int_\Omega \delta\big([y_1 - f(x_1,\omega)][y_2 - f(x_2,\omega)]\big) p_\omega(\omega)\,\mathrm{d}\omega \;, \tag{A.28}$$

where we want to assume that x_1 and x_2 are different because, if they are equal, the meaning of (A.28) reduces to that of (A.25). The n-point density functions are defined along the same line. We notice that the n-point density functions are symmetric in that permutations of the indices are immaterial.

When two random functions $f(x,\omega)$ and $g(x,\omega)$ over the same phase space are considered, the 2-point cross-cdf and -pdf are defined as

$$\begin{aligned} P_{fg}(y_1,y_2;x_1,x_2) &= \mathrm{Prob}\{f(x_1) \le y_1, g(x_2) \le y_2\} \\ p_{fg}(y_1,y_2;x_1,x_2) &= \frac{\partial^2 P_{fg}(y_1,y_2;x_1,x_2)}{\partial y_1 \partial y_2} \\ p_{fg}(y_1,y_2;x_1,x_2) &= \int_\Omega \delta\big([y_1 - f(x_1,\omega)][y_2 - g(x_2,\omega)]\big) p_\omega(\omega)\,\mathrm{d}\omega \end{aligned} \tag{A.29}$$

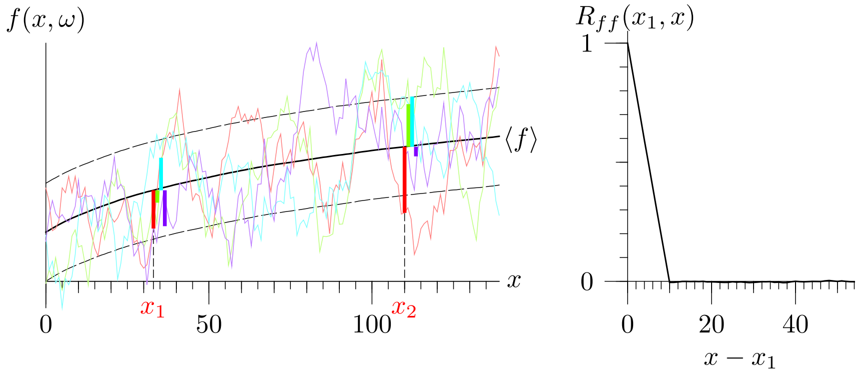


Figure A.4. Ensemble average $\langle f \rangle(x)$ (solid black line) of the random function $f(x,\omega)$ together with $\langle f \rangle(x) \pm \sqrt{C_{ff}(x,x)}$ (dashed black lines). The colored lines at x_1 and x_2 illustrate the calculation of the autocovariance function (A.31). They correspond to the deviation of the corresponding realization from the ensemble average. The deviations at two locations x_1 and x_2, multiplied and added over all realizations, lead to $C_{ff}(x_1, x_2)$. The normalized result of this calculation is shown in the graph on the right. The linear decrease reflects the construction of the random function which was obtained by a moving averaging with an interval $\Delta x = 10$ over a sequence of uniformly distributed uncorrelated random numbers. Again, 10^5 realizations have been used for these graphs, only four of which are shown (dimmed color lines).

with obvious extensions to n points and to more than two functions. We notice that the cross-density functions are in general not symmetric, that is, for instance, $p_{fg}(y_1, y_2; x_1, x_2) \neq p_{gf}(y_1, y_2; x_1, x_2)$.

Ensemble Averages In analogy to (A.17) we define the expectation function or the *ensemble average* of the random function $f(x,\omega)$ by

$$\langle f \rangle(x) := \int_\Omega f(x,\omega) p_\omega(\omega) \, \mathrm{d}\omega \tag{A.30}$$

which is an ordinary function. The ensemble average $\langle f \rangle$ is thus defined as the expectation value of the random variable $f(x,\omega)$ at each point x. We emphasis that $\langle f \rangle$ is not some "average member of the random function f": its form is generally quite different from that of the realizations (Figure A.4).

Higher moments of f are defined in complete analogy. Of particular interest is the *autocovariance function* C_{ff}, the second central moment, which is defined as

$$\begin{aligned} C_{ff}(x_1, x_2) &:= \langle f^0(x_1) f^0(x_2) \rangle \\ &= \int_\Omega \big[f(x_1,\omega) - \langle f \rangle(x_1) \big] \big[f(x_2,\omega) - \langle f \rangle(x_2) \big] p_\omega(\omega) \, \mathrm{d}\omega \ , \end{aligned} \tag{A.31}$$

where $f^0(x)$ is shorthand for $f(x) - \langle f \rangle(x)$. Sometimes, C_{ff} is called the *autocorrelation function*. However, we will only use this term for the normalized autocovariance function

$$R_{ff}(x_1, x_2) = \frac{C_{ff}(x_1, x_2)}{\sqrt{C_{ff}(x_1, x_1) C_{ff}(x_2, x_2)}} \ . \tag{A.32}$$

Clearly, C_{ff} and R_{ff} are symmetric, i.e., $C_{ff}(x_1, x_2) = C_{ff}(x_2, x_1)$. An example of R_{ff} is shown in Figure A.4.

Similar to the autocovariance and -correlation, we may define the *cross-covariance* and -correlation between two random functions $f(x, \omega)$ and $g(x, \omega)$ over the same phase space as

$$\begin{aligned} C_{fg}(x_1, x_2) &:= \langle f^0(x_1) g^0(x_2) \rangle \\ R_{fg}(x_1, x_2) &= \frac{C_{fg}(x_1, x_2)}{\sqrt{C_{ff}(x_1, x_1) C_{gg}(x_2, x_2)}} \ . \end{aligned} \tag{A.33}$$

Again, these functions are in general not symmetric.

Stationarity A random function is called stationary if its moments are invariant with respect to translation, i.e., $\langle f(x_1 + x) f(x_2 + x) \cdots f(x_n + x) \rangle = \langle f(x_1) f(x_2) \cdots f(x_n) \rangle$ for all values of n and x. It is called *weak-sense stationary* (wss) if this is true at least for $n \leq 2$.

Obviously, the expectation value of a stationary random function is a constant and the autocovariance function $C_{ff}(x_1, x_2)$ depends only on the difference $r = x_2 - x_1$, which is often called the *lag*. Autocovariance and -correlation may then be written as

$$\begin{aligned} C_{ff}(r) &= \langle f^0(x) f^0(x + r) \rangle \\ R_{ff}(r) &= C_{ff}(r) / C_{ff}(0) \end{aligned} \tag{A.34}$$

for arbitrary x. By definition, $R_{ff}(0) = 1$ and $-1 \leq R_{ff}(0) \leq 1$. Furthermore, for non-periodic functions, $R_{ff}(r)$ approaches 0 for sufficiently large values of r. This is because the correlation between the deviations from the mean at two locations decreases as the distance between them increases. A rough measure for the distance over which values are correlated is the *correlation length*

$$\ell := \int_0^\infty R_{ff}(r) \, \mathrm{d}r \ . \tag{A.35}$$

Notice that this integral need not exist. Then the correlation length is undefined, as in the case of periodic random functions. If it exists, it may be infinite indicating, if r is space, relations over arbitrary distances or, if r is time, influence of events over arbitrarily long times.

Stationarity greatly simplifies the analysis of a random function because it allows us to calculate the necessary statistics at a single location. Clearly, there remains the task of ascertaining that the function is indeed stationary.

This may be achieved based on some theoretical argument or by actually inspecting the function. Sometimes, however, it is merely postulated and justification is sought *a posteriori*. Whatever the foundation of the stationarity property, we recall that while the statistics may be calculated at a single location, we still require the entire ensemble of realizations.

Generating Realizations Sometimes, only a few averaged quantities are given for a random function and we wish to construct realizations that honor these quantities. We consider the simple but typical case of a real-valued, weakly stationary random function $f(x,\omega)$ on the finite interval $[0,X]$ with zero mean, autocovariance function $C_{ff}(r)$, and finite correlation length. One way to construct realizations is to add appropriate Fourier components.

We define the discrete Fourier sine transform of a stationary random function as

$$f(x,\omega) = \sum_{n=1}^{\infty} a_{n,\omega} \sin\Big(\frac{n\pi}{X}x + b_{n,\omega}\Big) \ , \tag{A.36}$$

where $a_{n,\omega}$ and $b_{n,\omega}$ are real-valued random variables. This somewhat uncommon definition with different phases $b_{n,\omega}$ in the Fourier transform has been chosen to accommodate stationary random functions. Without this, $f(0,\omega) = 0$ and also all the moments would vanish at $x = 0$. Notice that the functions $e_n(x) = \sin(n\pi x/X + b_n)$ with $n \in \mathbb{Z}^+$ still form an orthonormal basis on $[0,X]$ with respect to the inner product $(g,h) := [2/X]\int_0^X g(x)h(x)\,\mathrm{d}x$ since $(e_n,e_m) = \delta_{n,m}$. The inverse transform is then

$$a_{n,\omega} = \frac{2}{X}\int_0^X f(x,\omega)\sin\Big(\frac{n\pi}{X}x + b_{n,\omega}\Big)\,\mathrm{d}x \ . \tag{A.37}$$

Notice that these transforms operate on the individual realizations, hence on ordinary functions. Using the orthonormality of $\{e_n\}$, we easily obtain identity of *Parseval*

$$\frac{1}{X}\int_0^X f(x,\omega)^2\,\mathrm{d}x = \frac{1}{2}\sum_{n=1}^{\infty} a_{n,\omega}^2 \tag{A.38}$$

from (A.36). Taking the ensemble average and using stationarity – $\langle f(x,\omega)^2\rangle$ is a constant independent of x – we obtain from this

$$\langle f^2\rangle = \sum_{n=1}^{\infty}\frac{1}{2}\langle a_n^2\rangle \ . \tag{A.39}$$

This decomposes $\langle f^2\rangle$, the mean square of the fluctuation of $f(x,\omega)$, into the contributions from the single Fourier components which may also be expressed by the spectral density $S(k)$ of the fluctuations, which may be interpreted as the power

$$S(k)\,\mathrm{d}k = \sum_{k\le n\pi/X<k+\mathrm{d}k}\frac{1}{2}\langle a_n^2\rangle \tag{A.40}$$

of the fluctuations in the (spatial) frequency interval $[k, k + \mathrm{d}k]$. At this point, we invoke the *Wiener-Khinchin theorem* which relates $S(k)$ to the Fourier cosine transform of the autocovariance function $C_{ff}(r)$ of $f(x, \omega)$ by

$$S(k) = \frac{2}{\pi} \int_0^\infty \cos(kr) C_{ff}(r)\, \mathrm{d}r =: \widetilde{C}_{ff}(k) \tag{A.41}$$

and which may be proved by inserting (A.37) into (A.40).

For constructing a realization of $f(x, \omega)$, we choose $\mathrm{d}k = \pi/X$ in (A.40) such that the sum contains exactly one term, hence $\langle a_n^2 \rangle = 2S(k)\pi/X$, choose $a_n = \sqrt{\langle a_n^2 \rangle}$, and insert (A.41) to finally obtain, with $k = n\pi/X$,

$$a_n = \left[\frac{2\pi}{X} \widetilde{C}_{ff}\left(\frac{n\pi}{X} \right) \right]^{\frac{1}{2}} \tag{A.42}$$

Using these in (A.36), the realization is defined up to the phases b_n of the Fourier components which we have not used up to now. We choose them as realizations of an uncorrelated random variable with uniform probability distribution on the interval $[0, 2\pi]$. While the set $\{a_n\}$ determines the covariance function of the realization, the set $\{b_n\}$ picks a particular realization. Choosing the same set $\{b_n\}$ with different covariance functions leads to realizations that have a common large-scale structure but differ in small-scale details according to the prescribed covariance. An example is shown in Figure 5.11 on page 114.

Continuity of Realizations The autocovariance function $C_{ff}(r)$ quantifies the variance between two points of the same realization at a distance r. Hence the form of C_{ff} at the origin carries the information about the continuity of the realizations. To look into this, we consider a stationary random function over $\mathbb{R}$ with $\langle f \rangle = 0$. The generalization to non-stationary functions over $\mathbb{R}^n$ with $\langle f \rangle \neq 0$ is straightforward.

We first notice that continuity of $C_{ff}(r) = \langle f(x) f(x + r) \rangle$ at $r = 0$, that is $\lim_{r \to 0} C_{ff}(r) = C_{ff}(0)$, implies that the realizations $f(x)$ are continuous. Typical examples where this is not satisfied are random functions that contain a component with spatial white noise, i.e., uncorrelated random fluctuations. Realizations of such a function are nowhere continuous. Such a description is often used to account for some small-scale fluctuations whose correlation length is much smaller than the resolution at the scale of interest and hence cannot be represented accurately.

Odd-order derivatives of C_{ff} vanish at $r = 0$ because C_{ff} is symmetric. Hence, we next look at the second derivative of $C_{ff}(r)$, that is

$$\mathrm{d}_{rr} C_{ff}(r) = \mathrm{d}_{rr} \langle f(x) f(x + r) \rangle = \langle f(x) \mathrm{d}_{xx} f(x + r) \rangle \ . \tag{A.43}$$

Notice the change from d_{rr} to d_{xx} in the last step. We introduce the shorthand $C''_{ff}(r) = \mathrm{d}_{rr} C_{ff}(r)$ and $f(x)'' = \mathrm{d}_{xx} f(x)$, multiply both sides with $w(x) \subset C^\infty$, a regular but arbitrary function that decays rapidly for large $|x|$, and integrate over $\mathbb{R}$ to obtain

$$\begin{aligned}\int_{-\infty}^{+\infty} C''_{ff}(r)w(x)\,\mathrm{d}x &= \int_{-\infty}^{+\infty} \langle f(x)f''(x+r)\rangle w(x)\,\mathrm{d}x \\ &= \left\langle \int_{-\infty}^{+\infty} f(x)f''(x+r)w(x)\,\mathrm{d}x \right\rangle \\ &= -\left\langle \int_{-\infty}^{+\infty} f'(x)f'(x+r)w(x)\,\mathrm{d}x \right\rangle \\ &= -\int_{-\infty}^{+\infty} \langle f'(x)f'(x+r)\rangle w(x)\,\mathrm{d}x \ , \end{aligned} \tag{A.44}$$

where integration over x and ensemble average on the right hand side have been exchanged and integration by parts was used. Since (A.44) holds for arbitrary functions $w(x)$, we conclude that

$$C''_{ff}(r) = \langle f'(x)f'(x+r)\rangle \tag{A.45}$$

and with $r \to 0$ find that the first derivative of a realization is continuous if the second derivative of $C_{ff}(r)$ is continuous at the origin. One can show along the same line that a realization $f(x)$ is n times continuously differentiable everywhere if $C_{ff}(r)$ is $2n$ times continuously differentiable at the origin.

The conclusions about the continuity of realizations only hold “with probability 1” since an average is involved. This means that each realization may contain isolated points where it is not continuous and that some realizations, whose probability mass is 0, may be even more irregular.

Frozen Random Fields Random functions have been applied very successfully to turbulence theory [e.g., *Lumley* 1970]. As an example, consider the velocity field $\mathbf{v}(\mathbf{x}, t)$ in a mountain river. Its large-scale structure will clearly depend on boundary conditions like big rocks in the river bed while its small-scale structure will change rapidly in time due to the highly turbulent nature of the flow. Such a field may conveniently be described by a random function. We may actually consider time t as analogous to the random variable ω because different realizations of $\mathbf{v}(\mathbf{x})$ occur in the course of time. Ensemble averages of the velocity field are thus conveniently estimated from time series of appropriate measurements.

The situation is quite different for flow through porous media: While transport properties may vary wildly in space, and with them also the state variables and flow fields, all of them will typically be constant in time for constant boundary conditions. This leads to the notion of a *frozen random field* for which there clearly exists only a single realization. Hence, statistical properties must be estimated from the spatial structure. Let $f(\mathbf{x})$ be a

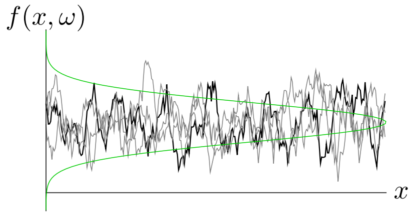


Figure A.5.
illustration of ergodic random function — to be completed —

quantity of interest, assume it to exist in $\mathbb{R}^3$, and define the spatial average as

$$\overline{f}(\mathbf{x}) := \int_{\mathbb{R}^3} f(\mathbf{x}+\boldsymbol{\xi})w(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} \ , \tag{A.46}$$

where the weight function $w(\mathbf{x}) \geq 0$ falls off sufficiently rapidly for large values of $|\mathbf{x}|$ and is normalized, i.e., $\int_{\mathbb{R}^3} w(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$. While such an average can be defined for arbitrary weight functions, it is useful only if it is objective in the sense that $\overline{f}(\mathbf{x})$ depends only weakly on the precise location or shape of $w(\mathbf{x})$. Hence, the support of $w(\mathbf{x})$ must not be larger than the REV for $f(\mathbf{x})$.[2] Higher order moments and covariance functions are defined in analogy to (A.46). For instance, the two-point covariance function, the autocovariance, is obtained as

$$\begin{aligned} &C_{ff}(\mathbf{x}_1,\mathbf{x}_2) \\ &\quad = \int_{\mathbb{R}^3} \left[f(\mathbf{x}_1+\boldsymbol{\xi}) - \overline{f(\mathbf{x}_1)}\right]\left[f(\mathbf{x}_2+\boldsymbol{\xi}) - \overline{f(\mathbf{x}_2)}\right] w(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} \ . \end{aligned} \tag{A.47}$$

In analogy to the notion introduced previously, the function $f(\mathbf{x})$ is called *stationary* if its statistics is translation invariant. It is called *weakly stationary* if this is only true up to second order statistics. Clearly, a prerequisite for stationarity is the existence of an REV for which case we will choose the weight function $w(\mathbf{x})$ such that its support equals the REV.

Ergodicity With the exception of molecular diffusion, there is nothing "random" about transport in a frozen random field. However, we often wish to account for the uncertainty in the transport phenomena that result from an insufficient knowledge of the field's detailed structure. To this end, we envisage a random function of which $f(\mathbf{x})$ is a realization. In particular, the ensemble statistics of this function shall be the same as that of $f(\mathbf{x})$. Hence, we assume that the ensemble averages of Section A.2.3 may be replaced by the corresponding averages over space, for instance Figure A.5

$$\overline{f}(\mathbf{x}) = \langle f\rangle(\mathbf{x}) \ . \tag{A.48}$$

Such a random function is called *ergodic*. We notice that in the context porous media, ergodicity is a concept that is invoked. It cannot be checked, since there exists only the one realization at hand.

---

[2] A second useful situation arises when $w(\mathbf{x})$ represents an instrument, $\overline{f}(\mathbf{x})$ thus corresponds to the measurements of f. Clearly however, measurements are useful to the extent that they are objective, hence encompasses at least an REV.

A.2.4
Geostatistics

Geostatistics may be considered as the application of the theory of random functions to problems in the geosciences. However, in the course of time a rather specialized framework has evolved that deals with the peculiarities of fields that appear there [*Matheron* 1971]. As a rule, the fields of interest are frozen, hence only a single realization is available. Consequently, averaging is almost exclusively done in space and, where necessary, the ergodicity hypothesis is invoked.

