

DISSERTATION

submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

put forward by
Dipl. Phys. Klaus Schneider-Zapp
born in Marburg/Lahn
Oral examination: 01.12.2009

Dynamics of Water Movement near Boundaries of the Vadose Zone

Referees: Prof. Dr. Kurt Roth
Prof. Dr. Peter Bastian

Dynamik der Wasserbewegung an den Grenzschichten der ungesättigten Zone

Die Prozesse an den Grenzschichten der ungesättigten Bodenwasserzone wurden untersucht: am oberen Rand Verdunstung an der Boden-Atmosphäre Grenzschicht und am oberen Rand der Kapillarsaum. Am oberen Rand wurde ein Verdunstungsexperiment auf der Skala des repräsentativen Elementarvolumens (REV) betrachtet. Ein Modell mit einer diffusiven Grenzschicht und einer 1D Richards-Beschreibung beschrieb die experimentellen Daten gut. Es zeigte ein grenzschichtlimitiertes Regime im feuchten Bereich und ein hydrauliklimitiertes Regime im trockenen Bereich. Das Modell wurde unter Verwendung eines Monte-Carlo-Levenberg-Marquardt-Verfahrens erfolgreich zur inversen Bestimmung der hydraulischen Bodeneigenschaften mittels des entsprechenden Verdunstungsexperiments benutzt. Für den unteren Rand wurden Lichttransmissions- und bildgebende NIR-Spektroskopieverfahren entwickelt und angewendet, um mikro- und makroskopisch die Wasserverteilung, die als Antwort auf transiente Randbedingungen entstand, in einem Semi-2D-Sandmedium mit hoher zeitlicher und räumlicher Auflösung zu messen. Die Analyse zeigte, dass gekoppelte Mehrphasen- und Nichtgleichgewichtseffekte sowie sub-REV-Prozesse eine wesentliche Rolle in der Wasserbewegung im dynamischen Kapillarsaum spielen.

Dynamics of Water Movement near Boundaries of the Vadose Zone

Processes at boundaries of the unsaturated soil water zone were investigated: At the upper boundary evaporation at the soil-atmosphere interface, and at the lower boundary the dynamic capillary fringe. For studying the upper boundary, an evaporation experiment at the representative elementary volume (REV) scale was considered and modelled numerically. A model with a diffusive boundary layer and a 1D Richards' description including vapour transport fitted well to experimental data. It showed a boundary layer dominated regime in the wet range and a regime where dynamics is controlled by soil hydraulic properties in the dry range. The model could successfully be used to determine soil hydraulic properties from the corresponding evaporation experiment by inverse modelling using a Monte-Carlo Levenberg-Marquardt approach. For the lower boundary, light transmission and NIR imaging spectroscopy methods were developed and employed to measure the micro- and macroscopic water distribution in response to transient boundary conditions in a semi-2D sand medium in a Hele-Shaw cell with high temporal and spacial resolution. The analysis showed that coupled multi-phase and dynamic non-equilibrium effects are essential to understand water movement in dynamic capillary fringes, and sub-REV processes play an important role in the dynamics.

Contents

1	Introduction	1
I	Theory	2
2	Fluid flow in porous media	3
2.1	Microscopic description	3
2.2	Macroscopic description	5
2.2.1	Macroscopic variables	6
2.2.2	Empirical flux law of Buckingham-Darcy	6
2.2.3	Material properties	7
2.3	Scales	9
2.4	Flow regimes in the subsurface	10
2.4.1	Degenerate multiphase regime	11
2.4.2	Continuous multiphase regime	12
2.4.3	Discontinuous multiphase regime	12
2.4.4	Singlephase regime	12
2.5	Water vapour transport	13
2.6	Dynamic effects	14
3	Hysteresis	16
3.1	Definition and characteristics of hysteresis in general	16
3.2	Mechanisms for hysteresis in porous media	17
3.2.1	Inkbottle effect	17
3.2.2	Entrapped air	19
3.2.3	Contact angle hysteresis	20
3.2.4	Pinning and snap-off	20
3.2.5	Surface roughness	20
3.3	Modelling hysteresis	21
3.3.1	The independent domain model	22
3.3.2	The dependent domain model	23
3.4	Model performance	25
4	Image analysis	28
4.1	Image representation	28
4.1.1	Homogeneous coordinates	28
4.2	Sampling and sampling theorem	29
4.3	Neighbourhood operations	31
4.3.1	Averaging	31
4.3.2	Edge detection and derivatives	32
4.4	Noise model for image sensors	32
4.5	Deconvolution	33
4.5.1	Determination of the PSF	35