Geostatistical Characterization of Structures The basic tenet of geostatistics is that many fields of interest are not stationary, not even weakly, and that as a consequence the autocovariance function (A.34) is not well suited for describing them. Instead, the *semivariogram*

$$\gamma(\mathbf{r}) := \frac{1}{2}\overline{[f(\mathbf{x}+\mathbf{r}) - f(\mathbf{x})]^2} \tag{A.49}$$

is introduced which is well-behaved for a larger class of functions than (A.34), namely for those with stationary increments. For a weakly stationary function, semivariogram and autocovariance yield equivalent descriptions since

$$\begin{aligned} \gamma(\mathbf{r}) &= \frac{1}{2}\overline{\Big[\underbrace{[f(\mathbf{x}+\mathbf{r}) - \overline{f}]}_{=:f^0(\mathbf{x}+\mathbf{r})} - \underbrace{[f(\mathbf{x}) - \overline{f}]}_{=:f^0(\mathbf{x})}\Big]^2} \\ &= \frac{1}{2}\left[\overline{f^0(\mathbf{x}+\mathbf{r})^2} + \overline{f^0(\mathbf{x})^2} - 2\overline{f^0(\mathbf{x}+\mathbf{r})f^0(\mathbf{x})}\right] \\ &= C(0) - C(\mathbf{r}) \ . \end{aligned} \tag{A.50}$$

In practical applications, estimation of the semivariogram is the most crucial and most difficult step: most crucial because it is often the only information extracted from the available data, hence all the subsequent analyses depend on it, most difficult because the datasets are typically rather small, irregularly spaced, at least partly correlated, and often come with a sizeable error. A simple-minded estimator for (A.49) is

$$\gamma^*(\mathbf{r}) = \frac{1}{4n(\mathbf{r})} \sum_{|\mathbf{x}_i - \mathbf{x}_j - \mathbf{r}| < \epsilon} [f(\mathbf{x}_i) - f(\mathbf{x}_j)]^2 \ , \tag{A.51}$$

where $n(\mathbf{r})$ is the number of pairs of measured locations whose separation is given by $\mathbf{r}$ with an accepted interval width of ϵ. The additional factor of 2 comes from counting entries ij and ji separately. Other estimators, particularly for small datasets, are given in the literature, e.g., in *Cressie* [1993]. Whatever the method, the estimates γ^* typical scatter quite strongly

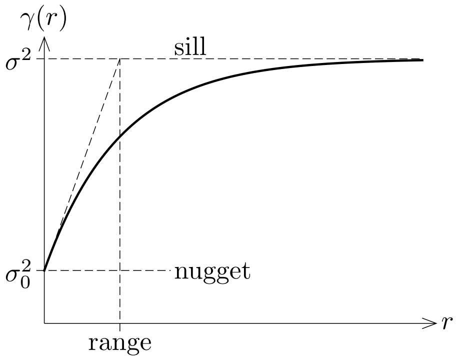


Figure A.6.
Typical semivariogram for a stationary quantity with finite correlation length (range). The variance σ_0^2 within the measurement support is called the "nugget" and the total variance σ^2 of the field is referred to as the "sill".

and it is mandatory to model the most appropriate semivariogram, i.e., to prescribe its functional form and estimate the pertinent parameters.

An often used model expression for an isotropic semivariogram for a stationary quantity with finite correlation length ℓ is

$$\gamma(r) = \sigma_0^2 + [\sigma^2 - \sigma_0^2]\left[1 - \exp\left(-\frac{r}{\ell}\right)\right] , \tag{A.52}$$

where σ_0^2 is the variance attributed to variation within the measurement's support, the so-called *nugget* and σ^2 is the variance of the entire field, the so-called *sill* (Figure A.6).

We notice that the semivariogram may be interpreted as a decomposition of the field's variance into different spatial scale.

Geostatistical Interpolation (Kriging) We consider a set of unbiased measurements $\{f_i,\ i = 1, \dots, n\}$ at locations $\{\mathbf{x}_i,\ i = 1, \dots, n\}$ and wish to obtain for the values of f at an arbitrary location $\mathbf{x}$ the **b**est **l**inear **u**nbiased **e**stimator, sometimes called a BLUE, of the form

$$f^*(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i f_i \tag{A.53}$$

that accounts properly for the spatial structure of the data. To this end, we interpret the set $\{f_i\}$ as originating from a realization of an ergodic random function $f(\mathbf{x};\omega)$. For $f^*(\mathbf{x})$ to be unbiased, we demand

$$\langle f^*(\mathbf{x}) - f(\mathbf{x};\omega)\rangle = 0 , \tag{A.54}$$

and for it to be the best estimator, we require the estimation variance to be minimal, hence

$$\sigma_{\text{est}}^2 := \left\langle [f^*(\mathbf{x}) - f(\mathbf{x};\omega)]^2 \right\rangle . \tag{A.55}$$

Notice that $\{f_i\}$ is a set of random variables, hence $f^*(\mathbf{x})$ is also a random variable.

Weakly Stationary Random Space Function In order to develop the ideas, we consider the simplest case where $f(\mathbf{x};\omega)$ is weakly stationary and we choose $\langle f(\mathbf{x};\omega)\rangle = 0$. For this case, (A.54) is always satisfied since, inserting (A.53),

$$\langle f^*(\mathbf{x}) - f(\mathbf{x};\omega)\rangle = \Big\langle \sum_{i=1}^{n} \alpha_i f_i - f(\mathbf{x};\omega)\Big\rangle = \sum_{i=1}^{n} \alpha_i \langle f_i\rangle - \langle f(\mathbf{x};\omega)\rangle \qquad \text{(A.56)}$$

and $\langle f_i\rangle = 0$ because the measurements are assumed to be unbiased. Thus, there remains (A.55) to be satisfied, that is, again inserting (A.53),

$$\sigma_{\text{est}}^2 = \sum_{i,j} \alpha_i\alpha_j \underbrace{\langle f_i f_j\rangle}_{C(\mathbf{x}_i-\mathbf{x}_j)} - 2\sum_i \alpha_i \underbrace{\langle f_i f(\mathbf{x})\rangle}_{C(\mathbf{x}_i-\mathbf{x})} + \underbrace{\langle f(\mathbf{x})^2\rangle}_{\sigma^2} \qquad \text{(A.57)}$$

shall be minimal. The weights α_i are thus chosen such that $\partial_{\alpha_i}\sigma_{\text{est}}^2 = 0$ which leads to

$$\sum_j \alpha_j C(\mathbf{x}_i - \mathbf{x}_j) = C(\mathbf{x}_i - \mathbf{x}) \ , \qquad i = 1,\ldots,n \ . \qquad \text{(A.58)}$$

This is a system of linear, inhomogeneous equations that may be written more compactly as $\mathsf{C}\boldsymbol{\alpha} = \mathbf{C}_0$ with $\boldsymbol{\alpha} = \{\alpha_1,\ldots,\alpha_n\}^{\mathrm{T}}$, $[\mathsf{C}]_{ij} = C(\mathbf{x}_i - \mathbf{x}_j)$, and $[\mathbf{C}_0]_i = C(\mathbf{x}_i - \mathbf{x})$. We notice that the matrix C inherits the symmetry from the autocovariance function. For the case where the locations $\mathbf{x}_i$ are all different and $C(0) > |C(\mathbf{r})|$ for $|\mathbf{r}| > 0$, C is positive definite. Then a single solution exists that is obtained readily using standard algorithms. Formally, it may be written as

$$\boldsymbol{\alpha} = \mathsf{C}^{-1}\mathbf{C}_0 \qquad \text{(A.59)}$$

We notice that C depends only on the location of the measurements through the known autocovariance model C. An important consequence of this is that the expensive inversion of C has to performed only once, even if interpolated values are required at many different locations.

Once the best weights are obtained, denote them by α_i^*, we may calculate the estimation variance. Inserting (A.58) into the first term of (A.57) leads to

$$\sigma_{\text{est}}^2 = \sigma^2 - \sum_i \alpha_i^* C(\mathbf{x}_i - \mathbf{x}) \ . \qquad \text{(A.60)}$$

This shows how information from nearby measurement locations reduces the variance of the estimated value. For locations that are many correlation lengths away from the nearest measurement, σ_{est}^2 approaches σ^2, the variance of the entire field (Figure A.7). Using (A.50), we may also express the estimation variance in terms of the semivariogram as

$$\sigma_{\text{est}}^2 = \sum_i \alpha_i^* \gamma(\mathbf{x}_i - \mathbf{x}) \qquad \text{(A.61)}$$

which has an analogous interpretation.

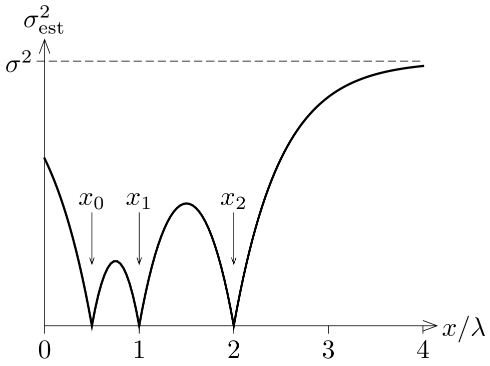


Figure A.7. Estimation variance (A.60) for the interpolation of a stationary random function from measurements at locations $x_0, \ldots, x_2$. The distance between the locations is $\ell/2$ and ℓ, respectively. The autocovariance model is assumed to be given by (A.52) with $\sigma_0 = 0$.

Notice that the variance of an estimate does not depend on the measured values, but only on their locations. Given the autocovariance model, an optimal sampling scheme that satisfies given quality criteria, possibly varying in space, may thus be designed beforehand.[3]

Random Space Function with Stationary Increments We now turn to the more common case of a random function that is not necessarily stationary itself, but whose increments are. The geostatistical literature sometimes refers to this as the "intrinsic case". In the following, we will consider semivariograms whose nugget vanishes, that is $\sigma_0 = 0$. Obviously, the nugget variance does not contain any information about spatial correlations and thus cannot be used for interpolation. It would only have to included in the end as an additive constant to the estimation variance.

We again assume $f^*(\mathbf{x})$ to be of the form (A.53) demand (A.54) and (A.55) to hold. While we found that (A.54) is always satisfied for weakly stationary functions, we now find from (A.56) that

$$\sum_{i=1}^{n} \alpha_i \overset{!}{=} 1 \tag{A.62}$$

in order to be able to represent a constant function.

---

[3] Notice that the assumption that the autocovariance model is given is a very strong stipulation in traditional applications where the field of interest is sampled and C is estimated from those measurements. Once we are in the position to obtain a good estimate for C or γ, we already know $f(\mathbf{x})$ pretty well because the measurements must lay well within a correlation length of each other in order to carry useful information.
An evolving alternative approach starts out from the premise that the most important parameter in (A.60) is the correlation length ℓ, at least for well-behaved covariance functions. The reason for this is that ℓ determines the "reach of information" while the functional form of C typically leads to higher order modifications. To obtain a reasonable estimate for ℓ, proxy variables may be used that are related to the quantity of interest, possibly not very strongly, but that may be measured with little effort and at a high spatial resolution. An example for this is ground penetrating radar (GPR) which allows to map dielectric subsurface structure that can eventually be related to hydraulic properties. Using such an estimate of ℓ, some covariance or semivariogram model is chosen, e.g., (A.52), and the optimal sampling scheme is estimated with (A.61).

Preliminary to looking into (A.55), we notice that the semivariance between two locations $\mathbf{x}_i$ and $\mathbf{x}_j$ may be distributed according to

$$\begin{aligned} \gamma(\mathbf{x}_i - \mathbf{x}_j) &:= \frac{1}{2}\Big\langle [f(\mathbf{x}_i) - f(\mathbf{x}_j)]^2 \Big\rangle = \frac{1}{2}\Big\langle \Big[[f_i - f(\mathbf{x})] - [f_j - f(\mathbf{x})]\Big]^2 \Big\rangle \\ &= \gamma(\mathbf{x}_i - \mathbf{x}) + \gamma(\mathbf{x}_j - \mathbf{x}) - \Big\langle [f_i - f(\mathbf{x})][f_j - f(\mathbf{x})] \Big\rangle \end{aligned} \tag{A.63}$$

by introducing a third location $\mathbf{x}$. Using that $f(\mathbf{x})$ may be written as $\sum_i \alpha_i f(\mathbf{x})$ because of (A.62), we obtain for the estimation variance (A.55)

$$\begin{aligned} \sigma^2_{\text{est}} &= \Big\langle \Big[\sum_i \alpha_i [f_i - f(\mathbf{x})]\Big]^2 \Big\rangle \\ &= \sum_{ij} \alpha_i \alpha_j \Big\langle [f_i - f(\mathbf{x})][f_j - f(\mathbf{x})] \Big\rangle \\ &\overset{(A.63)}{=} -\sum_{ij} \alpha_i \alpha_j \gamma(\mathbf{x}_i - \mathbf{x}_j) + \underbrace{\sum_{ij} \alpha_i \alpha_j \gamma(\mathbf{x}_i - \mathbf{x})}_{\sum_i \alpha_i \gamma(\mathbf{x}_i - \mathbf{x}) \sum_j \alpha_j} + \underbrace{\sum_{ij} \alpha_i \alpha_j \gamma(\mathbf{x}_j - \mathbf{x})}_{\sum_j \alpha_j \gamma(\mathbf{x}_j - \mathbf{x}) \sum_i \alpha_i} \\ &= 2\sum_i \alpha_i \gamma(\mathbf{x}_i - \mathbf{x}) - \sum_{ij} \alpha_i \alpha_j \gamma(\mathbf{x}_i - \mathbf{x}_j) \; . \end{aligned} \tag{A.64}$$

The optimal weights $\{\alpha_i\}$ are obtained from minimizing this expression under the constraint (A.62). The most elegant way to achieve this is the use of Lagrangian multipliers, hence to solve

$$\begin{aligned} \nabla_\alpha \sigma^2_{\text{est}} - 2\mu \nabla_\alpha \phi &= 0 \\ \phi &= 1 \end{aligned} \tag{A.65}$$

where $\phi(\boldsymbol{\alpha}) := \sum_{i=1}^n \alpha_i$. The factor -2 in front of μ is chosen for later convenience – to have G in (A.67) symmetric – but is otherwise immaterial since μ is a constant to be determined. Inserting (A.64) we finally obtain

$$\begin{aligned} \sum_j \alpha_j \gamma(\mathbf{x}_i - \mathbf{x}_j) + \mu &= \gamma(\mathbf{x}_i - \mathbf{x}) \; , \qquad i = 1, \ldots, n \\ \sum_j \alpha_j &= 1 \end{aligned} \tag{A.66}$$

which may be written more explicitly as

$$\underbrace{\begin{pmatrix} 0 & \gamma_{12} & \cdots & \gamma_{1n} & 1 \\ \gamma_{21} & 0 & \cdots & \gamma_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \cdots & 0 & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix}}_{\mathsf{G}} \underbrace{\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \\ \mu \end{pmatrix}}_{\boldsymbol{\beta}} = \underbrace{\begin{pmatrix} \gamma_{10} \\ \gamma_{20} \\ \vdots \\ \gamma_{n0} \\ 1 \end{pmatrix}}_{\mathbf{G}_0} \tag{A.67}$$

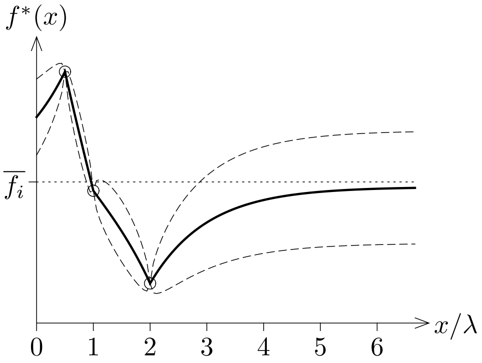


Figure A.8. Geostatistical interpolation (thick line) based on three data points f_i (symbols) together with standard deviation of estimate (dashed line) and average of data (dotted line). The setup of the data locations and the semivariogram model are the same as in Figure A.7.

with the shorthand $\gamma_{ij} = \gamma_{\mathbf{x}_i - \mathbf{x}_j}$ and $\gamma_{i0} = \gamma_{\mathbf{x}_i - \mathbf{x}}$. The diagonal of G is 0 because we assumed at the outset that $\gamma(0) = 0$. We notice that again G is symmetric because γ is an even function, that the solution may formally be written as

$$\boldsymbol{\beta} = \mathsf{G}^{-1}\mathbf{G}_0 \ , \tag{A.68}$$

and that the inverse G^{-1} needs to be calculated only once, even if $\boldsymbol{\beta}$ is to be calculated for many different locations.

As the last step, we calculate the estimation variance. We again denote the weight factors yielded by (A.68) by α^*, insert (A.66) into the second term of (A.64), and obtain

$$\sigma^2_{\text{est}} = \sum_i \alpha_i^* \gamma_{\mathbf{x}_i - \mathbf{x}} + \mu + \sigma_0^2 \ , \tag{A.69}$$

where σ_0^2 is the nugget variance that has been removed earlier and now has to be brought back in as an additive term.

Figure A.8 shows a typical example of an interpolated (kriged) dataset. We notice that the interpolation is very smooth between data points and that at distances significantly larger than a correlation length, the estimate f^* approaches the average $\overline{f_i}$ of the given data with the uncertainty approaching that of the data field.

Geostatistical Simulation Kriging attempts to estimate the average of all the realisations that honor the given set of data. Clearly, such an average is much smoother than a typical realization and it actually often belongs to a different class of functions. This is often acceptable, even desirable, if kriging is used to describe some measured quantity of interest, for instance the spatial structure of the groundwater table or of some solute concentration in the soil. The situation is quite different, however, when the set of data is used as the basis of some simulation study. Consider for instance the situation where we want to predict the migration of some dissolved contamination through groundwater. Typically, a few measurements of hydraulic conductivity will be available and we wish to estimate the conductivity field at a resolution

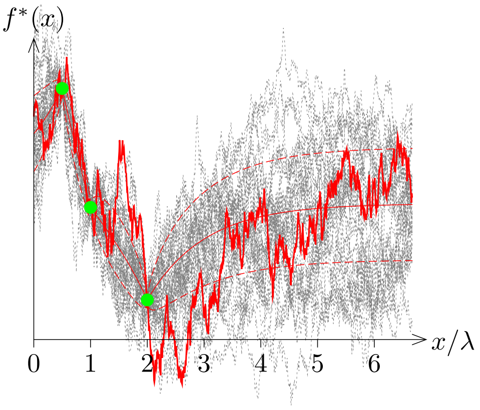


Figure A.9. Geostatistical simulations for the situation shown in Figure A.8. Green symbols represent data, thin solid and dashed red lines are kriged values and standard deviation of estimation, respectively, thick red line is one particular simulation and thin dotted black lines are 30 additional simulations.

that suffices for the simulation. Using the kriged values will clearly lead to a field that is unrealistically smooth, hence the simulated velocity field will be much too smooth and the dispersion will be underestimated. In order to deal with this difficulty, we require realizations of the random function, but only those that honor the measured data. Generating such realizations is referred to as "geostatistical simulation".

A number of algorithms have been developed to generate realizations that are conditioned on measured data. The most accurate ones start out from the given data and sequentially generate values at randomly chosen locations, typically on a grid, from kriging of all the previously assigned values [e.g., *Deutsch and Journel* 1998]. Obviously, this is a rather expensive undertaking and more simple, albeit less accurate methods often suffice.

In the simplest approach to geostatistical simulation, realizations of some weakly stationary random function with the desired spatial structure are modified such that their expectation equals that of the kriged dataset. Let $g(\mathbf{x})$ be such a realization – a possible approach to its generation is described on page 301f – and let $\{\mathbf{x}_i, f_i\}$ be the available dataset. Clearly, the covariance structure of $g(\mathbf{x})$ is chosen such as to represent the structure underlying the dataset. We define two interpolating functions: f^* obtained from kriging the dataset and g^* obtained from kriging the corresponding set $\{\mathbf{x}_i, g(\mathbf{x}_i)\}$. The two functions thus contain the analogous information, one from the dataset, the other from the realization. Finally, we define the geostatistical simulation of the dataset as

$$f(\mathbf{x}) := g(\mathbf{x}) - g^*(\mathbf{x}) + f^*(\mathbf{x}) \ . \tag{A.70}$$

Figure A.9 shows simulations of the dataset used in the previous figures. As expected, each of the simulations is much more irregular than the kriged data. Near the given data points, the variability between realizations is greatly reduced but it increases rapidly for distances larger than the correlation length.

We note in closing that there exist a number of excellent resources for geostatistical analysis. Most prominent are *GSlib* [*Deutsch and Journel* 1998], which is available from `ekofisk.stanford.edu/SCRFweb/supporting`, and *Gstat* [*Pebesma and Wesseling* 1998]

A.3
Laplace Transform

Linear partial differential equations (PDE) may often be solved efficiently by integral transforms which turn differentiation with respect to a chosen variable into an algebraic form. Thereby, the PDE can be transformed into an ordinary differential equation or even into an algebraic equation which can be solved readily. The result is then transformed back into normal space.