4.5.2	Band-limiting deconvolution	36
4.5.3	Richardson-Lucy deconvolution	37
4.5.4	Wiener filter	39
4.5.5	Ringings	40
5	Light transmission through porous media	41
5.1	Properties of light	41
5.1.1	Refraction and Reflection	42
5.1.2	Absorption	43
5.1.3	Geometric optics	44
5.2	Optical properties of water	45
5.3	Radiative transport through thin porous media	46
5.3.1	Visible and infrared light	46
5.3.2	X-rays	49
6	Numerical simulation and inverse modelling	50
6.1	Finite differences	50
6.2	Finite volumes	51
6.3	Inversion	52
II	The soil-atmosphere boundary during evaporation	54
7	Introduction	55
8	Materials and Methods	58
8.1	Setup of the Evaporation Experiment	58
8.2	Numerical Model	59
8.3	Numerical Experiments and Analysis	60
9	Results	63
9.1	Numerical study	63
9.1.1	Onestep experiment	63
9.1.2	Multistep experiments	66
9.1.3	Response surfaces	71
9.1.4	Convergence study	72
9.2	Inversion of experimental data	75
10	Conclusions	80
III	The capillary fringe	82
11	Introduction	83
12	Materials and methods	85
12.1	Experimental setup – overview	85
12.2	Porous medium	86
12.3	The Light Transmission Method	88
12.3.1	Introduction	88
12.3.2	Experimental setup	89

12.3.3	NIR imaging spectroscopy wavelength selection and test	93
12.3.4	Image processing	94
12.4	Consumer DSLR cameras	101
12.4.1	White balance and Bayer pattern averaging in LTM	103
12.5	Camera calibration	103
12.5.1	Calibration of the consumer camera	103
12.5.2	Calibration of the NIR camera	107
12.6	Deconvolution	109
12.7	X-ray measurements	115
12.7.1	Water content calibration	117
12.8	Additional instruments	121
12.8.1	Pumps	121
12.8.2	Balances	121
12.8.3	Temperature measurements	122
13	Results and discussion	123
13.1	Dynamics of a fluctuating water table	123
13.1.1	Measurement protocol	123
13.1.2	Evaluation	125
13.1.3	System response	125
13.1.4	Capillary fringe edge response	127
13.2	Dynamics of fluctuating infiltration	133
13.2.1	Measurement protocol	133
13.2.2	System response	134
13.3	Dynamics of fingered infiltration	135
14	Conclusions	139
	Summary and sidelook	140
	Appendix	143
A	Statistics	144
A.1	Random variables	144
A.2	Multi-dimensional random variables	145
A.3	Functions of random variables	145
B	Fourier transform	147
	Nomenclature	150

1 Introduction

For life, water is maybe the most important resource on earth. Although 71 % of earth's surface is covered by oceans, man's activity is mainly limited to land. Here, soils are the most important ingredient and it is essential to understand the dynamics of water movement in soils in order to comprehend the water cycle. In the last decades, large progress was made in understanding the different parts of earth's system as soils, ocean, or atmosphere. However, also the coupling between these components has a major influence on the dynamics. While in the past scientists often concentrated on their respective parts of interest, in recent years the complex interactions between the different systems have come more into focus. Yet these boundaries are much less understood than the parts themselves.

There are three major interfaces in earth's system: (i) the soil-atmosphere interface, (ii) the water-atmosphere interface, and (iii) the water-soil interface. In the broadest meaning, the third type also includes the interface between the water saturated groundwater zone and the unsaturated multi-phase zone (vadose zone) of the soil, the so-called capillary fringe. Two of those interfaces, namely the soil-atmosphere boundary and the capillary fringe, are investigated in the two respective parts of this thesis. For the soil-atmosphere boundary, the focus was on the evaporation process.

Part I

Theory

2 Fluid flow in porous media

Soils are porous materials. They consist of the soil material itself, i.e. the grains or particles the soil is made of, called soil matrix, and the intermediate space called pores. These pores can be filled with fluids such as water or air. Fluid dynamics in porous media is in general a multi-phase problem, since a change in one fluid content necessarily involves a change of fluid content of at least one other fluid.

Here we give an introduction to the concepts used in this thesis. For a detailed review and discussion of fluid flow and solute transport in porous media it is referred to Roth (2005).

2.1 Microscopic description

At the microscopic scale, fluid flow is governed by Navier-Stokes' equation, supplemented by appropriate initial and boundary conditions. Navier-Stokes' equation can be derived by writing Newton's law, $f = \rho \dot{v}$, where f is force density, ρ density and v velocity, for the fluid (note that all variables implicitly depend on space and time):

$$\rho \dot{v} = \rho \frac{\partial v}{\partial t} + \rho(v \cdot \nabla)v = f = \underbrace{-\nabla p}_{\text{pressure}} + \underbrace{\rho g}_{\text{gravity}} + \underbrace{\nabla \cdot \Pi}_{\text{friction}}. \quad (2.1)$$

The terms of the right correspond to the forces due to pressure, gravity, and friction, respectively. p denotes pressure, g the acceleration of gravity, and

$$\Pi_{ij} = \eta \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} - \frac{2}{3} \delta_{ij} \sum_l \frac{\partial v_l}{\partial x_l} \right) + \zeta \delta_{ij} \sum_l \frac{\partial v_l}{\partial x_l} \quad (2.2)$$

is the friction tensor (Landau and Lifschitz, 1991, §15). Here, η is the dynamic viscosity and ζ the second viscosity coefficient. The fluid is assumed to be Newtonian, i.e. friction only depends linearly on the first derivatives of the velocity and not on higher order derivatives.

The boundary conditions for the microscopic description must include the exact geometry of the porous domain and the fluid-fluid interfaces. For the latter, capillary forces must also be considered. These stem from inter-molecular forces which originate from electromagnetic interactions between atoms or molecules. While in gases the density is so low that intermolecular forces are only relevant when two particles collide, the density in liquids is high enough that these interactions play a fundamental role for the properties of the liquid. In the bulk of a liquid volume, the molecules are arranged such that forces average out. At the interfaces to other fluids however the forces cannot be compensated and thus a net force develops. On the other hand, moving molecules into the inside of the liquid would increase pressure. In equilibrium the liquid will occupy a state of minimal total energy. Thus, a surface energy density, the interfacial or capillary energy σ (energy per area) can be assigned to the surface. It depends on the properties of the two media at the interfaces. Notice that σ may become negative if the attraction between molecules within fluid 1 is smaller than with molecules of fluid 2 or a solid. This leads to dissolution of fluid 1 in 2 or, in case of a solid, complete coverage of the solid by the fluid.

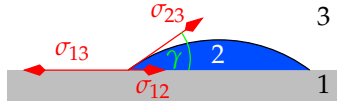


Figure 2.1: Tensions exerted by the three interfaces between a solid 1 and two fluids 2 and 3. The contact angle γ adjusts itself such that the forces balance.

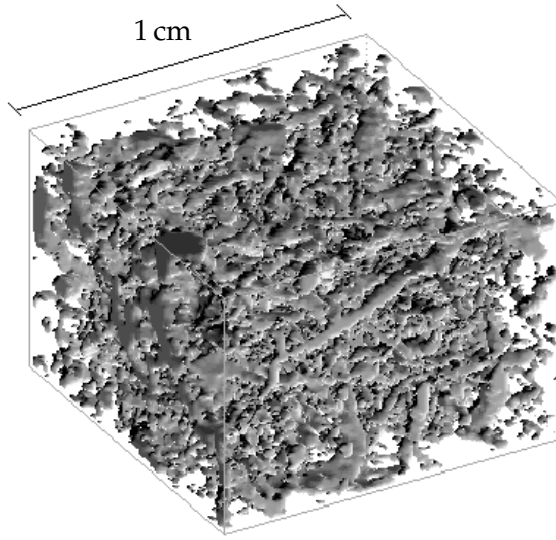


Figure 2.2: Three-dimensional microscopic reconstruction of a soil sample. Dark areas denote pores. Image courtesy of H.-J. Vogel.