An integral transform of some function, say $f(t)$, consists of an integration of $f(t)$, multiplied by a weighting factor, over the entire domain of the variable t. The weighting factor contains the product of t and a parameter, the so-called conjugate variable. The domain of t determines the type of integral transform that is optimal. For the case of a semi-infinite domain, this is the Laplace transform which is defined as

$$\mathcal{L} : f(t) \mapsto \widehat{f}(s) := \int_0^\infty f(t) \exp(-st)\, \mathrm{d}t \ . \tag{A.71}$$

This definition can be extended to a function $f(t, x_1, \ldots, x_n)$ of several variables,

$$\widehat{f}(s, x_1, \ldots, x_n) := \int_0^\infty f(t, x_1, \ldots, x_n) \exp(-st)\, \mathrm{d}t \ . \tag{A.72}$$

Many functions (or generalized functions) have Laplace transforms which can be obtained by direct integration of (A.71). Examples are:

- Heaviside's Step Function $H(t)$

$$\widehat{H}(s) = \int_0^\infty \exp(-st)\, \mathrm{d}t = -\frac{1}{s} \exp(-st)\Big|_{t=0}^{t=\infty} = \frac{1}{s} \tag{A.73}$$

- Dirac's Function $\delta(t-a)$

$$\widehat{\delta}(s) = \int_0^\infty \delta(t-a) \exp(-st)\, \mathrm{d}t = \begin{cases} \exp(-sa) & a > 0, \\ 0 & a < 0 \end{cases} \tag{A.74}$$

  Notice that (A.74) is undefined for $a = 0$. In the following, we will always use $\mathcal{L}\big(\delta(t)\big) = 1$.

---

[4] The sections on Laplace and Fourier transforms have been adapted from *Jury and Roth* [1990] with minor changes.

- Power Function t^N

$$\widehat{t^N}(s) = \int_0^\infty t^N \exp(-st)\,\mathrm{d}t = \frac{N!}{s^{N+1}} \tag{A.75}$$

A.3.1

Transforms of Derivatives and Integrals

One of the major uses of Laplace transforms is in the solution of ordinary or partial differential equations, where often the dependent variable is differentiated or integrated with respect to time. The Laplace transform operation (A.71) of multiplying by $\exp(-st)$ and integrating over t from 0 to ∞ can be applied to each side of a differential equation, and to the boundary conditions. As a preface to that exercise, the Laplace transform of a derivative and integral of a function with respect to t will be evaluated. In these exercises, as well as in future calculations, extensive use will be made of the formula for integrating by parts,

$$\int_a^b u(x)\,\mathrm{d}v(x) = u(x)v(x)\Big|_a^b - \int_a^b v(x)\,\mathrm{d}u(x) \ . \tag{A.76}$$

Transforming Partial Derivatives

$$\begin{aligned}\mathcal{L}\Big[\partial_t f(x,t)\Big] &= \int_0^\infty \partial_t f(x,t)\exp(-st)\,\mathrm{d}t \\ &= f(x,t)\exp(-st)\Big|_0^\infty + s\int_0^\infty f(x,t)\exp(-st)\,\mathrm{d}t \ , \\ &= -f(x,0) + s\widehat{f}(x;s) \end{aligned} \tag{A.77}$$

Note that the initial value $f(x,0)$ of the function appears as part of the transform of the time derivative of $f(x,t)$. The Laplace transform of the second time derivative is similar, except that two integrations by parts must be carried out, hence

$$\begin{aligned}\mathcal{L}\Big[\partial_{tt} f(x,t)\Big] &= \partial_t f(x,t)\exp(-st)\Big|_0^\infty + s\int_0^\infty \partial_t f(x,t)\exp(-st)\,\mathrm{d}t \ . \\ &= -\partial_t f(x,0) - sf(x,0) + s^2\widehat{f}(x;s) \ . \end{aligned} \tag{A.78}$$

If the function is differentiated by a variable other than the one being transformed, the order of differentiation and integration may be reversed.

$$\begin{aligned}\mathcal{L}\Big[\partial_x f(x,t)\Big] &= \int_0^\infty [\partial_x f(x,t)]\exp(-st)\,\mathrm{d}t = \partial_x\int_0^\infty f(x,t)\exp(-st)\,\mathrm{d}t \\ &= \partial_x\widehat{f}(x;s) \ . \end{aligned} \tag{A.79}$$

Note that if we regard s as a parameter, then the derivative of the Laplace transform in (A.79) may be written as a total derivative rather than as a partial one.

Transforming Integrals The integral of a function over t may be evaluated using the integration by parts (A.76), hence

$$\begin{aligned}\mathcal{L}\Big[\int_0^t f(x,t')\,\mathrm{d}t'\Big] &= \int_0^\infty \int_0^t f(x,t')\exp(-st)\,\mathrm{d}t'\,\mathrm{d}t \\ &= -\frac{\exp(-st)}{s}\int_0^t f(x,t')\,\mathrm{d}t'\Big|_{t=0}^{t=\infty} + \frac{1}{s}\int_0^\infty f(x,t)\exp(-st)\,\mathrm{d}t\ , \\ &= \frac{1}{s}\widehat{f}(x;s)\ . \end{aligned} \tag{A.80}$$

A.3.2
Transformed Solution of Differential Equations

The formulas given in (A.77)–(A.80) make it possible to calculate the Laplace transform of the solution to many ordinary and partial differential equations of interest in physics and engineering, as illustrated in the next two examples.

Example: Laplace Transform Solution of the Damped Harmonic Oscillator A spring of force constant k is attached at one end to a mass m lying on a table, and at the other end to a wall at the end of the table. The equilibrium length of the spring (at which it exerts no force on the mass) is x_0. At time $t = 0$, the spring is stretched out to a position $x_1 > x_0$ and released. We wish to calculate the motion $x(t)$ of the mass for $t > 0$, assuming that the table exerts a frictional force $F = -\beta \mathrm{d}_t x(t)$ on the mass.

By Newton's first law of motion,

$$\sum_i F_i = ma = m\mathrm{d}_{tt}x = -\beta\mathrm{d}_t x - k[x - x_0]\ , \tag{A.81}$$

where F_i are the forces (friction and spring) on the mass. Equation (A.81) is to be solved subject to the initial conditions $x(0) = x_1$ and $\mathrm{d}_t x(0) = 0$. Transforming (A.81) into Laplace space using (A.73), (A.77), (A.78), and the initial conditions yields

$$ms^2\widehat{x} - msx_1 + \beta s\widehat{x} - \beta x_1 + k\widehat{x} - \frac{kx_0}{s} = 0\ . \tag{A.82}$$

This is merely an algebraic equation for the function $\widehat{x}(s)$, the transform of $x(t)$. Thus

$$\widehat{x}(s) = \frac{msx_1 + \beta x_1 + kx_0/s}{ms^2 + \beta s + k}\ , \tag{A.83}$$

which is the Laplace transform of the solution to (A.81) for the given initial conditions. Interested readers may evaluate the inverse transform using the table of transforms in Section A.3.5.

Example: Laplace Transform Solution of the Heat Equation Transforming a two-dimensional partial differential equation into Laplace space yields an ordinary differential equation which is much easier to solve. This is illustrated for the one-dimensional heat flow equation which may be written as [*Carslaw and Jaeger* 1990]

$$\partial_t T(x,t) = D_h \partial_{xx} T(x,t) \tag{A.84}$$

where D_h is the thermal diffusivity. We consider the semi-infinite medium $0 < x < \infty$ with zero initial temperature, whose inlet surface at $x = 0$ is held at the constant value T_0 for all times $t > 0$. Transforming this into Laplace space and considering for the time being the conjugate variable s as a parameter produces the ordinary differential equation

$$\mathrm{d}_{xx}\widehat{T}(x) - q^2\widehat{T}(x) = 0 \ , \quad \widehat{T}(\infty) = 0 \ , \quad \widehat{T}(0) = \frac{T_0}{s} \ , \tag{A.85}$$

where $q := \sqrt{s/D_h}$. Substituting the trial solution $\widehat{T}(x) = \exp(-mx)$ yields

$$[m^2 - q^2]\exp(-mx) = 0 \ , \tag{A.86}$$

which for arbitrary x can be satisified only if $m = \pm q$. Therefore, the general solution to (A.85) becomes

$$\widehat{T}(x) = A\exp(-qx) + B\exp(qx) \ , \tag{A.87}$$

where A and B are constants which may depend on the variable s but not on x. Since $q > 0$, the condition for $T(0)$ requires $A = T_0/s$ and the one for $T(\infty)$ leads $B = 0$, hence

$$\widehat{T}(x) = \frac{T_0}{s}\exp\Big(-x\sqrt{\frac{s}{D_h}}\Big) \ , \tag{A.88}$$

which is the Laplace transform of the solution to (A.84) with the given intial and boundary conditions.

A.3.3
Statistical Moments

Solving amenable differential equations in Laplace space is often easy. The more difficult is invariably the transformation of the solution back into real space. However, for some applications this solution may not be required and knowing the statistical moments of the solution may already suffice. This is in particular the case for many transport problems. The statistical moments of a pdf, $\widehat{p}(s)$ in Laplace space, may be calculated as

$$\langle t^n \rangle = [-1]^n \frac{\partial^n \widehat{p}(s)}{\partial s^n}\Big|_{s=0} \ . \tag{A.89}$$

This is most easily demonstrated by directly inserting the definition of $\widehat{p}(s)$ into (A.89). Taking the nth derivative with respect to s – notice that t and $p(t)$ are not affected – yields $\int_0^\infty p(t)[-t]^n \exp(-st)\,\mathrm{d}t$. Evaluating this at $s = 0$ and recalling definition (A.19) leads to (A.89).

A.3.4
Inverse Laplace Transformation

Direct Integration The transformation from Laplace space back into normal space may be formulated as

$$\mathcal{L}^{-1} : \widehat{f}(s) \mapsto f(t) := \frac{1}{2\pi \mathrm{i}} \int_{\gamma-\mathrm{i}\infty}^{\gamma+\mathrm{i}\infty} \widehat{f}(s) \exp(st)\, \mathrm{d}s \ , \tag{A.90}$$

where $\mathrm{i} := \sqrt{-1}$ and γ is a positive constant such that all singularities of $\widehat{f}(s)$ lie to the left. The integral is then evaluated using the calculus of residues [e.g., *Arfken and Weber* 1995]. This integration is often rather cumbersome, and will not be considered here any further. However, there exist several alternatives to accomplish the inversion: direct inversion using a computer mathematics program like *Mathematics* or *Maple*, numerical inversion, and inversion tables.

Numerical Inversion Numerical evaluation of (A.90) is possible in many cases if the function $f(t)$ is reasonably smooth. A Fortran program for a versatile inversion method is given in *Jury and Roth* [1990]. Numerical inversion is often computationally more efficient than direct evaluation of the analytic expressions found in inversion tables, or by contour integration. However, it is often unstable.

Inversion Tables The most common method of inversion is to use the tables of standard forms furnished in numerous mathematical handbooks. A set of the most commonly encountered transforms in solute transport is provided in Section A.3.5.

For example, the inverse transform of (A.88) may be calculated from (LT.9) as

$$T(x,t) = T_0 \mathrm{erfc}\left(\frac{x}{2\sqrt{D_h t}}\right) \ . \tag{A.91}$$

In many cases, a Laplace transform which is not in the table can be converted to one of the standard forms by various transform operations. The most useful of these are given below.

Some Theorems for the Inverse Laplace Transform

Shifting Theorem

$$\widehat{f}(s) = \mathcal{L}[f(t)]$$
$$\Downarrow$$
$$\mathcal{L}^{-1}\big[\widehat{f}(as+b)\big] = \exp\Big(-\frac{bt}{a}\Big)\mathcal{L}^{-1}\big[\widehat{f}(as)\big] = \frac{1}{a}\exp\Big(-\frac{bt}{a}\Big) f\Big(\frac{t}{a}\Big) \tag{A.92}$$

Example: Inversion of Travel Time PDF for Convection-Dispersion Solving the convection-dispersion equation produces for the travel time pdf

$$\widehat{p_t}(s;x) = \exp\Big(\frac{vx}{2D}[1-\xi]\Big) , \qquad \xi = \sqrt{1+\frac{4sD}{v^2}} . \tag{A.93}$$

This function does not appear in the table of transforms. However, using (A.92), we may transform it as follows

$$\begin{aligned} \widehat{p_t}(s;x) &= \exp\Big(\frac{vx}{2D}\Big[1-\sqrt{1+\frac{4sD}{v^2}}\Big]\Big) = \exp\Big(\frac{vx}{2D}\Big)\exp\Big(-\frac{vx}{2D}\sqrt{\frac{4D}{v^2}}\sqrt{\frac{v^2}{4D}+s}\Big) \\ &= \exp\Big(\frac{vx}{2D}\Big)\exp\Big(-\frac{x}{\sqrt{D}}\sqrt{\frac{v^2}{4D}+s}\Big) . \end{aligned} \tag{A.94}$$

This expression contains s only in the form $s + v^2/4D$. Therefore, letting $x = x/\sqrt{D}$, the inverse transform can be obtained with (A.92) as

$$p_t(t;x) = \mathcal{L}^{-1}\big[\widehat{p_t}(s;x)\big] = \exp\Big(\frac{vx}{2D}-\frac{v^2t}{4D}\Big)\mathcal{L}^{-1}\big[\exp(-x\sqrt{s})\big] \tag{A.95}$$

which may be completed with (LT.7) to obtain

$$\begin{aligned} p_t(t;x) &= \exp\Big(\frac{vx}{2D}-\frac{v^2t}{4D}\Big)\frac{x}{2\sqrt{\pi D t^3}}\exp\Big(-\frac{x^2}{4Dt}\Big) \\ &= \frac{x}{2\sqrt{\pi D t^3}}\exp\Big(-\frac{[x-vt]^2}{4Dt}\Big) . \end{aligned} \tag{A.96}$$

Convolution Theorem

$$\begin{gathered} \widehat{f_1}(s) = \mathcal{L}\big[f_1(t)\big] \quad \text{and} \quad \widehat{f_2}(s) = \mathcal{L}\big[f_2(t)\big] \\ \Downarrow \\ \widehat{f_1}(s)\widehat{f_2}(s) = \mathcal{L}\Big[\textstyle\int_0^t f_1(\tau)f_2(t-\tau)\,\mathrm{d}\tau\Big] \end{gathered} \tag{A.97}$$

where we used $\int_0^t f_1(\tau)f_2(t-\tau)\,\mathrm{d}\tau = \int_0^\infty f_1(\tau)f_2(t-\tau)\,\mathrm{d}\tau$ because, by definition, $f_2(t) = 0$ for $t < 0$. Also notice that $\int_0^t f_1(\tau)f_2(t-\tau)\,\mathrm{d}\tau = \int_0^t f_1(t-\tau)f_2(\tau)\,\mathrm{d}\tau$.

Example: Inversion of the Transform of the Travel Time CDF of the CDE The solute flux $j_s(t;x)$ through the plane at x in response to the arbitrary flux $j_s(t;0)$ through the plane at 0 is given by

$$j_s(t;x) = \int_0^t j_s(\tau;0)p_t(t-\tau;x)\,\mathrm{d}\tau , \tag{A.98}$$

hence,

$$\widehat{j_s}(s;x) = \widehat{j_s}(s;0)\widehat{p_t}(s;x) . \tag{A.99}$$

We consider as an example the response to the input flux step $j_s(t;0) = H(t)$. With (A.98), this response may be written as

$$j_s(t;x) = \int_0^t \frac{z}{2\sqrt{\pi D\tau^3}}\exp\Big(-\frac{[z-v\tau]^2}{4D\tau}\Big)\,\mathrm{d}\tau , \tag{A.100}$$

which is a difficult integral. However, using the convolution theorem with the transforms (A.73) and (A.93), we may also express this in Laplace space as

$$\widehat{j}_s(s;x) = \frac{1}{s}\exp\left(\frac{vx}{2D}\left[1-\sqrt{1+\frac{4sD}{v^2}}\right]\right) . \tag{A.101}$$

Inverting this using (A.92) and (LT.11) yields

$$\begin{aligned} j_s(t;x) &= \exp\Big(\frac{vx}{2D}-\frac{v^2t}{4D}\Big)\mathcal{L}^{-1}\Big[\frac{1}{s-V^2/4D}\exp(-x\sqrt{s})\Big] \\ &= \frac{1}{2}\left[\operatorname{erfc}\Big(\frac{x-vt}{2\sqrt{Dt}}\Big)+\exp\Big(\frac{vx}{D}\Big)\operatorname{erfc}\Big(\frac{x+vt}{2\sqrt{Dt}}\Big)\right] . \end{aligned} \tag{A.102}$$

Generalized Convolution Theorem [Walker 1987]

$$\begin{gathered} f(t_1,t_2) = \mathcal{L}^{-1}_{s_1}\Big[\mathcal{L}^{-1}_{s_2}\big[g(s_1,s_2)\big]\Big] \\ \Downarrow \\ \textstyle\int_0^t f(\tau,t-\tau)\,\mathrm{d}\tau = \mathcal{L}^{-1}\big[g(s,s)\big] . \end{gathered} \tag{A.103}$$

This is an extremely valuable tool for inverting complex transforms, by allowing the s-dependence of the transform to be split up into two terms which are inverted separately. This procedure will be illustrated in the next example.

Example: Inversion of the Travel Time PDF of the CDE Under Rate Limited Adsorption The travel time pdf of the CDE under rate limited adsorption in Laplace space is given in dimensionless form by (7.49) on page 212. In dimensional form, it may be written as

$$\widehat{p}_t(s_1,s_2;x) = \exp\Big(\frac{vx}{2D}\Big)\exp\left(-\frac{vx}{2D}\sqrt{1+\frac{4D}{v^2}\Big[s_1+\frac{s_2\beta[R-1]}{s_2+\beta}\Big]}\right) , \tag{A.104}$$

where $\beta = \alpha/\rho_b$. In the first inversion with respect to s_1, the parameter s_2 may be treated as a constant. Therefore, we may use the shifting theorem (A.92) to produce

$$\begin{aligned} &\mathcal{L}^{-1}_{s_1}\big[\widehat{p}_t(s_1,s_2;x)\big] \\ &\quad= \exp\Big(-\frac{s_2\beta[R-1]t_1}{s_2+\beta}\Big)\underbrace{\mathcal{L}^{-1}_{s_1}\left[\exp\Big(\frac{vx}{2D}\Big[1-\sqrt{1+\frac{4Ds_1}{v^2}}\Big]\Big)\right]}_{p_t^c(t_1,x)\ ,\ \text{see(A.96)}} \\ &\quad= \exp\Big(-\frac{s_2\beta[R-1]t_1}{s_2+\beta}\Big)p_t^c(t_1,x) \\ &\quad= \exp(-\beta[R-1]t_1)\exp\Big(\frac{\beta^2[R-1]t_1}{s_2+\beta}\Big)p_t^c(t_1,x) , \end{aligned} \tag{A.105}$$

where $p_t^c(t_1, x)$ is the travel time pdf of a conservative tracer expressed in terms of t_1 rather than t. The second inversion with respect to s_2 involves only the middle term, whose inverse transform is given by (LT.25). Thus

$$\begin{aligned} f(t_1, t_2) &= \mathcal{L}_{s_2}^{-1}\mathcal{L}_{s_1}^{-1}\left[\widehat{p}_t(s_1, s_2; x)\right] \\ &= \mathcal{A}(t_1)p_t^c(t_1, x)\Big[\delta(t_2) + \mathcal{T}(t_1, t_2)\mathcal{B}(t_2)\Big] , \end{aligned} \tag{A.106}$$

where

$$\begin{aligned} \mathcal{A}(t) &= \exp(-\beta[R-1]t) \\ \mathcal{B}(t) &= \exp(-\beta t) \\ \mathcal{T}(t_1, t_2) &= \beta\sqrt{\frac{[R-1]t_1}{t_2}}I_1\big(2\beta\sqrt{[R-1]t_1t_2}\big) \end{aligned} \tag{A.107}$$

and I_1 is the modified Bessel function of order 1 [*Abramowitz and Stegun* 1970]. With (A.103) we finally obtain

$$p_t(t; z) = \mathcal{A}(t)p_t^c(t; x) + \int_0^t \mathcal{A}(\tau)\mathcal{T}(\tau, t-\tau)\mathcal{B}(t-\tau)p_t^c(\tau; x)\,\mathrm{d}\tau \ . \tag{A.108}$$

A.3.5
Table of Laplace Transform Pairs

This table is adapted from *Jury and Roth* [1990] who compiled it from entries found in *van Genuchten and Alves* [1982], *Abramowitz and Stegun* [1970], and *Walker* [1987]. The first reference provided the source for the abbreviations

$$\begin{aligned} \mathcal{A} &= \frac{1}{\sqrt{\pi t}}\exp\Big(-\frac{x^2}{4t}\Big) , \qquad \mathcal{B} = \operatorname{erfc}\Big(\frac{x}{2\sqrt{t}}\Big) \\ \mathcal{C}^{\pm} &= \exp\big(a^2t \pm ax\big)\operatorname{erfc}\Big(\frac{x}{2\sqrt{t}} \pm a\sqrt{t}\Big) \ . \end{aligned}$$

I_N is the modified Bessel function of integer order N; a and b are constants.