If three different fluids or two fluids and a solid are brought into contact, at the contact point of all three phases, the surfaces have a certain angle γ , the so-called contact angle. This is illustrated in figure 2.1 for a solid. In equilibrium, the forces balance and one obtains

$$\sigma_{23} \cos \gamma = \sigma_{13} - \sigma_{12} \quad (2.3)$$

The pressure at a curved fluid-fluid interface is discontinuous due to the capillary energy. The pressure difference is

$$p_1 - p_2 = \sigma \left(\frac{1}{r_1} + \frac{1}{r_2} \right), \quad (2.4)$$

where r_1 and r_2 are the major radii of curvature of the interface between fluids 1 and 2 (Landau and Lifschitz, 1991, §61). Equation (2.4) is called Young-Laplace equation.

The boundary condition for the interface between the two fluids 1 and 2 reads

$$\mathbf{n} \left\{ (p_1 - p_2) - \sigma \left(\frac{1}{r_1} + \frac{1}{r_2} \right) \right\} = (\Pi_1 - \Pi_2) \cdot \mathbf{n} + \nabla \sigma, \quad (2.5)$$

where \mathbf{n} denotes the normal unit vector pointing into the inner of fluid 1. Note the Young-Laplace term describing the effects of capillarity. $\nabla \sigma$ on the right describes the tangential force exerted by capillarity.

In principle equations (2.1) and (2.5) can be used to solve the flow problem. A first step into this direction was done by Heimann (2009). However, the exact geometry of a soil is very complicated (figure 2.2) and has structures at many different scales, with typical dimensions ranging from micrometres (loam particles) to centimetres (larger constituents) and even metres. The exact geometry is difficult to obtain, especially for larger samples. Also, it requires an incredibly large amount of data for larger samples. In a pore geometry with relevant structures of about some micrometres, just 1 m^3 would already have about 10^8 degrees of freedom.

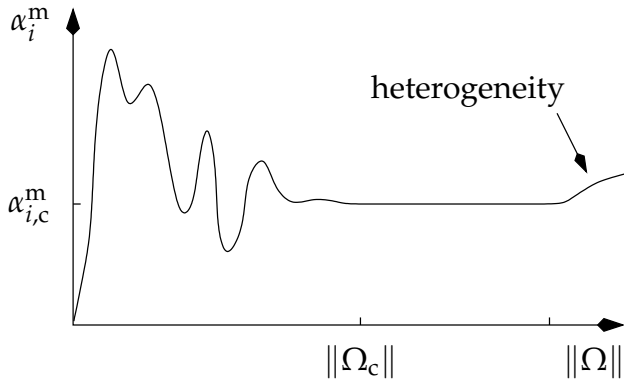


Figure 2.3: Sketch of the variation of the averaged macroscopic quantity α_i^m in phase i depending on the extend of the averaging volume $\|\Omega\|$. If $\|\Omega\|$ is small, α_i^m depends crucially on the choice of κ . If the averaging volume becomes larger than a critical value $\|\Omega_c\|$, it eventually becomes stable. If $\|\Omega\|$ is increased even more, at some point macroscopic heterogeneities become visible if the soil is not homogeneous.

2.2 Macroscopic description

In most cases the detailed geometry is not of interest, but only the effective macroscopic dynamics. In analogy to thermodynamics of gases, where macroscopic properties like pressure are of interest and the trajectories of the particles do not matter, some averaging can be done. One scale is picked out, and a description is chosen which emphasises the aspects of that scale.

Let $\chi_i(x)$ be the phase indicator function for phase i , where i might be soil matrix (m), water (w), or air (a),

$$\chi_i(x) := \begin{cases} 1, & x \in \text{phase } i \\ 0, & x \notin \text{phase } i \end{cases} \quad (2.6)$$

Averaging a microscopic quantity α^μ leads to the averaged (macroscopic) quantity α_i^m in phase i :

$$\alpha_i^m(x, t) := \int_{\Omega} \alpha^\mu(x - \xi, t) \chi_i(x - \xi) \kappa(\xi) d\xi = \langle \alpha^\mu \chi_i \rangle, \quad (2.7)$$

where κ is the averaging weight function and Ω the support of κ (Roth, 2005). Naturally $\kappa(x) \geq 0$ for all x and $\int_{\Omega} \kappa(x) dx = 1$ are required. Popular choices for κ are box functions which are constant inside a cube or sphere and 0 outside. Then κ can be interpreted geometrically as the averaging volume.

In general, the macroscopic property α_i^m will depend on the choice of κ which makes a meaningful definition of α_i^m impossible. However, if the porous medium is sufficiently uniform and κ is non-zero in a volume which is large enough, eventually the value of α_i^m will not depend on the actual size or shape of the averaging volume anymore (figure 2.3) and reach the value $\alpha_{i,c}^m$ (convergence). Such an averaging volume is called a representative elementary volume (REV). The existence of an REV is a prerequisite for a macroscopic description. Note that the REV is defined as the *smallest* possible volume where the averaging converges. Due to the multi-scale nature of soils (Roth, 2005, sect. 3.1), larger volumes might exist where the averaging converges to a different value $\alpha_i^m \neq \alpha_{i,c}^m$.

Obviously, to calculate macroscopic variables, χ_i and thus the detailed microscopic geometry must be known. As outlined above, this is normally not the case, therefore these values have to be estimated experimentally.

In the following it is assumed that the averaging converges and thus macroscopic variables exist. For simplicity the superscript ^m will be dropped from the quantities. The compressibility of water in the pressure range which can be found in soils is so small that the water phase can be assumed to be incompressible, $\rho_w = \text{const}$. The porous medium is assumed water-wettable¹,

¹ This means that $\sigma_{ws} < 0$. Water-wettable surfaces are always covered by a thin film of water, if enough liquid is available. An example is clean glass.

which is a good assumption for almost all soils, and rigid.

2.2.1 Macroscopic variables

The macroscopic variable $\theta_i := \langle \chi_i \rangle$ is called the fluid content of fluid i . It can be interpreted as volume fraction of fluid i . $\phi := 1 - \langle \chi_m \rangle = \sum_{i \in \text{fluids}} \theta_i$ is called porosity and denotes the total volume fraction of pore space. If only water is considered, the abbreviation $\theta := \theta_w$ is used. Note that while at the pore scale every point belongs exactly to one phase, in the macroscopic description one point belongs to several phases. The state of a fluid element of a fluid i in a porous medium is defined by its height z_i , its pressure p_i , its temperature T_i and the concentrations C_{ij} of dissolved chemicals j .

Potential

The density of potential energy $\psi_i(\mathbf{x})$ of a fluid i is defined by the energy that is necessary to move a unit volume of fluid from the reference state $z = z_0$, $p = p_0$, $T = T_0$, $C_{ij} = 0$ to the state at position \mathbf{x} . p_0 is normally chosen to be the ambient air pressure. In an isothermal situation with no dissolved chemicals, only pressure and gravitation contribute to the energy density:

$$\psi_i(\mathbf{x}) = \underbrace{p_i(\mathbf{x}) - p_0}_{=: \psi_{ip}} - \underbrace{\int_{z_0}^z \rho_i(\zeta) g \, d\zeta}_{=: \psi_{ig}}. \quad (2.8)$$

By convention, the z direction points downwards into the ground.

In an incompressive medium, the density ρ_i is constant, and the second term, the potential due to gravity, simplifies to $\psi_{ig}(\mathbf{x}) = (z - z_0)\rho_i g$. The pressure term ψ_{ip} may consist of several components and can be of complex structure. In case of neglectable air pressure gradients, it corresponds to the pressure jump at the interface, $p_w - p_a$, and is also called matric potential ψ_m . Further details follow below in section 2.4.