$f(t)$	$\widehat{f}(s)$	
$\delta(t)$	1	(LT.1)
1	$\frac{1}{s}$	(LT.2)
t^N	$\frac{N!}{s^{N+1}}$	(LT.3)
$\exp(-at)$	$\frac{1}{s+a}$	(LT.4)
$\frac{\sin(at)}{a}$	$\frac{1}{s^2+a^2}$	(LT.5)

$\cos(at)$	$\dfrac{s}{s^2+a^2}$	(LT.6)
$\dfrac{x}{2t}\mathcal{A}$	$\exp(-x\sqrt{s})$	(LT.7)
$\mathcal{A}$	$\dfrac{\exp(-x\sqrt{s})}{\sqrt{s}}$	(LT.8)
$\mathcal{B}$	$\dfrac{\exp(-x\sqrt{s})}{s}$	(LT.9)
$2t\mathcal{A}-x\mathcal{B}$	$\dfrac{\exp(-x\sqrt{s})}{s\sqrt{s}}$	(LT.10)
$\dfrac{\mathcal{C}^-+\mathcal{C}^+}{2}$	$\dfrac{\exp(-x\sqrt{s})}{s-a^2}$	(LT.11)
$\dfrac{\mathcal{C}^--\mathcal{C}^+}{2a}$	$\dfrac{\exp(-x\sqrt{s})}{\sqrt{s}[s-a^2]}$	(LT.12)
$\mathcal{A}-a\mathcal{C}^+$	$\dfrac{\exp(-x\sqrt{s})}{\sqrt{s}+a}$	(LT.13)
$\mathcal{C}^+$	$\dfrac{\exp(-x\sqrt{s})}{\sqrt{s}[\sqrt{s}+a]}$	(LT.14)
$\dfrac{\mathcal{B}-\mathcal{C}^+}{a}$	$\dfrac{\exp(-x\sqrt{s})}{s[\sqrt{s}+a]}$	(LT.15)
$t\mathcal{A}+\dfrac{\mathcal{C}^-}{4a}-\dfrac{\mathcal{C}^+}{4a}[1+2ax+4a^2t]$	$\dfrac{\exp(-x\sqrt{s})}{[s-a^2][\sqrt{s}+a]}$	(LT.16)
$-\dfrac{t\mathcal{A}}{a}+\dfrac{\mathcal{C}^-}{4a^2}+\dfrac{\mathcal{C}^+}{4a^2}[-1+2ax+4a^2t]$	$\dfrac{\exp(-x\sqrt{s})}{\sqrt{s}[s-a^2][\sqrt{s}+a]}$	(LT.17)
$\dfrac{t\mathcal{A}}{a^2}-\dfrac{\mathcal{B}}{a^3}+\dfrac{\mathcal{C}^-}{4a^3}+\dfrac{\mathcal{C}^+}{4a^3}[3-2ax-4a^2t]$	$\dfrac{\exp(-x\sqrt{s})}{s[s-a^2][\sqrt{s}+a]}$	(LT.18)
$[1+ax+2a^2t]\mathcal{C}^+-2at\mathcal{A}$	$\dfrac{\exp(-x\sqrt{s})}{[\sqrt{s}+a]^2}$	(LT.19)
$2t\mathcal{A}-[x+2at]\mathcal{C}^+$	$\dfrac{\exp(-x\sqrt{s})}{\sqrt{s}[\sqrt{s}+a]^2}$	(LT.20)
$\dfrac{\mathcal{C}^+}{a^2}[-1+ax+2a^2t]+\dfrac{\mathcal{B}}{a^2}-\dfrac{2t}{a}\mathcal{A}$	$\dfrac{\exp(-x\sqrt{s})}{s[\sqrt{s}+a]^2}$	(LT.21)
$\dfrac{\mathcal{C}^-}{8a^2}-\dfrac{t\mathcal{A}}{2a}[1+ax+2a^2t]$ $+\dfrac{\mathcal{C}^+}{8a^2}\left[-1+2ax+8a^2t+2a^2[x+2at]^2\right]$	$\dfrac{\exp(-x\sqrt{s})}{[s-a^2][\sqrt{s}+a]^2}$	(LT.22)

$I_0\big(2\sqrt{at}\big)$	$\dfrac{1}{s}\exp\Big(\dfrac{a}{s}\Big)$	(LT.23)
$I_0\big(2\sqrt{at}\big)\exp(-bt)$	$\dfrac{1}{s+b}\exp\Big(\dfrac{a}{s+b}\Big)$	(LT.24)
$\delta(t) + I_1\big(2\sqrt{at}\big)\sqrt{\dfrac{a}{t}}\exp(-bt)$	$\exp\Big(\dfrac{a}{s+b}\Big)$	(LT.25)
$1 - \displaystyle\int_0^a \exp(-y-bt) I_0\big(2\sqrt{ayt}\big)\,\mathrm{d}y =: J(a, bt)$ (Goldstein's J-function)	$\dfrac{1}{s}\exp\Big(-\dfrac{as}{s+b}\Big)$	(LT.26)

A.4
Fourier Transform

The Fourier transform is optimal for functions with an unbounded support and may be defined as[5]

$$\mathcal{F} : f(x) \mapsto \widetilde{f}(k) := \int_{-\infty}^{\infty} f(x)\exp(-\mathrm{i}kx)\,\mathrm{d}x \tag{A.109}$$

with the corresponding inverse

$$\mathcal{F}^{-1} : \widetilde{f}(k) \mapsto f(x) := \frac{1}{2\pi}\int_{-\infty}^{\infty} \widetilde{f}(k)\exp(\mathrm{i}kx)\,\mathrm{d}k\ , \tag{A.110}$$

where k is the wave vector, the variable conjugate to x, and $\mathrm{i} := \sqrt{-1}$. In contrast to the inverse integral (A.90) of the Laplace transform, the inverse Fourier transform is an ordinary definite integral. Therefore, any tabulation of definite integrals can be used to invert it and in addition the numerical inversion is stable and can be implemented efficiently through the FFT (Fast Fourier Transform) [e.g., *Press et al.* 2002].

One may also define the n-dimensional Fourier transform as [*Arfken and Weber* 1995]

$$\begin{aligned}
\mathcal{F} : f(\mathbf{x}) \mapsto \widetilde{f}(\mathbf{k}) &:= \int_{-\infty}^{\infty} f(\mathbf{x})\exp(-\mathrm{i}\mathbf{k}\cdot\mathbf{x})\,\mathrm{d}\mathbf{x}\ , \\
\mathcal{F}^{-1} : \widetilde{f}(\mathbf{k}) \mapsto f(\mathbf{x}) &:= \frac{1}{[2\pi]^n}\int_{-\infty}^{\infty} \widetilde{f}(\mathbf{k})\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{x})\,\mathrm{d}\mathbf{k}\ .
\end{aligned} \tag{A.111}$$

---

[5] A general definition is $\widetilde{f}(k) := \sqrt{|b|/[2\pi]^{1-a}}\int_{-\infty}^{\infty} f(x)\exp(\mathrm{i}bkx)\,\mathrm{d}x$, where popular choices for $\{a, b\}$ are $\{0, 1\}$ (physics), $\{1, -1\}$ (mathematics), $\{-1, 1\}$ (classical physics), and $\{0, 2\pi\}$ (signal processing). We choose $\{a, b\} = \{1, -1\}$ in order to obtain the same structure as for the Laplace transform introduced above.

Example: Fourier Transform of the Exponential Covariance Function The three-dimensional, isotropic, exponential covariance function may be written as

$$C_{yy}(\mathbf{r}) = \sigma_y^2 \exp(-\mathbf{r}/\ell_y) \ , \tag{A.112}$$

where $\mathbf{r}$ is the separation between locations of the random variable y. Its Fourier transform is most easily evaluated in spherical coordinates, hence, with $r = |\mathbf{r}|$ and $k = |\mathbf{k}|$,

$$\begin{aligned}
\mathcal{F}\big[C_{yy}(\mathbf{r})\big] &= \sigma_y^2 \int_0^\infty \int_0^\pi \int_0^{2\pi} \exp\Big(-\frac{\mathbf{r}}{\ell_y} + \mathrm{i}kr\cos(\theta)\Big) r^2 \sin(\theta)\,\mathrm{d}r\mathrm{d}\theta\mathrm{d}\phi \\
&= \frac{2\pi\sigma_Y^2}{\mathrm{i}k} \int_0^\infty \Big[\exp\Big(r\Big[\mathrm{i}k - \frac{1}{\ell_y}\Big]\Big) - \exp\Big(-r\Big[\mathrm{i}k + \frac{1}{\ell_y}\Big]\Big)\Big] r\,\mathrm{d}r \\
&= \frac{8\pi\sigma_Y^2\ell_y^3}{\big[1+[k\ell_y]^2\big]^2} \ ,
\end{aligned} \tag{A.113}$$

where we have used $\int_0^\infty x^n \exp(-ax)\,\mathrm{d}x = n!/a^{n+1}$.

The Fourier transform of the autocorrelation function appears in the formulation of the macrodispersion coefficients on stochastic continuum theory [*Gelhar and Axness* 1983; *Dagan* 1984]. Notice that the expression (A.113) differs by a constant factor from the one given by *Dagan* [1984] because of the difference in the definition of the Fourier transform.

Example: Travel Distance PDF for the CDE in Infinite Soil One can shown that the Fourier transform of the travel distance pdf for the CDE in infinite soil may be written as

$$\widetilde{f}_x(k;t) = \exp(-\mathrm{i}kvt - k^2Dt) \ . \tag{A.114}$$

Transforming this back into normal space yields

$$\begin{aligned}
f_x(x,t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-k^2Dt + \mathrm{i}k[x-vt])\,\mathrm{d}k \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\Big(-Dt\Big[k - \frac{\mathrm{i}[x-vt]}{2Dt}\Big]^2 - \frac{[x-vt]^2}{4Dt}\Big)\,\mathrm{d}k \\
&= \frac{1}{2\pi\sqrt{Dt}} \exp\Big(-\frac{[x-vt]^2}{4Dt}\Big) \int_{-\infty}^{\infty} \exp(-y^2)\,\mathrm{d}y \\
&= \frac{1}{2\sqrt{\pi Dt}} \exp\Big(-\frac{[x-vt]^2}{4Dt}\Big) \ ,
\end{aligned} \tag{A.115}$$

where the substitution $y := \sqrt{Dt}\big[k - \mathrm{i}[x-vt]/2Dt\big]$ was used. This shows that the travel distance pdf for the CDE in an unbounded domain is a Gaussian.

A.5
Differential Equations

A.5.1
Ordinary Differential Equations

A.5.2
Partial Differential Equations

Partial differential equations (PDEs) typically arise from the description of the dynamics of spatially extended systems. Their solution is not straightforward and a number of instruments are available to approach them. While integral transforms (Laplace, Fourier,. . .) are very powerful instruments for many linear PDEs a solution may be gained more easily by guessing an appropriate ansatz. This is demonstrated in the following for some important cases.

Diffusion Equation with Periodic Forcing As a characteristic example we consider the one-dimensional form

$$\partial_t T(z;t) - D\partial_{zz} T = 0 \tag{A.116}$$

of the heat conduction equation (8.15). In order to solve it, we employ the ansatz

$$T(z;t) = f(t)g(z) \ , \tag{A.117}$$

which presumes that temperature can be decomposed into the product of a time- and a space-function. Inserting this into (A.116) and rearranging under the premise that neither f nor g vanishes anywhere decomposes (A.116) into a time- and a space-problem, namely

$$\frac{\partial_t f(t)}{f(t)} = D\frac{\partial_{zz} g(z)}{g(z)} \ . \tag{A.118}$$

Apparently, both sides of this equation must be constant since the left hand side depends only on t and the right hand side only on z, both of which vary independently. Let this constant be γ, possibly a complex number. First solve the z-problem $\partial_{zz} g(z) = [\gamma/D] g(z)$ and obtain

$$g(z) = \beta_1 \exp\bigl(z\sqrt{\gamma/D}\bigr) + \beta_2 \exp\bigl(-z\sqrt{\gamma/D}\bigr) \ , \tag{A.119}$$

with constants β_1 and β_2.

Semi-Infinite Uniform Medium The appropriate boundary conditions for a semi-infinite medium are $g(0) = 1$ and $\lim_{z\to\infty} g(z) = 0$. This leads to

$$g(z) = \exp\bigl(-z\sqrt{\gamma/D}\bigr) \ . \tag{A.120}$$

Next, we consider the t-problem $\partial_t f(t) = \gamma f(t)$ which is solved by

$$f(t) = \alpha \exp(\gamma t) \ , \tag{A.121}$$

where α is again a constant. Inserting $f(t)$ and $g(z)$ into (A.117) finally yields

$$T(z;t) = \alpha \exp\Big(\gamma t - z\sqrt{\frac{\gamma}{D}}\Big) \ . \tag{A.122}$$

We aim at describing a system that is forced periodically at $z = 0$. As a first step we impose $T(0;t) = \alpha \exp(\mathrm{i}\omega t)$. This apparently leads to $\gamma = \mathrm{i}\omega$ and further to

$$T(z;t) = \alpha \exp\Big(\mathrm{i}\omega t - z\sqrt{\frac{\mathrm{i}\omega}{D}}\Big) = \alpha \exp\Big(\mathrm{i}\omega t - z\sqrt{\frac{\omega}{2D}}[1+\mathrm{i}]\Big) \ . \tag{A.123}$$

Introducing the penetration depth

$$z_0(\omega) = \sqrt{\frac{2D}{\omega}} \tag{A.124}$$

this may be written as

$$T(z;t) = \alpha \exp\Big(-\frac{z}{z_0(\omega)}\Big) \exp\Big(\mathrm{i}\Big[\omega t - \frac{z}{z_0(\omega)}\Big]\Big) \tag{A.125}$$

As a final step, consider the arbitrary forcing $T_0(t)$ at $z = 0$ and, with (A.110), write its Fourier representation as

$$T_0(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widetilde{T}_0(\omega) \exp(\mathrm{i}\omega t)\, \mathrm{d}\omega \ . \tag{A.126}$$

Since heat conduction equation (A.116) is linear, the principle of superposition is applicable and, employing (A.125), the solution to the above forcing may be written as

$$T(z;t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widetilde{T}_0(\omega) \exp\Big(-\frac{z}{z_0(\omega)}\Big) \exp\Big(\mathrm{i}\Big[\omega t - \frac{z}{z_0(\omega)}\Big]\Big)\, \mathrm{d}\omega \ . \tag{A.127}$$

We notice that (i) $\widetilde{T}_0(\omega)$ is in general complex which leads to a corresponding shift in the phase, (ii) for a periodic forcing the integral reduces to a sum, and (iii) this expression may be interpreted as the projection of the surface forcing to depth z.

Uniform Medium with $T = 0$ at Finite Lower Boundary Let the lower boundary be at $z = \ell$. The appropriate boundary conditions for (A.119) are then $g(0) = 1$ and $g(\ell) = 0$. This leads to

$$\beta_1 = \frac{1}{1-\kappa} \ , \qquad \beta_2 = -\frac{\kappa}{1-\kappa} \ , \qquad \kappa = \exp\big(2\ell\sqrt{\gamma/D}\big) \ . \tag{A.128}$$

Inserting this, together with $\gamma = i\omega$, into (A.119) and further into (A.117) leads to the somewhat complicated expression

$$T(z;t) = \textit{do this with Mathematica} \ . \tag{A.129}$$

[see hand-written notes]

B
Numerical Methods

The aim of this appendix is to offer a glimpse at some important numerical methods. More extensive and authoritative treatments are available and should be consulted whenever numerical approaches are used. A good starting point for general numerical algorithms is *Press et al.* [2002].

B.1
Partial Differential Equations

Partial differential equations (pde) play a dominating role in modeling dynamical systems. However, only a very small class of them can be solved analytically. This has stirred early interest in numerical methods for approximate solutions. Pioneering work was done by Euler and Gauss in the eighteenth century and some of todays methods were already in routine use a hundred years ago [*Richardson* 1910, e.g.,]. Numerical methods began to spread rapidly with the advent of digital computers and their importance increased ever since. In the past few decades, they also facilitated deeper insight into the highly nonlinear processes of flow, transport, and interaction phenomena in soils and aquifers [*Remson et al.* 1971; *Huyakorn and Pinder* 1983; *Helmig* 1997].

Before looking into some of the numerical methods for solving a pde, we recall that they typically arise as a formulation of the state or evolution of some dynamic system. For a complete description, they have to be supplemented with boundary conditions which formulate the external forcing and possibly with initial conditions if we consider an evolving system. As we will find, the necessity for such conditions arises naturally when implementing numerical solutions.

The fundamental challenge for the numerical solution of a pde is the fact that computers can handle only limited amounts of information but that an infinite amount is required for describing a continuous function in general. The solution to this is discretization, typically in space and time but sometimes also in the variable (particle tracking) or in frequency (spectral methods). We will study two widely employed approaches – finite differences

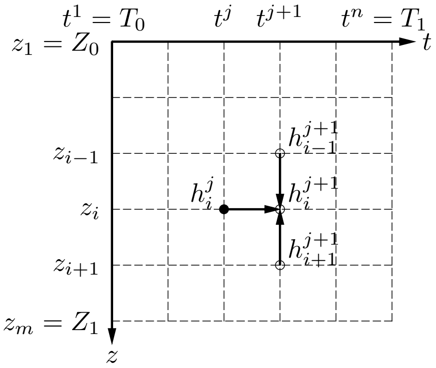


Figure B.1. Discretization of the domain $\{(z,t) \mid z \in [Z_0, Z_1],\ t \in [T_0, T_1]\}$ for the finite difference method. The arrows indicate the flow of information for an implicit formulation.

and finite elements – and use as an example the one-dimensional Richards equation in the form

$$C(h)\partial_t h - \partial_z\big[K(h)[\partial_z h - 1]\big] = 0\ . \tag{B.1}$$

Solving (B.1) means finding the function $h(z,t)$ which (i) satisfies the pde in the interior of the domain $\{(z,t) \mid z \in [Z_0, Z_1],\ t \in [T_0, T_1]\}$ and (ii) satisfies the imposed initial and boundary conditions, e.g., $h(z, T_0) = I(z)$, $h(Z_0, t) = B_0(t)$, and $h(Z_1, t) = B_1(t)$. Both, finite differences and finite elements, first discretize space and time.

B.1.1
Finite Differences

An intuitive approximation of a pde replaces derivatives by finite differences on a discrete grid with grid constant Δz in space and Δt in time (Figure B.1).

Adding and subtracting Taylor series expansions of $h(z+\Delta z)$ and $h(z-\Delta z)$, yields the approximations

$$\begin{aligned}\partial_z h(z) &\approx \frac{h(z+\Delta z) - h(z-\Delta z)}{2\Delta z}\\ \partial_{zz} h(z) &\approx \frac{h(z+\Delta z) - 2h(z) + h(z-\Delta z)}{\Delta z^2}\ .\end{aligned} \tag{B.2}$$

For $\partial_t h(t)$, we use the analogous but lower order approximation

$$\partial_t h(t) \approx \frac{h(t+\Delta t) - h(t)}{\Delta t}\ . \tag{B.3}$$

We denote a point on the grid by (z_i, t^j) and further introduce the abbreviations $\Delta t = t^{j+1} - t^j$, $h_i^j = h(z_i, t^j)$, and $C_i^{j+1} = C(h(z_i, t^{j+1}))$. With this, we approximate the first term of (B.1) by

$$C(h)\partial_t h(z_i, t^{j+1}) \approx C_i^{j+1}\frac{h_i^{j+1} - h_i^j}{\Delta t}\ . \tag{B.4}$$

The choice to evaluate C at t^{j+1} leads to a so-called implicit formulation of the finite difference approximation. It has the advantage of a higher numerical stability although at the cost of a lower accuracy. Alternatives are the Crank-Nicholson formulation $C_i^{j+\frac{1}{2}}$ which is higher order accurate but less stable and the explicit formulation C_i^j which does not require the solution of a system of equations, as we will see in the following, but which is only conditionally stable.

For the second term of (B.1), we find, after some algebraic manipulations,

$$\partial_z\big[K(h)[\partial_z h - 1]\big] \approx \frac{K^j_{i+\frac{1}{2}}[h^j_{i+1} - h^j_i] - K^j_{i-\frac{1}{2}}[h^j_i - h^j_{i-1}]}{\Delta z^2} - \frac{K^j_{i+\frac{1}{2}} - K^j_{i-\frac{1}{2}}}{\Delta z} \,. \tag{B.5}$$

Inserting (B.4)–(B.5) into (B.1) finally yields the approximation

$$C_i^{j+1}\frac{h_i^{j+1} - h_i^j}{\Delta t} - \frac{K^{j+1}_{i+\frac{1}{2}}[h^{j+1}_{i+1} - h^{j+1}_i] - K^{j+1}_{i-\frac{1}{2}}[h^{j+1}_i - h^{j+1}_{i-1}]}{\Delta z^2} + \frac{K^{j+1}_{i+\frac{1}{2}} - K^{j+1}_{i-\frac{1}{2}}}{\Delta z} = 0 \tag{B.6}$$

for (B.1). The choice to evaluate (B.5) at t^{j+1} again represents a fully implicit formulation. The flow of information for this scheme is indicated by the arrows in Figure B.1. The decision on where to evaluate C and K, and on how to approximate their values at internodal points, e.g., $K^{j+1}_{i+\frac{1}{2}}$, is crucial for the performance of the resulting method and is discussed extensively in textbooks on numerical techniques [*Helmig* 1997, e.g.,].

We notice that (B.6) cannot be evaluated at the boundary nodes z_1 and z_m because values at the nonexistent nodes z_0 and z_{m+1} are required. This is naturally resolved by introducing the boundary conditions $B_0(t)$ and $B_1(t)$ which automatically yield the values of h_1^j and h_m^j for all j.

The approximation (B.6) consists of $m-2$ equations, one for the matric head h_i^{j+1} at each node in space except for the boundary nodes, where the head is already known. These equations are nonlinear because the material properties C and K depend on the matric head h. To solve one of them, the head h_i^j for the earlier time t^j is required. In addition, also the heads at the new time t^{j+1} must be known at adjacent positions z_{i-1} and z_{i+1}. This leads to the coupling of the $m-2$ equations and requires their simultaneous solution. Efficient algorithms are readily available for solving coupled linear equations. However, they are not immediately applicable to solving (B.6) because of the nonlinearity. A common solution to this problem is to guess the head at time t^{j+1} based on the known values at t^j and to iteratively improve this guess until (B.6) is satisfied to some given accuracy.

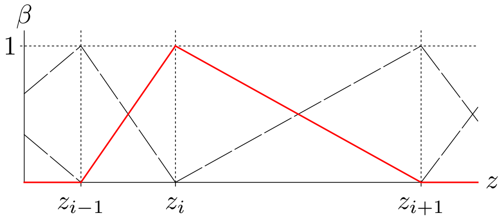


Figure B.2. Basis functions for linear finite elements. The red line is the basis function $\beta_i(z)$.

The numerical solution of (B.1) thus starts at $t^1 = T_0$ with the given initial condition $I(z)$ and iteratively calculates the matric head at $t^2 = T_0 + \Delta t$, performing a so-called time step. Repeating time steps propagate the solution through the domain of interest to $t^n = T_1$.

B.1.2
Finite Elements

A popular alternative to finite differences are finite elements which are more flexible for higher dimensional problems and for irregularly shaped domains. However, the underlying theory is more demanding and the resulting codes are more involved. Finite elements are typically used for solving the spatial part of the problem only with the temporal evolution still handled by finite differences. To illustrate the concept of finite elements, we consider stationary flow for which (B.1) simplifies to

$$\partial_z \big[K(h)[\partial_z h - 1]\big] = 0 \; . \tag{B.7}$$

With finite differences, an approximation for the value of h is sought at grid nodes and h is undefined at other locations. These values can be interpolated, of course, but this is not part of the numerical solution and hence not quality-controlled during the simulation. In contrast, the finite element formulation divides the domain of interest into small subregions, so-called elements, typically low-order simplexes. These include one-dimensional line segments, two-dimensional triangles or quadrangles, and three-dimensional tetrahedrons. Within each element, the function h is approximated by basis functions. These are typically low-order polynomials, in the most simple case linear functions.

For the one-dimensional problem (B.7), element i is the line segment between z_{i-1} and z_i, and the basis function $\beta_i(z)$ is a piecewise linear function which equals 1 at z_i and is 0 at all other nodes (Figure B.2). The β_i form a basis – not an orthogonal one – of the space of continuous, piecewise linear functions. An approximation of the solution of (B.7) from this space may thus be written as

$$\widehat{h}(z) = \sum_i h_i \beta_i(z) \; , \tag{B.8}$$

where h_i is the value of h at z_i and where the index i runs through all nodes. In general, $\widehat{h}(z)$ does not satisfy (B.7) and the residual

$$\mathcal{R}(\widehat{h}(z)) := \partial_z[K(\widehat{h}(z))[\partial_z \widehat{h}(z) - 1]] \tag{B.9}$$

is different from 0. The aim then is to find $\widehat{h}(z)$ such that $\mathcal{R}(\widehat{h})$ is minimal. Since $\mathcal{R}(\widehat{h})$ is a function, however, there exist various ways to define "minimal". A popular approach is to again use the basis functions β_i for weighing $\mathcal{R}(\widehat{h})$ and to require the resulting integral over the entire domain Ω to vanish. For our simple example, we thus want

$$\int_\Omega \beta_i(z)\partial_z \left[K(\widehat{h}(z))[\partial_z \widehat{h}(z) - 1]\right] \mathrm{d}z = 0 \ , \quad \forall i \ . \tag{B.10}$$

This approach is known as *Galerkin's method*. The first step in evaluating (B.10) is an integration by parts which leads to

$$\beta_i(z)K(\widehat{h}(z))[\partial_z \widehat{h}(z) - 1]\Big|_{\partial\Omega} - \int_\Omega \beta_i'(z)K(\widehat{h}(z))[\partial_z \widehat{h}(z) - 1] \, \mathrm{d}z = 0 \ , \tag{B.11}$$

where $\partial\Omega$ denotes the boundary of the domain Ω and $\beta_i' := \partial_z \beta_i$. This integration removes the problems that arise from using the second derivative of a function that is only piecewise linear, here the basis function β_i. It furthermore provides a natural way of including flux-type boundary conditions, since the first term in (B.11) is the negative of the water flux j_w across the boundary, weighted with the basis function. Notice that for most i, this term vanishes since β_i is only nonzero at the boundary for the first and the last basis function, respectively.

Inserting (B.8) into (B.11) and rearranging finally yields

$$\beta_i(z) j_w\Big|_{\partial\Omega} - \int_\Omega K\beta_i'(z) \, \mathrm{d}z + \sum_j \widehat{h}_j \int_\Omega K\beta_i'(z)\beta_j'(z) \, \mathrm{d}z = 0 \ , \tag{B.12}$$

where the index j runs through all nodes. Notice with Figure B.2 that most terms in the sum vanish since only overlapping basis functions contribute. If we neglect the dependence of K on h, the integrals are constant and (B.12) is a system of linear equations for the unknown heads $\widehat{h}_j$. This is for instance the case for saturated conditions. For unsaturated flow, we are again faced with the problem to interpolate K within the element and to solve a nonlinear system of equations. This has already been discussed for the method of finite differences.

B.2
Nonlinear Parameter Estimation

An inverse problem can be solved by parameterizing the underlying physics with a model $\mathcal{M}$ that depends on m parameters $\mathbf{p} = \{p_1, \ldots, p_m\}$. It then reduces to finding $\mathbf{p}$ such that the given dataset $\mathbf{d} = \{d_1, \dots, d_n\}$ for generalized locations $\mathbf{x} = \{x_1, \dots, x_n\}$ is optimally described by the corresponding model predictions $\mathcal{M}(\mathbf{x}; \mathbf{p})$. This is often accomplished by minimizing the cost function

$$\chi^2(\mathbf{p}) = \sum_{i=1}^{n} \Big[\frac{d_i - \mathcal{M}(x_i; \mathbf{p})}{\sigma_i}\Big]^2 , \tag{B.13}$$

where σ_i is the uncertainty of data d_i, where we have assumed for simplicity that the data are statistically independent. Generalized locations can be space, time, or any other independent variable. Examples of inverse problems can be found in Section 6.4.

There exist a number of approaches to minimize (B.13). In the following, we consider the workhorse of nonlinear parameter estimation, the *Levenberg-Marquardt* algorithm. It is an example of the so-called gradient methods. These are typically highly efficient but are only able to find a local minimum. If there exists only one minimum, as is the case for a well-designed pair of experiment and model, then this obviously suffices.

In general, the function $\chi^2(\mathbf{p})$ is very complicated with many local minima. However, in the following we consider only the simple case with a single well-defined minimum. If there are multiple minima, this just means that the starting point must be sufficiently near to the desired minimum. We approximate $\chi^2(\mathbf{p})$ locally by the quadratic function

$$\chi^2(\mathbf{p}) \approx \chi_0^2 + \frac{1}{2}[\mathbf{p} - \mathbf{p}^{\text{opt}}]^{\mathrm{T}}\, \mathsf{A}\, [\mathbf{p} - \mathbf{p}^{\text{opt}}] , \tag{B.14}$$

where A with elements $a_{ij} = \partial^2\chi^2/[\partial p_i \partial p_j]$ evaluated at $\mathbf{p}^{\text{opt}}$ is the constant Hessian matrix and $\mathbf{p}^{\text{opt}}$ is the parameter array that minimizes χ^2. We notice that by its definition, A is symmetric and positive definite, i.e., $\mathbf{v}^{\mathrm{T}}\, \mathsf{A}\, \mathbf{v} > 0$ for all vectors $\mathbf{v}$ with $|\mathbf{v}| > 0$. Apparently, in the approximation (B.14) the optimal parameter array $\mathbf{p}^{\text{opt}}$ can be obtained in a one-step procedure,

$$\mathbf{p}^{\text{opt}} = \mathbf{p} - \mathsf{A}^{-1}\nabla\chi^2(\mathbf{p}) , \tag{B.15}$$

from an arbitrary starting point $\mathbf{p}$. Towards the edge of the validity of the quadratic approximation, (B.15) still yields a very efficient iteration rule for calculating $\mathbf{p}^{\text{opt}}$. The matrix A then obviously has to be approximated by calculating it at $\mathbf{p}$.

If the initial estimate for $\mathbf{p}^{\text{opt}}$ is too far off, however, then (B.15) is quite useless and the best we can do is to follow χ^2 down-gradient. Naively, we could try an iteration like $\mathbf{p}^{\text{next}} = \mathbf{p}^{\text{act}} - \lambda\nabla\chi^2(\mathbf{p}^{\text{act}})$, where λ is some number. However, the components of $\mathbf{p}$ are typically quantities with completely different units and their numerical values may differ by orders of magnitude. Look at Table 6.3 on page 179 for an example where $\mathbf{p} = \{\theta_r, \theta_s, \alpha_d, \alpha_w, n, a, K_0\}^{\mathrm{T}}$. Obviously, the naive iteration is then dimensionally wrong and even if it is

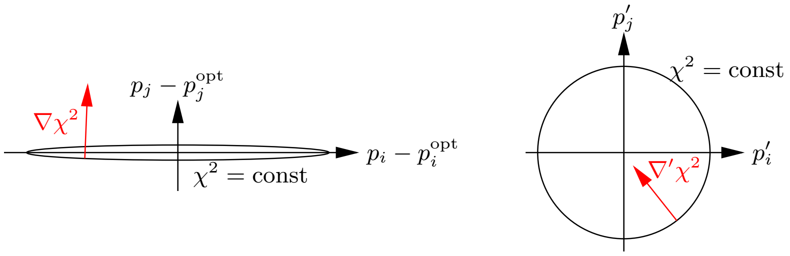


Figure B.3. Sketch for the distortion of the original parameter space $\{p_i\}$ and its remediation by the scaling $\{p'_i\} = \{\sqrt{a_{ii}}[p_i - p_i^{\rm opt}]\}$, where a_{ii} is a diagonal component of the Hessian matrix A.

reduced to the numerical values alone, the metric of the parameter space is wildly distorted (Figure B.3). To remediate this problem, we realize with (B.15) that this distortion is determined by the Hessian matrix A and introduce the dimensionless parameters

$$p'_i = \sqrt{a_{ii}}[p_i - p_i^{\rm opt}] \ . \tag{B.16}$$

The down-gradient iteration then becomes

$$\mathbf{p}'^{\rm next} = \mathbf{p}'^{\rm act} - \lambda\nabla'\chi^2(\mathbf{p}'^{\rm act}) \ , \tag{B.17}$$

where $\nabla' = \{\partial_{p'_i}\} = \{a_{ii}^{-1/2}\partial_{p_i}\}$. Translating this back into normal space finally yields the iteration

$$\mathbf{p}^{\rm next} = \mathbf{p}^{\rm act} - \lambda\mathbf{a}^{-1}\mathsf{I}\nabla\chi^2(\mathbf{p}^{\rm act}) \ , \tag{B.18}$$

where $\mathbf{a}^{-1} = \{1/a_{11}, \ldots, 1/a_{mm}\}$, I is the identity matrix, and λ is a yet undetermined number.

We thus have two different iteration rules, (B.15) for small distances from the (unknown) optimal array $\mathbf{p}^{\rm opt}$ and (B.18) for large ones. We notice that a convenient smooth transition between these two is accomplished by

$$\mathbf{p}^{\rm next} = \mathbf{p}^{\rm act} - [\mathsf{A}^{-1} + \lambda\mathbf{a}^{-1}\mathsf{I}]\nabla\chi^2(\mathbf{p}^{\rm act}) \ , \tag{B.19}$$

where λ varies between 0 near $\mathbf{p}^{\rm opt}$ and some large value far from it. How should we choose λ? This is a difficult question which determines the speed of convergence, but not the final result. A simple approach, suggested by *Press et al.* [2002], is to choose a small value to start with, say $\lambda = 0.001$ and to check if $\chi^2(\mathbf{p}^{\rm next}) < \chi^2(\mathbf{p}^{\rm act})$. If this is the case, reduce λ by a large factor, say 10, and continue the iteration. Whenever this condition is not satisfied, we are apparently too far away from the minimum and should shift more towards moving down-gradient, hence increase λ by a factor of 10.

Finally, we need stopping criteria for the iteration. There are basically two avenues available. The first one is based on the χ^2-statistics and decides if a further reduction of χ^2 is statistically significant. Indeed, with χ^2 being a sum of n independent Gaussian variates, $\langle\chi^2\rangle = n$ and $\mathrm{var}(\chi^2) = n$. Applying this to (B.13) and noticing that n data points to estimate m model parameters leads to $n - m$ degrees of freedom leads to $\langle\chi^2\rangle = n - m$ and $\mathrm{var}(\chi^2) = n - m$, provided model $\mathcal{M}$ is a correct description of the data and $\mathbf{p}$ is the correct parameter vector. The correctness of the model is a tough premiss, however. A more robust heuristic criterion for stopping the iteration is to look at absolute and relative changes of $\mathbf{p}$ and to set corresponding break points.

A final comment on the calculation of $\nabla\chi^2$ and of A is in order. For the components of the gradient we obtain with (B.13)

$$\frac{\partial\chi^2(\mathbf{p})}{\partial p_j} = -2\sum_{i=1}^{n}\frac{1}{\sigma_i^2}[d_i - \mathcal{M}(x_i;\mathbf{p})]\underbrace{\frac{\partial\mathcal{M}(x_i;\mathbf{p})}{\partial p_j}}_{=:S_j(x_i;\mathbf{p})}\ , \tag{B.20}$$

where $S_j(x_i;\mathbf{p})$ is the so-called *model sensitivity* for parameter p_j at generalized location x_i. The components of A result from taking a further derivative,

$$\begin{aligned}\frac{\partial^2\chi^2(\mathbf{p})}{\partial p_j\partial p_k} &= 2\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\left[S_j(x_i;\mathbf{p})S_k(x_i;\mathbf{p}) - [d_i - \mathcal{M}(x_i;\mathbf{p})]\frac{\partial^2\mathcal{M}(x_i;\mathbf{p})}{\partial p_j\partial p_k}\right]\\ &\approx 2\sum_{i=1}^{n}\frac{1}{\sigma_i^2}S_j(x_i;\mathbf{p})S_k(x_i;\mathbf{p})\ . \end{aligned}\tag{B.21}$$

The approximation is motivated by the following facts: (i) Calculating the second derivative is computationally expensive. (ii) In the neighborhood of $\mathbf{p}^{\mathrm{opt}}$ and for a correct model, the deviations $d_i - \mathcal{M}(x_i;\mathbf{p})$ approach random noise that is not correlated with the model. This part of the sum may thus be expected to be very small. (iii) Approximations of A, even very crude ones, only affect the path to $\mathbf{p}^{\mathrm{opt}}$, but not its final value.

C
Modeling Exercises

Soil physical processes almost invariably lead to complicated, often non-linear problems once they are studied beyond the most fundamental level. This is certainly true for all practical issues like optimal irrigation with minimal salinization, water harvesting in semi-arid regions, or contamination and remediation of soils and aquifers, to just mention some. In addition, also more conceptual considerations often lead to representations whose phenomenology is hard to visualize let alone to quantify. These challenges invariably lead to numerical simulations, an approach which in the past decades rapidly evolved into a third pillar of our scientific understanding of the World, next to experiment and theory.

This appendix collects a series of modeling exercises that expand the main themes of these lecture notes and make them more tangible. These exercises are separated into individual sections which follow the same structure: (i) The thematic background is briefly outlined. (ii) A model representation is proposed, (iii) simulated numerically, and presented. (iv) Some threads for further exploration are suggested.

The exercises in this appendix go well beyond the traditional paper-pencil-calculator approach and require more evolved tools. Given those tools, however, and their appropriate command, the exercises are no more difficult or time-consuming than those accessible to paper-pencil-calculator. There are three classes of tools that are useful: (i) computer mathematics systems, (ii) systems for technical computing, and (iii) dedicated solvers for partial differential equations. The separations between these three classes are by no means sharp and they indeed evolve rapidly. Prominent examples for the first class are *Mathematica*[1] and *Maple*,[2] which both cover large parts of mathematics and go well into the field of numerical simulations. A well-known representative of the second class is *MATLAB*,[3] a generic tool for data analysis and simulations. There exists an open source clone, *Octave*[4], and

---

[1] www.wolfram.com/mathematica

[2] www.maplesoft.com/products/Maple

[3] www.mathworks.com/products/matlab

[4] www.gnu.org/software/octave

the similar, but incompatible, open source tool *Scilab*[5]. An example for the third class, dedicated pde-solvers, is *COMSOL Multiphysics*[6]. The DUNE project's *dune-pdelab*[7] offers a research oriented open platform for developing pde-solvers. Highly specialized, and correspondingly efficient solvers for Richards' equation include *MuPhi* ($\mu\varphi$)[8] and *HYDRUS*,[9] with the latter only available for the Windows operating system, however.