2.2.2 Empirical flux law of Buckingham-Darcy

With the prerequisite that (a) the external forcing (i.e. the influence of external forces to the system) changes so slowly that the flow may be considered stationary, (b) the flux is so small that inertia can be neglected, (c) the dimensions of the system are small enough that gravitation can be neglected compared to viscosity, and (d) the fluid is considered incompressible, Navier-Stokes' equation can be simplified to Darcy's Law

$$\mathbf{j} = -K \nabla p \quad (2.9)$$

using a similarity analysis (Roth, 2005). Due to the small dimensions and low velocities, there is no turbulent flow inside the medium, hence the linear dependence of the flow on the driving force. Note the analogy to other flow equations, e.g. in electrodynamics. K denotes the hydraulic conductivity, in general this is a second rank tensor, as flow and driving force are not necessarily in parallel.

If gravitation cannot be neglected, an additional driving force adds up to the pressure gradient:

$$\mathbf{j} = -K(\nabla p - \rho \mathbf{g}) \quad (2.10)$$

In a homogeneous medium, the tensor K reduces to a scalar K , the flow is parallel to the driving forces. $\nabla p - \rho \mathbf{g}$ can be identified with $\nabla \psi_i = \nabla[(p_i - p_0) - (z - z_0)\rho_i g]$, because p_0 and z_0 are constants.

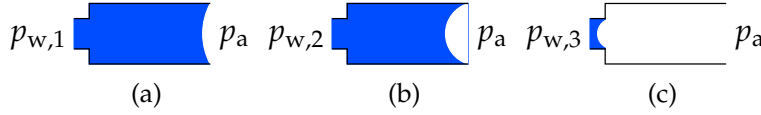


Figure 2.4: Illustration of the dependence of water content on potential for a simple pore. The right of the pore is connected to atmosphere. When reducing water pressure p_w at the left of the pore, the radius of the interface adjusts according to Young-Laplace's equation (2.4) (b). If the corresponding radius drops below the radius of the pore, the interface cannot be sustained and the pore is drained (c). The water recedes to the next position where an interface is energetically possible.

If more than one fluid is considered, everything gets much more complicated. In general, one falls back to Buckingham's conjecture. It states that Darcy's law remains valid, but the conductivity of the fluid i will be a function of the fluid fraction θ_i :

$$j_i = -K_i(\theta_i) \nabla \psi_i. \quad (2.11)$$

This law is called Buckingham-Darcy law. It is an empirical flux law, because it describes measured processes reasonably well but cannot be deduced from first principles. Onsager's theorem would suggest an equation of the form $j_i = -\sum_l K_{il}(\theta_l) \nabla \psi_l$, i. e. a sum of all phases l instead of a dependence on the same fluid alone. However, this issue has not yet been explored (Roth, 2005).

2.2.3 Material properties

Soil water characteristic

The pressure jump at the interface between water and air, $\psi_m = p_w - p_a$, is connected with the curvature of the interface via Young-Laplace's relation (2.4). The radius of the interface cannot be smaller than the radius of the pore the interface is located in. If ψ_m is reduced (e.g. by applying an external pressure), the radius $|r|$ of the interface must become smaller according to equation (2.4). If $|r|$ gets smaller than the radius \bar{r}_i of pore i , $|r| < \bar{r}_i$, the interface cannot be sustained anymore, and that pore will be drained (figure 2.4). This point is called the air entry point.

This line of thoughts reveals that the water content θ_w is a function of the matric potential ψ_m , $\theta_w = \theta_w(\psi_m)$. It is called soil water characteristic or water retention curve. The function depends on the geometry of pore space and is thus a material property, which is fixed (but hysteretic) for rigid porous media. When the potential becomes more negative, large pores are drained first since $1/|r|$ is smaller for large $|r|$.