In the following, we will predominantly employ *Mathematica*, *dune-pdelab*, and *MuPhi* and corresponding recipes are provided. The examples are formulated in a generic way, however, such that they may be adapted easily to other appropriate tools.

---

[5] www.scilab.org

[6] www.comsol.com/products/multiphysics

[7] www.dune-project.org/pdelab

[8] Olaf Ippisch, personal communication

[9] www.pc-progress.com/en/Default.aspx?hydrus-3d and [*Radcliffe and Šimůnek* 2010]

C.1

Stationary Water Flow in Layered Soil

Water flow in the vadose zone may often be described as stationary with episodic strong fluctuation. Situations where stationary flow prevails include (i) extended stretches of time with constant forcing, precipitation or evaporation, when only the surface layer is of interest and (ii) the vadose zone at a sufficient distance from the surface when fluctuations of the forcing have abated.

The subsurface architecture is typically heterogenous and the surface is not flat. For simplicity, we focus on situations with uniform and horizontal layers. This is appropriate for many settings like agricultural fields in large basins.

Model Formulation

Richards equation (6.2) is the appropriate formulation for the dynamics of soil water. However, for stationary flow and layered soils, it suffices to just integrate the ordinary differential equation (6.15) obtained from requiring the Buckingham-Darcy flux to be constant. This approach is also much more economical in terms of computing resources.

For the soil architecture, we consider a stack of uniform horizontal layers with each of the materials described by the simplified Mualem-van Genuchten parameterization (3.45) and (3.56) with parameters given in Table 3.1.

As boundary conditions, consider a range of infiltration and evaporation fluxes. Recall from Section 6.2.1 that the infiltrating flux is limited by the saturated hydraulic conductivity of the soil and that the maximum evaporation flux that can be sustained indefinitely is reached as $h_m \to \infty$ at the soil surface. Depending the hydraulic properties of the layers near the surface, this may already happen at very small fluxes.

Numerical Simulation

We choose to employ *Mathematica* for the integration of the ODE (6.15) with the code shown in Figure C.1. We consider two soil layers, sand on top of silt, each 1 m thick, and define the conductivity function of the corresponding stack with

```
k[h_, z_] := Piecewise[{{kmvg[h, p1], z<-1}, {...}}]
```

where `kmvg[h, p]` is the Mualem-van Genuchten parameterization with parameters `p` $= \{\alpha, n, a, K_0\}$ and `{...}` represents the second and possibly further layers.

```
(* define the conductivity function for a stack of two materials... *)
k[h_, z_] := Piecewise[{
                {kmvg[h, p1], z<-1},
                {kmvg[h, p2], z>=-1}
            }]

(* ...with Mualem-van Genuchten parameterization... *)
kmvg[h_, p_]:=p[[4]] (1+(p[[1]] h)^p[[2]])^(-p[[3]] (1-1/p[[2]]))
              (1-(p[[1]] h)^(p[[2]]-1)(1+(p[[1]] h)^p[[2]])^(-1+1/p[[2]]))^2

(* ...and parameters *)
alpha1 = -2.0; n1 = 4.00; a1 = 0.5; k01 = 10^(-4); p1={alpha1, n1, a1, k01};
alpha2 = -0.5; n2 = 1.33; a2 = 0.5; k02 = 10^(-5); p2={alpha2, n2, a2, k02};

(* set effective conductivity as reference for boundary fluxes *)
k0eff  = 1.81818 10^(-5);

(* specify set of dimensionless boundary fluxes to solve ODE for,
   negative for evaporation, positive for infiltration *)
xi     = {-6.8 10^(-6), -5.0 10^(-6), -2.5 10^(-6),
          10^(-5), 10^(-4), 10^(-3), 10^(-2), 0.026, 0.1, 0.3};

(* calculate dimensional boundary flux *)
jw     = xi k0eff;

(* numerically solve ODE in depth interval [-2 m, 0] for set of boundary
   fluxes and groundwater table at z=0 *)
sol    = Table[
            NDSolve[{h'[z] == 1 - jw[[i]]/k[h[z], z], h[0] == 0},
                h, {z, 0, -2}
            ],
            {i, 10}
         ];

(* plot the resulting matric heads... *)
Plot[Evaluate[h[z] /. sol], {z, 0, -2}, PlotRange -> All]

(* ...and export them to a file for later processing *)
Export["stat-flow-layered-sl.dat",
    Table[Flatten[{z, Flatten[Evaluate[h[z] /. sol]]}], {z, 0, -2, -0.05}]
]

(* clean up *)
ClearAll[jw, xi, k0eff, p1, p2, alpha1, n1, a1, k01, alpha2, n2, a2, k02]
```

Figure C.1. *Mathematica* code for calculating the matric head $h_m(z)$ for stationary flow in layered medium. The graphical output is shown in Figure C.2. The exported data were used to generate Figure 6.9 on page 149.

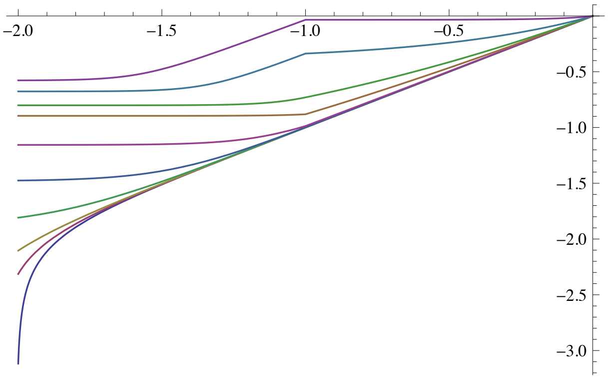


Figure C.2. Matric head $h_m(z)$ for various stationary water fluxes in layered medium obtained with *Mathematica* code shown in Figure C.1.

Finally assume a constant water table at depth $z = 0$ and choose the same values for the dimensionless fluxes $\xi := j_w/K_0^{\text{eff}}$ as those in Figure 6.9 on page 149.

At the heart of the simulation is the solver `NDSolve` which produces the numerical solution for the ODE with the given boundary condition $h_m(0) = 0$ and for the interval $[-2\text{ m}, 0]$. A somewhat compact notation was used in that the list `{xi}` is used to hand down the entire set of fluxes to the solver instead of looping over its elements individually. The result, stored in `sol` for later use, consists of a set of interpolating functions which can be evaluated for any $z \in [-2, 0]$. This is done by the operator `Evaluate` for plotting as well as for exporting the data. Looking at the command

```
Plot[Evaluate[h[z] /. sol], {z, 0, -2}, PlotRange -> All]
```

in some more detail, it requests to evaluate the interpolating functions `h[z]` contained in solution `sol` and to plot them in the interval $[-2, 0]$ with the plot range chosen such that all values are shown (Figure C.2).

Further Suggestions

- Change the order of materials such that silt is the top layer (see Figure 6.7 on page 147).
- Experiment with different thicknesses, different material parameters and parameterizations, and add a third layer.
- What happens if the requested fluxes exceed their limits, i.e., too high an infiltration or evaporation?

C.2

Simple Atmospheric Forcing of Soil Water Flow

Soil water flow is primarily driven at the soil surface by precipitation and evaporation. We represent this forcing in a simplified manner this forcing by a spatially uniform flux j_{rain} for precipitation and a spatially uniform soil matric head $h_{m_{\mathrm{surf}}} < 0$ for evaporation. Thereby, we neglect all the complications that arise from (i) the coupling between water and energy fluxes through the latent heat of evaporation, (ii) the vegetation which extracts water from deeper soil layers through their roots and transport it higher into the land-atmospheric boundary layer through their leaves, and (iii) the coupling between soil and atmosphere by the water flux which lowers/raises the chemical potential of water in the atmosphere, thereby modifying the driving force all the way to switching from precipitation to evaporation and *vice versa*.

Model Formulation

We represent the dynamics of water flow by Richards' equation (6.2). For the soil architecture, we focus on situations with a horizontal soil surface and consider (i) a uniform medium, (ii) horizontal layers, and (iii) heterogeneous media, where the simplified Mualem-van Genuchten parameterization (3.45) and (3.56) with parameters given in Table 3.1 describe the hydraulic properties of the constituting materials. Apparently, the first two architectures may be studied with one-dimensional simulations, while the third one requires an at least two-dimensional representation.

Numerical Simulation

1. Use Figure 6.7 and Figure 6.9 as guides for the construction of the one-dimensional representation of the uniform and of the horizontally layered media as well as for the upper boundary condition. Choose z to point downwards with $z = 0$ at the depth of the water table. As lower boundary condition choose a constant water table and as initial condition hydrostatic equilibrium, i.e., $h_m(z) = z$.

2. Prescribe a constant upper boundary condition (infiltration or evaporation) which is turned on at time $t = 0$ and observe/discuss the transition from the static to the dynamic equilibrium. Address in particular:

   - the formation and propagation of an infiltration front depending on soil hydraulic properties and infiltration flux,
   - the formation of an evaporation zone, again depending on material properties and matric potential,

- the effect of a thin coarse-textured layer at the soil surface for both infiltration and evaporation,
- for layered media, the behavior of θ and h_m at the interface.

It may be useful to represent the simulation results – $\theta(z;t)$, $h_m(z;t)$, and $j_w(t;z)$ – as contours in a space-time plot.

Further Suggestions

1. Generate a two-dimensional heterogeneous medium, either using an appropriate random number generator or by drawing it by hand, and explore its phenomenology along the same lines as above.
2. Calculate and discuss horizontal averages of θ and h_m.
3. Explore materials with other hydraulic parameters.

C.3
Fluctuating Water Table

A fluctuating water table may originate from a phreatic aquifer that is strongly influenced by some surface waters like a river or by ocean tides, or by the operation of pumping wells. It may also result from agricultural operations like irrigation in regions with less permeable soil layers at greater depths. Possible effects of such fluctuations range from surface salinization and, conversely, drying up vegetation, to modifications of microbial activities like soil carbon decomposition and denitrification, and further to modifications of gaseous environmental tracers like noble gases.

Model Formulation

We recall from Section 6.1 that the dynamics of the capillary fringe immediately above the water table is beyond the Richards regime and from Section 3.4 that the hysteresis of the soil water characteristic is most pronounced near saturation. While both issues are addressed in current research, they are not yet understood firmly and quantitatively. Modeling thus demands some bold assumptions and we will assume the same formulation for the dynamics of the water phase as in Section C.2, Richards equation, and also consider the same architectures. This obviously precludes an accurate simulation of the capillary fringe, but it will still lead to insight into the dynamics of the water phase above the capillary fringe.

Numerical Simulation

Further Suggestions

C.4
Solute Transport with Stationary Water Flow

Solutes originate from a multitude of sources. They are deposited at the soil surface as agrochemicals, dusts, and chemicals dissolved in rainwater and fog, they are present in groundwater, or they are released by biogeochemical processes within soils and aquifers. Their transport is primarily driven by the water flow – neglect transport paths through the soil air which is important for a number of substances –, but interrupted at the soil surface where water evaporates, leaving the solutes behind. Focussing on transport proper, we only consider stationary flows and conservative solutes.

Model Formulation

Consider the media and the stationary flow regimes explored in Section C.2 and let solute transport be described locally by the convection-dispersion equation (4.53) with the effective dispersion coefficient given by the sum of molecular diffusion, parameterized by one of the Millington-Quirk models (4.55), and of hydromechanic dispersion represented by the Scheidegger tensor (4.56). If values for the required dispersivities are not available, choose $\lambda_\ell = \ell/100$, where ℓ is the length of the flow domain, in the direction of the main flow, and choose $\lambda_t = \lambda_\ell$. Notice that for stationary flow, all parameters – θ and $\mathbf{v}$, with this also D_{eff} – are constant in time, but not necessarily uniform in space.

Numerical Simulation: Uniform Media

1. Consider a one-dimensional situation and observe the evolution of a solute concentration that is highly localized at time $t = 0$, ideally of the form $\delta(z)$,
    - at the soil surface, when water infiltrates,
    - at the water table, for evaporation.
2. As functions of time, calculate the first two spatial moments

$$\begin{aligned} \langle z\rangle &:= \int_0^\ell z p_z(z;t)\,\mathrm{d}z \ , \\ \operatorname{var}(z) &:= \int_0^\ell \big[z - \langle z\rangle\big]^2 p_z(z;t)\,\mathrm{d}z \end{aligned}$$

with

$$p_z(z;t) := \frac{C_t(z;t)}{m_0} \ , \quad m_0 = \int_0^\ell C_t(z;t)\,\mathrm{d}z \ ,$$

where $\ell < 0$ is the position of the soil surface and $p_z(z;t)$ is the travel distance pdf. Use the *method of moments* to quantify the transport and calculate the corresponding apparent transport parameters defined through (4.26) and (4.27), respectively. Specifically, address:

- for a uniform medium and the region with gravity flow, the agreement between expected and simulated values of the transport parameters and their dependence on the flow velocity,
- the qualitative changes of the pulse form as the respective outflow boundary is approached,
- the qualitative changes of the pulse form at a layer interface of a multi-layer soil, paying particular attention to the differences between C_t and C_w.

3. Consider a soil that is initially free of solutes, with constant evaporation, and with the constant solute concentration $C_{w_0} > 0$ in the groundwater. Predict, at least qualitatively, the evolution of the concentration profile $C_t(z;t)$ and then observe it in the numerical simulation. How does $C_t(z;t)$ depend on (i) the extent of the flow domain and (ii) on the hydraulic parameters.

Numerical Simulation: Heterogeneous Media

For the following, consider two different materials, e.g., the sand and the silt used before, to construct heterogeneous architectures. Study solute transport (i) with gravity flow and (ii) with a constant water table at the lower boundary for the following geometries:

1. Two vertical slabs with thicknesses d_1 and d_2, respectively, as a generalization of the MIM model. Assume no-flow at the outer boundaries.

2. Two horizontal slabs, again with thicknesses d_1 and d_2 and with vertical flow.

3. Random distribution with correlation lengths λ_h and λ_v in the horizontal and vertical, respectively.

In particular calculate D_{eff}/D_m as a function of the Peclet number Pe for the gravity flow regime and plot it as a function of Pe and, alternatively, of θ.

Choosing Miller- or Warrick-scaling and continuous distributions of scaling factors, the architectures used above may be further generalized. It is instructive to study to what extent this modifies the results obtained from the discrete architectures.

Further Suggestions

1. Generate a two-dimensional heterogeneous medium with vertical correlation length $\lambda_z = \ell/100$, where ℓ is the height of the simulated flow domain, and horizontal correlation length $\lambda_x = 10\lambda_z$. Assume the vertical boundaries to be periodic, i.e., what flows out at one end flows in at the other. Calculate the horizontally averaged concentration profile
$$\overline{C}_t(z;t) := \int_0^{\ell_x} C_t(x,z;t)\,\mathrm{d}x \ ,$$
where ℓ_x is the horizontal extent of the flow domain, and address some of the issues already raised above. In particular observe the shape of the concentration profile and its dependence on the flow velocity.

2. Generate a two-dimensional medium that is uniform except for a circular inclusion with radius r_0 at z_i. Again choose periodic boundaries for the vertical.
   - Simulate various stationary flow fields – infiltration and evaporation with different fluxes – and observe how the hydraulic character of the inclusion changes. You may want to play with the dimensionless numbers r_0/ℓ_x and αz_i, where α is the scaling parameter in the van Genuchten parameterization (3.44) of the embedding material.
   - Study solute transport along the lines explored above, in particular illuminate the definition of apparent/effective transport parameters as functions of v, D_{eff}, r_0, and z_i.

C.5

Flow and Transport under Irrigated Field

The production of food and biomass for energy production in semi-arid and arid regions is an important field of application for soil physics. A key issue is the optimal control of salinization with the minimal consumption of water. As a starter for this very complicated problem, we neglect all the really difficult issues like plants and their development or the dynamics of the soil, which includes cracking and formation of surface crusts, and only consider (i) sequences of irrigations followed by evaporation and (ii) the associated transport of solutes. As the only complicating factor, we look at the impact of soil heterogeneity.

Model Formulation

We envisage a moderately permeable silty soil that is managed through flood irrigation (see for instance front picture of these lecture notes). We then assume flow and transport to be described locally by the Richards and the convection-dispersion equation, respectively, as was done already in Sections C.2 and C.4. For simplicity, disregard hysteresis and the fact that Richards equation is not applicable near saturation. For the soil, consider a uniform soil and the two architectures described und "Further Suggestions" in Section C.4.

Hydraulic forcing of the system is by cycles of (i) an irrigation event, modeled with $h_m = 0.02$ m at the soil surface for a duration that leads to a mean infiltration of some 3 mm/d, and (ii) an evaporation period, modeled with $h_m = -10$ m for the remainder of the cycle. Treat the cycle time T as a variable, which would eventually be optimized but choose $T = 2$ d for a start. At the lower boundary, let $h_m = 0$. For the two-dimensional simulations, the vertical boundaries are periodic.

For studying transport, let the mass density m_0 be uniform at the soil surface at time $t = 0$, hence the vertical profile of the form $\delta(x)$. Let the upper and lower boundaries be impermeable for solute, with the vertical boundaries again periodic. We thus study the fluctuation and redistribution of the initially present mass, which is conserved over time.

For quantitative analyses, we again employ the method of moments on the horizontally averaged total concentration $\overline{C}_t(z;t)$.

Numerical Simulation

Choose two materials to be employed – recall that the embedding material must be moderately conductive, else flood irrigation could not be used (why not?) – and roughly estimate the irrigation schedule.

1. Observe the evolution of the solute pulse through the irrigation cycles and compare the different architectures. Play with the various system parameters to understand their influence.
2. Calculate $\langle z \rangle$ and $\mathrm{var}(z)$ as functions of time.

Further Suggestions

1. Calculate the autocovariance function of the total concentration and the crosscovariance between total concentration and material as they evolve in time.

E
Some Constants and Material Properties

E.1
General Constants

Name	Symbol	Value
Avogadro constant	N_A	$6.0221 \cdot 10^{23}$ mol^{-1}
Boltzmann constant	k	$1.3807 \cdot 10^{-23}$ $\mathrm{J\,K}^{-1}$
universal gas constant $N_A k$	R	8.3144 $\mathrm{J\,mol}^{-1}\mathrm{K}^{-1}$
acceleration of gravity on Earth	g	9.81 $\mathrm{m\,s}^{-2}$

E.2
Material Properties of Water

Density ρ_w, vapor pressure p^{vapor}, dynamic viscosity μ_w, surface tension water-air σ_{wa}, and heat capacity C_p as functions of temperature T at standard pressure $p_0 = 10^5$ Pa. Data from *Lide* [1994], page **6**-10.

T °C	ρ_w $\mathrm{kg\,m}^{-3}$	p^{vapor} kPa	μ_w 10^{-3} Pa s	σ_{wa} $\mathrm{J\,m}^{-2}$	C_p $\mathrm{J\,kg}^{-1}\,\mathrm{K}^{-1}$
0	999.84	0.6113	1.793	0.07564	4218
10	999.70	1.2281	1.307	0.07423	4192
20	998.21	2.3388	1.002	0.07275	4182
30	995.65	4.2455	0.798	0.07120	4178
40	992.22	7.3814	0.653	0.06960	4179
50	988.03	12.344	0.547	0.06794	4181
60	983.20	19.932	0.467	0.06624	4184
70	977.78	31.176	0.404	0.06447	4190
80	971.82	47.373	0.354	0.06267	4196
90	965.35	70.117	0.315	0.06082	4205
100	958.40	101.325	0.282	0.05891	4216

Other material properties of water

Name	Symbol	Value	Unit
enthalpy of fusion (0°C)	H_{sl}	0.333	$\mathrm{MJ\,kg^{-1}}$
enthalpy of vaporisation (0°C)	H_{lv}	2.503	$\mathrm{MJ\,kg^{-1}}$
enthalpy of vaporisation (100°C)	H_{lv}	2.259	$\mathrm{MJ\,kg^{-1}}$

E.3
Material Properties of Air

Name	Symbol	Value	Unit
density (dry air, 0°C, 101’325 Pa)	ρ_a	1.293	$\mathrm{kg\,m^{-3}}$
dynamic viscosity (10°C)	μ_a	$1.77 \cdot 10^{-5}$	$\mathrm{Pa\,s}$

E.4
Properties of Geologic Materials

Name	Symbol	Value	Unit
quartz minerals ($\mathrm{SiO_2}$)			
density	ρ_m	$2.65 \cdot 10^3$	$\mathrm{kg\,m^{-3}}$
heat capacity (300 K)	C	733	$\mathrm{J\,kg^{-1}\,K^{-1}}$

Dictionary English-Deutsch and Glossary

A

abstraction Entnahme
of groundwater

AD nach Christus
anno Domini

advection Advektion
flow of something from one region to another, e.g., of groundwater from a high- to a low-pressure region; ↗ convection

alluvial alluvial (geol.), angeschwemmt
related to flowing surface water, e.g., alluvial deposits or alluvial fan

anomaly Anomalie, Abweichung
deviation from some reference which may be a scalar, vector, or function

aquiclude Grundwasserstauer
less permeable than aquitard

aquifer Grundwasserleiter

aquitard Grundwasserstauer
more permeable than aquiclude

B

BC vor Christus
before Christ

C

capillary fringe Kapillarsaum
in *sensu stricto* the water-saturated zone above a water table, the so-called *satiated* zone where the matric potential is already negative but the largest pore not yet air-filled; in *sensu lato* the zone above a water table where the air phase is not yet continuous

clay Ton
small-sized grain size fraction of soils; depending on classification, the size fraction ranges up to about 2 μm; ↗ silt, ↗ sand

conductivity Leitfähigkeit

confined aquifer gespannter Grundwasserleiter

convection Konvektion
transport of something *with* something else, e.g., of solutes or heat with groundwater flow; ↗ advection

core drilling Kernbohrung
drilling that yields undisturbed core

D

diagenetic diagenetisch (geol.)
physical, chemical, or biological alteration of sediments into sedimentary rock at relatively low temperatures and pressures

drainage Entwässerung

F

facies Fazies (geol.)
sum of characteristics of a rock, sedimenatry, or other unit

fluid Fluid (die umgangssprachliche "Flüssigkeit" ist ein Beispiel eines Fluids, deckt aber die Bedeutung bei weitem nicht vollständig ab)

forcing (externer) Antrieb

G

groundwater Grundwasser

H

head h Höhe
: height of water column equivalent to certain potential: matric head, hydraulic head,...

I

igneous magmatisch (geol.)
: originating from molten rock; opposed to sedimentary

imbibition Benetzung

irrigation Bewässerung

L

loam Lehm
: soil consisting of about 40% clay, 40% silt, and 20% sand

M

MAGT (mean annual ground temperature) mittlere Jahrestemperatur des Bodens

metamorphic metamorph (geol.)
: rock alterated by high temperatures or pressures, higher than with diagenetic alterations, or by contact with different fluids

momentum Impuls

P

parameterization Parametrisierung
: (i) class of functions to approximate the relation between variables, e.g., the van Genuchten parameterization for the relation between water content and matric head; (ii) a macroscale model for microscale processes that are not represented in detail, e.g., Darcy's flux law at the continuum scale for Stokes flow at the pore scale

permeability k Durchlässigkeit
: tensorial quantity that describes the geometry of a porous medium with respect to fluid flow, scalar in isotropic media; not to be confounded with conductivity $\mathsf{K} = \mathsf{k}/\mu$, where μ is the dynamic viscosity

piezometer Piezometer
: tube inserted into the ground to measure height of watertable, the piezometric head

R

residual Rest
: deviation from some model or exact value (math)

S

sand Sand
: large-sized grain size fraction of soils; depending on classification, the size fraction ranges from about 63 μm to 2 mm; still larger fractions are referred to as gravel; ↗ clay, ↗ silt

silt Schluff (Staub)
: medium-sized grain size fraction of soils; depending on classification, the size fraction ranges from about 2 μm to about 63 μm; ↗ clay, ↗ sand

soil Boden

stationary stationär
: (i) dynamics: all partial derivatives with respect to time vanish, (ii) statistics: all statistical moments are translation invariant

structure Struktur
: features that must be represented explicitly and in detail in order to obtain a faithful description; ↗ texture

T

table Spiegel, Oberkante
: for groundwater or permafrost

transmissivity Transmissivität
: vertically integrated hydraulic conductivity

texture Textur

features, for which a statistical representation suffices, whose contribution to a process of interest is parameterized; ↗ structure

U

unconfined aquifer freier Grundwasserleiter

V

vadose zone vadose, wasserungesättigte Zone

W

water table Grundwasserspiegel

Bibliography

Abramowitz, M. and I. A. Stegun, 1970: *Handbook of Mathematical Functions*, Dover Publishing Co., New York.

Arfken, G. B. and H. J. Weber, 1995: *Mathematical Methods for Physicists*, Academic Press, New York, 4. edition.

Aris, R., 1956: On the dispersion of solute in a fluid flowing through a tube, *Proc. R. Soc. London*, **235A**, 67–77.

Arya, L. M. and J. F. Paris, 1981: A physicoempirical model to predict the soil moisture characteristic from particle-size distribution and bulk density data, *Soil Sci. Soc. Am. J.*, **45**, 1023–1030.

Bayer, A., H.-J. Vogel, O. Ippisch and K. Roth, 2005: Do effective properties for unsaturated weakly layered porous media exist? An experimental study, *Hydrology and Earth System Sciences*, **9**, 517–522, SRef–ID: 1607–7938/hess/2005–9–517.

Bear, J., 1961: On the tensor form of dispersion in porous media, *J. Geophys. Res.*, **66**, (4), 1185–1197.

Bear, J., 1972: *Dynamics of Fluids in Porous Media*, Elsevier, New York.

Bird, R. B., W. E. Stewart and E. N. Lightfoot, 1960: *Transport Phenomena*, John Wiley & Sons, New York.

Bittelli, M., M. Flury and G. S. Campbell, 2003: A thermodielectric analyzer to measure the freezing and moisture characteristic of porous media, *Water Resour. Res.*, **39**, (2), 1041, doi:10.1029/2001WR000930.

Blackman, E. G. and G. B. Field, 2003: A new approach to turbulent transport of a mean scalar, *Phys. Fluids*, **15**, L73, doi: 10.1063/1.1613281.

Bonn, D., J. Eggers, J. Indekeu, J. Meunier and E. Rolley, 2009: Wetting and spreading, *Rev. Mod. Phys.*, **81**, (2), 739–805, doi: 10.1103/RevModPhys.81.739.

Brooks, R. H. and A. T. Corey, 1966: Properties of porous media affecting fluid flow, *J. Irrigation and Drainage Div., Proc. Am. Soc. Civil Eng. (IR2)*, **92**, 61–88.

Brown, R., 1828: On the existence of active molecules in organic and inorganic bodies, *Phil. Mag.*, **4**, 162–173.

Brown, E., N. A. Forman, C. S. Orellana, H. Zhang, B. W. Maynor, D. E. Betts, J. M. DeSimone and H. M. Jaeger, 2010: Generality of shear thickening in dense suspensions, *Nature Materials*, **9**, 220–224, doi: 10.1038/NMAT2627.

Buck, A. L., 1981: New equations for computing vapor pressure and enhancement factor, *J. Appl. Meteor.*, **20**, (12), 1527–1532.

Buckingham, E., 1907: Studies on the movement of soil moisture, Bulletin 38, U.S. Department of Agriculture, Bureau of Soils, Washington, DC.

Burdine, N. T., 1953: Relative permeability calculations from pore-size distribution data, *Petr. Trans., Am. Inst. Mining Metall. Eng.*, **198**, 71–77.

Burgess, W. G., M. A. Hoque, H. A. Michael, C. I. Voss, G. N. Breit and K. M. Ahmed, 2010: Vulnerability of deep groundwater in the Bengal aquifer system to contamination by arsenic, *Nature Geoscience*, **3**, 83–87, doi: 10.1038/ngeo750.

Campbell, G. S., J. D. Jungbauer, W. R. Bidlake and R. D. Hungerford, 1994: Predicting the effect of temperature on soil thermal conductivity, *Soil Sci.*, **158**, (5), 307–313.

Carslaw, H. S. and J. C. Jaeger, 1990: *Conduction of Heat in Solids*, Oxford University Press, Oxford, 2. revised edition.

Cary, J. W. and H. F. Mayland, 1972: Salt and water movement in unsaturated frozen soil, *Soil Sci. Soc. Am. Proc.*, **36**, (4), 549–555.

Cass, A., G. S. Campbell and T. L. Jones, 1984: Enhancement of thermal water vapor diffusion in soil, *Soil Sci. Soc. Am. J.*, **48**, 25–32.

Childs, E. C. and N. Collis-George, 1950: The permeability of porous materials, *Proc. R. Soc. London*, **201A**, 392–405.

Cho, C. H., J. Urquidi, S. Singh and G. W. Robinson, 1999: Thermal offset viscosities of liquid H_2O, D_2O, and T_2O, *J. Phys. Chem. B*, **103**, (11), 1991–1994.

Churkina, G. and S. W. Running, 1998: Contrasting climatic controls on the estimated productivity of global terrestrial biomes, *Ecosystems*, **1**, 206–215.

Coats, K. H. and B. D. Smith, 1964: Dead-end pore volume and dispersion in porous media, *Soc. Petr. Eng. J.*, **4**, 73–84.

Cornelis, W. M., J. Ronsyn, M. V. Meirvenne and R. Hartmann, 2001: Evaluation of pedotransfer functions for predicting the soil moisture retention curve, *Soil Sci. Soc. Am. J.*, **65**, (3), 638–648, doi: 10.2136/sssaj2001.653638x.

Costanza-Robinson, M. S., K. H. Harrold and R. M. Lieb-Lappen, 2008: X-ray microtomography determination of air-water interfacial area-water saturation relationships in sandy porous media, *Environ. Sci. Technol.*, **42**, 2949–2956, doi: 10.1021/es072080d.

Cousin, I., P. Levitz and A. Bruand, 1996: Three-dimensional analysis of a loamy-clay soil using pore and solid chord distributions, *Europ. J. Soil Sci.*, **47**, (4), 439–452.

Crank, J., 1975: *The Mathematics of Diffusion*, Oxford University Press, Oxford, 2. edition.

Cressie, N., 1993: *Statistics for Spatial Data*, John Wiley & Sons.

Dagan, G., 1984: Solute transport in heterogeneous porous formations, *J. Fluid Mech.*, **145**, 151–177.

Dagan, G., 1989: *Flow and Transport in Porous Formations*, Springer Verlag, Berlin.

Darcy, H., 1856: *Les Fontaines de la Ville de Dijon*, Dalmont, Paris.

de Gennes, P. G., 1985: Wetting: statics and dynamics, *Rev. Mod. Phys.*, **57**, (3), 827–863.

Dentz, M., H. Kinzelbach, S. Attinger and W. Kinzelbach, 2000: Temporal behavior of a solute cloud in a heterogeneous porous medium 1. point-like injection, *Water Resour. Res.*, **36**, (12), 3591–3604.

Deutsch, C. V. and A. Journel, 1998: *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, New York, 2. edition.

de Vries, D. A., 1963: Thermal properties of soils, in *Physics of Plant Environments*, edited by W. R. van Wijk, 210–235, North-Holland.

de Vries, D. A., 1975: Heat transfer in soils, in *Heat and Mass Transfer in the Biosphere. Part 1: Transfer Processes in the Plant Environment*, edited by D. A. de Vries and N. H. Afgan, 5–28, Wiley&Sons, New York.

DiCarlo, D. A., 2004: Experimental measurements of saturation overshoot on infiltration, *Water Resour. Res.*, **40**, W04215, doi:10.1029/2003WR002670.

Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki and N. Hanasaki, 2006: Gswp-2: Multimodel analysis and implications for our perception of the land surface, *Bull. Am. Met. Soc.*, **87**, (10), 1381–1397, doi: 10.1175/BAMS–87–10–1381.

Dussan, E. B., 1979: On the spreading of liquids on solid surfaces: static and dynamic contact lines, *Annu. Rev. Fluid Mech.*, **11**, 371–400.

Eching, S. O., J. W. Hopmans and O. Wendroth, 1994: Unsaturated hydraulic conductivity from transient multistep outflow and soil water pressure data, *Soil Sci. Soc. Am. J.*, **58**, 687–695.

Einstein, A., 1905: Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen, *Ann. der Physik*, **17**, 549–560.

Eliassi, M. and R. J. Glass, 2001: On the continuum-scale modeling of gravity-driven fingers in unsaturated porous media: The inadequacy of the Richards equation with standard monotonic constitutive relations and hysteretic equations of state, *Water Resour. Res.*, **37**, (8), 2019–2035.

Eloy, C., 2011: Optimal Strouhal number for swimming animals, *arXiv e-prints, physics.flu-dyn*, 1102.0223.

EPICA, 2004: Eight glacial cycles from an antarctic ice core, *Nature*, **429**, 623–628, doi:10.1038/nature02599.

European Environment Agency, 2007: *Europe's Environment: The Forth Assessment*, Environmental Assessment Report, `http://www.eea.europa.eu/publications/state_of_environment_report_2007_1`, Office for Official Publications of the European Communities.

European Environment Agency, 2010: *The European Environment: State and Outlook 2010 – Freshwater Quality*, Publications Office of the European Union, `http://www.eea.europa.eu/soer`, doi: 10.2800/60214.

Falkenmark, M. and J. Rockström, 2006: The new blue and green water paradigm: Breaking new ground for water resources planning and management, *J. Water Resour. Plng. and Mgmt*, **132**, (3), 129–132.

Farouki, O. T., 1981: The thermal properties of soils in cold regions, *Cold Reg. Sci. Technol.*, **5**, (1), 67–75, doi:10.1016/0165–232X(81)90041–0.

Fick, A., 1855: Über Diffusion, *Poggendorff's Annalen der Physik*, **94**, 59–86.

Fisher, L. R., R. A. Gamble and J. Middlehurst, 1981: The Kelvin equation and the capillary condensation of water, *Nature*, **290**, 575–576, doi: 10.1038/290575a0.

Flury, M., H. Flühler, W. A. Jury and J. Leuenberger, 1994: Susceptibility of soils to preferential flow of water: A field study, *Water Resour. Res.*, **30**, (7), 1945–1954.

Forchheimer, P., 1901: Wasserbewegung durch Boden, *Z. Ver. Deutsch. Ing.*, **45**, 1782–1888.

Fredlund, M. D., G. W. Wilson and D. G. Fredlund, 2002: Use of the grain-size distribution for estimation of the soil-water characteristic curve, *Can. Geotech. J.*, **39**, (5), 1103–1117, doi: 10.1139/t02–049.

Frisch, U., 1995: *Turbulence*, Cambridge University Press, Cambridge, UK.

Fu, G., S. Chen, C. Liu and D. Shepard, 2004: Hydro-climatic trends of the Yellow River basin for the last 50 years, *Climate Change*, **65**, (1-2), 149–178, doi: 10.1023/B:CLIM.0000037491.95395.bb.

Gelhar, L. W. and C. L. Axness, 1983: Three-dimensional stochastic analysis of macrodispersion in aquifers, *Water Resour. Res.*, **19**, 161–180, doi:10.1029/WR019i001p00161.

Gelhar, L. W., C. Welty and K. R. Rehfeldt, 1992: A critical review of data on field-scale dispersion in aquifers, *Water Resour. Res.*, **28**, (7), 1955–1974.

Glass, R. J. and M. J. Nicholl, 1996: Physics of gravity fingering of immiscible fluids within porous media: An overview of current understanding and selected complicating factors, *Geoderma*, **70**, (2-4), 133–163.

Goldenfeld, N. and L. P. Kadanoff, 1999: Simple lessons from complexity, *Science*, **284**, 87–89.

Good, R. J., 1992: Contact angle, wetting, and adhesion: a critical review, *J. Adhesion Sci. Technol.*, **6**, (12), 1269–1302, doi: 10.1163/156856192X00629.

Gray, W. G. and S. M. Hassanizadeh, 1998: Macroscale continuum mechanics for multiphase porous-media flow including phases, interfaces, common lines and common points, *Adv. Water Res.*, **21**, (4), 261–281.

Gruber, N. and J. N. Galloway, 2008: An Earth-system perspective of the global nitrogen cycle, *Nature*, **451**, 293–296, doi: 10.1038/nature06592.

Haines, W. B., 1930: Studies in the physical properties of soils: V. The hysteresis effect in capillary properties and the modes of moisture distribution associated therewith, *J. Agric. Sci.*, **20**, 97–116.

Hardy, C., 1977: A grain boundary groove measurement of the surface tension between ice and water, *Phil. Mag.*, **35**, (2), 471–484, doi: 10.1080/14786437708237066.

Helmig, R., 1997: *Multiphase Flow and Transport Processes in the Subsurface*, Springer Verlag, Berlin.

HGK, 1999: Hydrogeologische Kartierung und Grundwasserbewirtschaftung Rhein-Neckar-Raum, technical report, Ministerium für Umwelt und Verkehr Baden-Württemberg, Hessisches Ministerium für Umwelt, Landwirtschaft und Forsten Ministerium für Umwelt und Forsten Rheinland-Pfalz.

Ho, C. K. and S. W. Webb, 1996: *A Review of Porous Media Enhanced Vapor-Phase Diffusion Mechanisms, Models, and Data – Does Enhanced Vapor-Phase Diffusion Exist?*, SAND96-1198, UC-403, Geohydrology Department, Sandia National Laboratories, Albuquerque, NM 87185-1324.

Hoekstra, A. Y. and P. Q. Hung, 2002: Virtual water trade: A quantification of virtual water flow between nations in relation to international crop trade, Research report series no. 11, IHE Delft.

Homsy, G. M., 1987: Viscous fingering in porous media, *Ann. Rev. Fluid Mech.*, **19**, 271–311, doi:10.1146/annurev.fl.19.010187.001415.

Huang, J. and L. S. Bartell, 1995: Kinetics of homogeneous nucleation in the freezing of large water clusters, *J. Phys. Chem.*, **99**, (12), 3924–3931, doi: 10.1021/j100012a010.

Huyakorn, P. S. and G. F. Pinder, 1983: *Computational Methods in Subsurface Flow*, Academic Press, San Diego.

Ippisch, O., H.-J. Vogel and P. Bastian, 2006: Validity limits for the van Genuchten-Mualem model and implications for parameter estimation and numerical simulation, *Adv. Water Res.*, **29**, 1780–1789, doi:10.1016/j.advwatres.2005.12.011.

Jenny, H., 1941: *Factors of Soil Formation – A System of Quantitative Pedology*, McGraw-Hill, New York; reprinted by Dover Publications, 1994.

Jensen, H. J., 1998: *Self-Organized Criticality*, vol. 10 of *Cambridge Lecture Notes in Physics*, Cambridge University Press, Cambridge, UK.

Jerison, E. R., Y. Xu, L. A. Wilen and E. R. Dufresne, 2011: Deformation of an elastic substrate by a three-phase contact line, *Phys. Rev. Lett.*, **106**, 186103, doi: 10.1103/PhysRevLett.106.186103.

Jin, Y. and W. A. Jury, 1996: Characterizing the dependence of gas diffusion coefficient on soil properties, *Soil Sci. Soc. Am. J.*, **60**, 66–71.

Johannessen, O. M., L. Bengtsson, M. W. Miles, S. I. Kuzmina, V. A. Semenov, G. V. Alekseev, A. P. Nagurnyi, V. F. Zakharov, L. P. Bobylev, L. H. Pettersson, K. Hasselmann and H. P. Cattle, 2004: Arctic climate change: observed and modelled temperature and sea-ice variability, *Tellus A*, **56**, (4), 328–341, doi:10.1111/j.1600–0870.2004.00060.x.

Johansen, O., 1977: Thermal conduction of soils, Draft translation 637, Norwegian Technical University, Cold Regions Research and Engineering Laboratory (CRREL), Hanover, New Hampshire.

Jury, W. A. and K. Roth, 1990: *Transfer Functions and Solute Movement through Soils. Theory and Applications*, Birkhäuser Verlag, Basel, Switzerland.

Kandhai, D., D. Hlushkou, A. G. Hoekstra, P. M. A. Sloot, H. V. As, and U. Tallarek, 2002: Influence of stagnant zones on transient and asymptotic dispersion in macroscopically homogeneous porous media, *Phys. Rev. Lett.*, **88**, (23), 234501 1–4, doi:10.1103/PhysRevLett.88.234501.

Kersting, A. B., D. W. Efurd, D. L. Finnegan, D. J. Rokop, D. K. Smith and J. L. Thompson, 1999: Migration of plutonium in ground water at the Nevada Test Site, *Nature*, **397**, (6714), doi:10.1038/16231, 56–59.

Kessler, M. A. and B. T. Werner, 2003: Self-organization of sorted patterned ground, *Science*, **299**, 380–383.

Kiehl, J. T. and K. E. Trenberth, 1997: Earth's annual global mean energy budget, *Bull. Amer. Meteor. Soc.*, **78**, (2), 197–208.

Kirkpatrick, R. J., A. G. Kalinichev and J. Wang, 2005: Molecular dynamics modelling of hydrated mineral interlayers and surfaces: structure and dynamics, *Mineralogical Magazine*, **69**, (3), 289–308.

Kolmogorov, A. N., 1941: The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers, *Dokl. Akad. Nauk SSSR*, **30**, (4), translation 1991 in *Proc. R. Soc. Lond. A*, **434**, 9–13.

Kool, J. B. and J. C. Parker, 1987: Development and evaluation of closed-form expressions for hysteretic soil hydraulic properties, *Water Resour. Res.*, **23**, 105–114, doi:10.1029/WR023i001p00105.

Kool, J. B., J. C. Parker and M. T. van Genuchten, 1985: Determinig soil hydraulic properties from one-step outflow experiments by parameter estimation: I. Theory and numerical studies, *Soil Sci. Soc. Am. J.*, **49**, 1348–1354.

Kozeny, J., 1927: Über die kapillare Leitung des Wassers im Boden, *Berichte Wiener Akademie*, **136A**, 271–306.

Kreft, A. and A. Zuber, 1978: On the physical meaning of the dispersion equation and its solutions for different initial and boundary conditions, *Chem. Eng. Sci.*, **33**, 1471–1480.

Kung, K.-J. S., 1990*a*: Preferential flow in a sandy vadose zone: 1. Field observation, *Geoderma*, **46**, 51–58.

Kung, K.-J. S., 1990*b*: Preferential flow in a sandy vadose zone: 2. Mechanisms and implications, *Geoderma*, **46**, 59–71.

Landau, L. D. and E. M. Lifschitz, 1981: *Lehrbuch der theoretischen Physik. Band VI: Hydrodynamik*, Akademie Verlag, Berlin, 4. Auflage.

Landau, L. D. and E. M. Lifschitz, 1984: *Lehrbuch der theoretischen Physik. Band V: Statistische Physik, Teil 1*, Akademie Verlag, Berlin, 6. Auflage.

Lawes, J. B., J. H. Gilbert and R. Warington, 1882: On the amount and composition of the rain and drainage waters collected at Rothamsted. III. The drainage water from land cropped and manured, *J. Roy. Agric. Soc. Eng.*, **18**, 1–71.

Levine, D., 1996: Users guide to the PGAPack parallel genetic algorithm library, technical report 95/18, Argonne National Laboratory, 9700 South Cass Avenue, Argonne Il 60439.

Lide, D. R. (Ed.), 1994: *CRC Handbook of Chemistry and Physics*, CRC Press, Boca Raton, FL, 75. edition.

Lord Rayleigh, 1915: Æolian tones, *Phil. Mag.*, **29**, (172), 433–444, doi: 10.1080/14786440408635325.

Luckner, L., M. T. van Genuchten and D. R. Nielsen, 1989: A consistent set of parametric models for the two-phase flow of immiscible fluids in the subsurface, *Water Resour. Res.*, **25**, 2187–2193, doi:10.1029/WR025i010p02187.

Lu, S., T. Ren, Y. Gong and R. Horton, 2007: An improved model for predicting soil thermal conductivity from water content at room temperature, *Soil Sci. Soc. Am. J.*, **71**, (1), 8–14, doi: 10.2136/sssaj2006.0041.

Lu, S., T. Ren, Z. Yu and R. Horton, 2011: A method to estimate the water vapour enhancement factor in soil, *Europ. J. Soil Sci.*, **62**, 498–504, doi: 10.1111/j.1365–2389.2011.01359.x.

Lumley, J. L., 1970: *Stochastic Tools in Turbulence*, Academic Press, New York.

Mann, M. E. and P. D. Jones, 2003: Global surface temperatures over the past two millennia, *Geophys. Res. Lett.*, **30**, (15), 1820, doi:10.1029/2003GL017814.

Manz, B., P. Alexander and L. F. Gladden, 1999: Correlations between dispersion and structure in porous media probed by nuclear magnetic resonance, *Phys. Fluids*, **11**, (2), 259–267.

Massman, W. J., 1998: A review of the molecular diffusivities of H2O, CO2, CH4, CO, O3, SO2, NH3, N2O, NO, and NO2 in air, O2 and N2 near STP, *Atmos. Env.*, **32**, (6), 1111–1127, doi: 10.1016/S1352–2310(97)00391–9.

Matheron, G., 1971: *The Theory of Regionalized Variables and its Applications*, Ecole de Mines, Fontainbleau, France.

McGuire, V. L. and B. C. Fischer, 1999, Water-level changes, 1980 to 1997, and saturated thickness, 1996–97, in the High Plains aquifer, U.S. Geological Survey, Fact Sheet 124-99.

Mei, C. C. and J.-L. Auriault, 1991: The effect of weak inertia on flow through a porous medium, *J. Fluid Mech.*, **222**, 647–663.

Miller, E. E. and R. D. Miller, 1956: Physical theory for capillary flow phenomena, *J. Appl. Phys.*, **27**, 324–332.

Millington, R. J., 1959: Gas diffusion in porous media, *Science*, **130**, 100–102.

Millington, R. J. and J. P. Quirk, 1961: Permeability of porous solids, *Trans. Faraday. Soc.*, **57**, 1200–1207.

Mualem, Y., 1976: A new model for predicting the hydraulic conductivity of unsaturated porous media, *Water Resour. Res.*, **12**, 513–522.

Mualem, Y. and G. Dagan, 1978: Hydraulic conductivity of soils: Unified approach to the statistical models, *Soil Sci. Soc. Am. J.*, **42**, 392–395.

Nelson, F. E., O. A. Anisimov and N. I. Shiklomanov, 2001: Subsidence risk from thawing permafrost, *Nature*, **410**, 889–890, doi:10.1038/35073746.

Neuweiler, I. and H. Eichel, 2006: Effective parameter functions for the Richards equation in layered porous media, *Vadose Zone J.*, **5**, 963–977, doi: 10.2136/vzj2005.0127.

Nkedi-Kizza, P., J. W. Biggar, H. M. Selim, M. T. van Genuchten, P. J. Wierenga, J. M. Davidson and D. R. Nielsen, 1984: On the equivalence of two conceptual models for describing ion exchange during transport through an aggregated oxisol, *Water Resour. Res.*, **20**, 1123–1130, doi:10.1029/WR020i008p01123.

Okabe, H. and M. J. Blunt, 2007: Pore space reconstruction of vuggy carbonates using microtomography and multiple- point statistics, *Water Resour. Res.*, **43**, W12S02, doi:10.1029/2006WR005680.

Papoulis, A., 1984: *Probability, Random Variables and Stochastic Processes*, McGraw-Hill Book Company, Auckland, 2. edition.

Parker, J. C., J. B. Kool and M. T. van Genuchten, 1985: Determinig soil hydraulic properties from one-step outflow experiments by parameter estimation: II. Experimental studies, *Soil Sci. Soc. Am. J.*, **49**, 1354–1359.

Pebesma, E. J. and C. G. Wesseling, 1998: Gstat, a program for geostatistical modelling, prediction and simulation, *Computers & Geosciences*, **24**, (1), 17–31.

Pedroso, D. M., D. Sheng and J. Zhao, 2009: The concept of reference curves for constitutive modelling in soil mechanics, *Comp. Geotech.*, **36**, 149–165.

Peters-Lidard, C. D., E. Blackburn, X. Liang and E. F. Wood, 1998: The effect of soil thermal conductivity parameterization on surface energy fluxes and temperatures, *J. Agric. Sci.*, **55**, 1209–1224.

Petit, J. R., J. Jouzel, D. Raynaud, N. I. Barkov, J.-M. Barnola, I. Basile, M. Bender, J. Chapellaz, M. Davis, G. Delaygue, M. Delmotte, V. M. Kotlyakov, M. Legrand, V. Y. Lipenkov, C. Lorius, L. Pépin, C. Ritz, E. Saltzmann and M. Stievenard, 1999: Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica, *Nature*, **399**, (6735), 429–436.

Pfannkuch, H. O., 1963: Contribution à l'étude des déplacements de fluides miscible dans un milieu poreux, *Rev. Inst. Fr. Petrol.*, **18**, 215–270.

Philip, J. R., 1957*a*: The theory of infiltration: 1. The infiltration equation and its solution, *Soil Sci.*, **83**, 345–357.

Philip, J. R., 1957*b*: The theory of infiltration: 2. The profile of infinity, *Soil Sci.*, **83**, 435–448.

Phythian, R., 1975: Dispersion by random velocity fields, *J. Fluid Mech.*, **67**, 145–153.

Pollack, H. N., S. Huang and P.-Y. Shen, 1998: Climate change record in subsurface temperatures: A global perspective, *Science*, **282**, 279–281.

Postel, S. L., G. C. Daily and P. R. Ehrlich, 1996: Human appropriation of renewable fresh water, *Science*, **271**, 785–788.

Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, 2002: *Numerical Recipes in C++. The Art of Scientific Computing*, Cambridge University Press, Cambridge, UK, 2. edition.

Prunty, L. and F. X. M. Casey, 2002: Soil water retention curve description using a flexible smooth function, *Vadose Zone J.*, **1**, (1), 179–185, doi: 10.2136/vzj2002.1790.

Radcliffe, D. E. and J. Šimůnek, 2010: *Soil Physics with HYDRUS: Modeling and Applications*, CRC Press, Taylor & Francis Group.

Remson, I., G. M. Hornberger and F. J. Molz, 1971: *Numerical Methods in Subsurface Hydrology*, John Wiley & Sons, New York.

Renard, P. and G. de Marsily, 1997: Calculating equivalent permeability: a review, *Adv. Water Res.*, **20**, (5-6), 253–278.

Renn, J., 2005: Die atomistische Revolution, *Physik Journal*, **4**, (3), 53–59.

Richards, L. A., 1931: Capillary conduction of liquids through porous mediums, *Physics*, **1**, 318–333.

Richardson, L. F., 1910: The approximate arithmetical solution by finite differences of physical problems involving differential equations with an application to the stresses in a masonry dam, *Phil. Trans. Royal Soc. A*, **210A**, 307–357.

Richardson, L. F., 1922: *Weather Prediction by Numerical Process*, Cambridge University Press.

Roth, K. and J. Boike, 2001: Quantifying the thermal dynamics of a permafrost site near Ny-Ålesund, Svalbard, *Water Resour. Res.*, **37**, (12), 2901–2914, doi:10.1029/2000WR000163.

Roth, K. and K. Hammel, 1996: Transport of conservative chemical through an unsaturated two-dimensional Miller-similar medium with steady-state flow, *Water Resour. Res.*, **32**, (6), 1653–1663, doi: 10.1029/96WR00756.

Roth, K. and W. A. Jury, 1993: Linear transport models for adsorbing solutes, *Water Resour. Res.*, **29**, (4), 1195–1203, doi: 10.1029/92WR02537.

Roth, K., R. Schulin, H. Flühler and W. Attinger, 1990: Calibration of time domain reflectometry for water content measurement using a composite dielectric approach, *Water Resour. Res.*, **26**, (10), 2267–2273, doi: 10.1029/90WR01238.

Saffman, P. G. and G. Taylor, 1958: The penetration of a fluid into a porous medium or Hele-Shaw cell containing a more viscous liquid, *Proc. R. Soc. London A*, **245**, 312–329.

Scheidegger, A. E., 1961: General theory of dispersion in porous media, *J. Geophys. Res.*, **66**, (10), 3273–3278.

Schuster, H. G. and W. Just, 2005: *Deterministic Chaos: An Introduction*, John Wiley & Sons, 4. edition.

Sharon, E., M. G. Moore, W. D. McCormick and H. L. Swinney, 2003: Coarsening of fractal viscous fingering patterns, *Phys. Rev. Lett.*, **91**, 205504, 10.1103/PhysRevLett.91.205504.

Shokri, N., P. Lehmann and D. Or, 2009: Critical evaluation of enhancement factors for vapor transport through unsaturated porous media, *Water Resour. Res.*, **45**, W10433, 1–9, doi: 10.1029/2009WR007769.

Shraiman, B. I. and E. D. Siggia, 2000: Scalar turbulence, *Nature*, **405**, (6787), 639–646.

Šimůnek, J., T. Vogel and M. T. van Genuchten, 1994: The SWMS_2D code for simulating water flow and solute transport in two-dimensional variably saturated media, Research Report 132, U.S. Salinity Laboratory, USDA/ARS, Riverside, CA.

Skinner, J. L., 2010: Following the motions of water molecules in aqueous solutions, *Science*, **328**, 985–986, doi: 10.1126/science.1190093.

Skinner, L. M. and J. R. Sambles, 1972: The Kelvin equation – a review, *J. Aerosol Sci.*, **3**, (3), 199–210, doi: 10.1016/0021–8502(72)90158–9.

Smedley, P. L. and D. G. Kinniburgh, 2002: A review of the source, behaviour and distribution of arsenic in natural waters, *Appl. Geochem.*, **17**, (5), 517–568.

Stefan, J., 1891: Über die Theorie der Eisbildung, insbesondere über Eisbildung im Polarmeer, *Ann. der Physik*, **42**, 269–286.

Stiopkin, I. V., C. Weeraman, P. A. Pieniazek, F. Y. Shalhout, J. L. Skinner and A. V. Benderskii, 2011: Hydrogen bonding at the water surface revealed by isotopic dilution spectroscopy, *Nature*, **474**, 192–195, doi:10.1038/nature10173.

Stöhr, M., K. Roth and B. Jähne, 2003: Measurement of 3D pore-scale flow in index-matched porous media, *Experiments in Fluids*, **35**, (doi 10.1007/s00348-003-0641-x), 159–166.

Stöhr, M., 2003: *Analysis of Flow and Transport in Refractive Index Matched Porous Media*, Ph.D. thesis, Faculty of Physics and Astronomy, Heidelberg University, www.ub.uni-heidelberg.de/archiv/3733.

Sudicky, E. A., 1986: A natural gradient experiment on solute transport in a sand aquifer: Spatial variability of hydraulic conductivity and its role in the dispersion process, *Water Resour. Res.*, **22**, 2069–2082.

Taylor, G. I., 1953: Dispersion of soluble matter in solvent flowing slowly through a tube, *Proc. R. Soc. London*, **219A**, 186–203.

Taylor, G. K., R. L. Nudds and A. L. R. Thomas, 2003: Flying and swimming animals cruise at a Strouhal number tuned for high power efficiency, *Nature*, **425**, 707–711, doi: 10.1038/nature02000.

Thompson, P. A. and M. O. Robbins, 1989: Simulations of contact-line motion: Slip and the dynamic contact angle, *Phys. Rev. Lett.*, **63**, (7), 766–769, doi: 10.1103/PhysRevLett.63.766.

Toorman, A. F., P. J. Wierenga and R. G. Hills, 1992: Parameter estimation of hydraulic properties from one-step outflow data, *Water Resour. Res.*, **28**, (11), 3021–3028.

Topp, G. C. and E. E. Miller, 1966: Hysteretic moisture characteristics and hydraulic conductivities for glass-bead media, *Soil Sci. Soc. Am. Proc.*, **30**, 156–162.

Tritton, D. J., 1988: *Physical Fluid Dynamics*, Oxford University Press, 2. edition.

United Nations, 2006: *United Nations World Water Development Report 2*, United Nations Educational, Scientific and Cultural Organization (UNESCO).

van Dam, J. C., J. N. M. Stricker and P. Droogers, 1992: Inverse method for determining soil hydraulic functions from one-step outflow experiments, *Soil Sci. Soc. Am. J.*, **56**, 1042–1050.

van Dam, J. C., J. N. M. Stricker and P. Droogers, 1994: Inverse method to determine soil hydraulic functions from multistep outflow experiments, *Soil Sci. Soc. Am. J.*, **58**, 647–652.

van Everdingen, R. (Ed.), 1998: *Multi-Language Glossary of Permafrost and Related Ground-Ice Terms*, University of Calgary, Calgary, Canada.

van Geen, A., Y. Zheng, R. Versteeg, M. Stute, A. Horneman, R. Dhar, M. Steckler, A. Gelman, C. Small, H. Ahsan, J. H. Graziano, I. Hussain and K. M. Ahmed, 2003: Spatial variability of arsenic in 6000 tube wells in a 25 km^2 area of Bangladesh, *Water Resour. Res.*, **39**, (5), 1140, doi:10.1029/2002WR001617.

van Genuchten, M. T., 1980: A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Sci. Soc. Am. J.*, **44**, 892–898.

van Genuchten, M. T. and W. J. Alves, 1982: Analytical solutions of the one-dimensional convection dispersion solute transport equation, Tech. Bull. 1661, US Dept. of Agriculture, Salinity Lab., Riverside, CA.

van Genuchten, M. T. and P. J. Wierenga, 1976: Mass transfer studies in sorbing porous media. 1. Analytical solutions, *Soil Sci. Soc. Am. J.*, **40**, 473–480.

van Kampen, N. G., 1981: *Stochastic Processes in Physics and Chemistry*, North-Holland, Amsterdam.

Vereecken, H., J. Feyen, J. Maes and P. Darius, 1989: Estimating the soil moisture retention characteristic from texture, bulk density, and carbon content, *Soil Sci.*, **148**, 389–403.

Vogel, H. J., 2000: A numerical experiment on pore size, pore connectivity, water retention, permeability, and solute transport using network models, *Europ. J. Soil Sci.*, **51**, (1), 99–105.

Wagner, W., A. Saul and A. Pruss, 1994: International equations for the pressure along the melting and along the sublimation curve of ordinary water substance, *J. Phys. Chem. Ref. Data*, **23**, (3), 524–527, doi:10.1063/1.555947.

Walker, G. R., 1987: Solution to a class of coupled linear partial differential equations, *IMA J. Appl. Math.*, **38**, 35–48.

Weitz, D. A., J. P. Stokes, R. C. Ball and A. P. Kushnick, 1987: Dynamic capillary pressure in porous media: Origin of the viscous-fingering length scale, *Phys. Rev. Lett.*, **59**, (26), 2967–2970.

Wollschläger, U., 2003: *Kopplung zwischen Oberflächengewässer und Grundwasser: Modellierung und Analyse von Umwelttracern (in German)*, Ph.D. thesis, Faculty of Chemistry and Geosciences, Heidelberg University, www.ub.uni-heidelberg.de/archiv/3393.

Woodbury, A. D. and E. A. Sudicky, 1991: The geostatistical characteristics of the Borden aquifer, *Water Resour. Res.*, **27**, (4), 533–546.

Wösten, J. H. M., Y. A. Pachepsky and W. J. Rawls, 2001: Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics, *J. Hydrol.*, **251**, 123–150.

Young, T., 1805: An essay on the cohesion of fluids, *Proc. R. Soc. London*, **95**, 65–87.

Zachmann, D. W., P. C. Duchateau and A. Klute, 1981: The calibration of the Richards flow equation for a draining column by parameter identification, *Soil Sci. Soc. Am. J.*, **45**, 1012–1015.

Zhang, Y., W. Chen, S. L. Smith, D. W. Riseborough and J. Cihlar, 2005: Soil temperature in Canada during the twentieth century: Complex responses to atmospheric climate change, *J. Geophys. Res.*, **110**, D03112, doi:10.1029/2004JD004910.

Index