Conductivity

Pores which are filled with air are not available for liquid water transportation since this would require moving water-air interfaces. Typically gradients are not large enough for this. Changing water-air interfaces normally involves changing the water content. Therefore, the greater the water content θ , the larger the hydraulic conductivity K . Large connections are better conductors than narrow necks: according to Hagen-Poiseuille's law, the conductivity for laminar flow in pipes of radius R increases with R^4 (remember that soil water flow does not exhibit turbulence). With decreasing water content, large pores are drained first and the remaining water is located in smaller pores with smaller conductivity. As a result, the hydraulic conductivity of (liquid) water K_w decreases rapidly with decreasing water content. Additionally, the path between two points A and B which only consists of water filled pores becomes longer when the medium gets drier. This effect is called tortuosity and also decreases the effective

(macroscopic) hydraulic conductivity K_w . Consequently, K_w is a function of the water content, $K_w = K_w(\theta_w)$. As θ_w is a function of ψ_m , K_w also depends on ψ_m .

Parametrisation

If the exact geometry of pore space is not available, which is normally the case, the hydraulic properties cannot be calculated or deduced from first principles. Thus, they have to be measured. Hence, some kind of parametrisation is needed to be fitted to experimental data.

Commonly, these parametrisations are formulated in terms of the water saturation

$$\Theta := \frac{\theta - \theta_r}{\theta_s - \theta_r}, \quad (2.12)$$

where θ_r and θ_s are the residual and the saturated water content, respectively, and the subscript w has been dropped. Based on a power-law distribution of pore radii with a finite upper end, [Brooks and Corey \(1966\)](#) introduced the parametrisation

$$\Theta(\psi_m) = \begin{cases} [\psi_m / \psi_e]^{-\lambda} & , \psi_m < \psi_e \\ 1 & , \psi_m \geq \psi_e \end{cases}, \quad (2.13)$$

where ψ_e is the air entry potential and $\lambda > 0$ an empirical parameter. [Van Genuchten \(1980\)](#) developed a parametrisation whose shape is similar to the one of [Brooks and Corey \(1966\)](#) but which has continuous derivatives:

$$\Theta(\psi_m) = [1 + (\alpha \psi_m)^n]^{-m} \quad (2.14)$$

with parameters $\alpha > 0$, $n > 1$ and $m > 0$. It is more popular because the capacity $C = \frac{\partial \theta}{\partial \psi_m}$ which can be found in Richards' equation (2.19) is continuous. Often, $m = 1 - 1/n$ is used and the number of free parameters thereby reduced by one.

Newer experiments revealed that in some cases these parametrisations are not flexible enough to describe measured $\Theta(\psi_m)$ curves accurately ([Zurmühl and Durner, 1998](#)). Therefore, more flexible approaches like spline parameterisations are currently being investigated.

Using a simple model of pore space, a relation between K and Θ can be deduced. [Mualem \(1976\)](#) used an isotropic medium with randomly connected stacks of capillary bundles. In each capillary, knowing its radius, the conductivity can be calculated according to Hagen-Poiseuille's law. The matric potential is related to the radius of the interface according to equation (2.4). Only the fraction of capillaries which have a radius equal or smaller than the corresponding matric potential are water filled and thus contribute to the conductivity. With this idea, he obtained the relation

$$K(\Theta) = K_s \Theta^\tau \left[\frac{\int_0^\Theta \psi_m(\xi)^{-1} d\xi}{\int_0^1 \psi_m(\xi)^{-1} d\xi} \right]^2, \quad (2.15)$$

where K_s denotes the saturated conductivity and the factor Θ^τ describes tortuosity. Inserting the inverted [Brooks and Corey \(1966\)](#) parametrisation (2.13) yields

$$K(\Theta) = K_s \Theta^{\tau+2+2/\lambda} \quad (2.16)$$

which is called Mualem-Brooks-Corey parametrisation. Applying (2.15) to the [van Genuchten \(1980\)](#) parametrisation (2.14) leads to the Mualem-van Genuchten parametrisation

$$K(\Theta) = K_0 \Theta^\alpha \left[1 - \left(1 - \Theta^{n/(n-1)} \right)^{1-1/n} \right]^2. \quad (2.17)$$

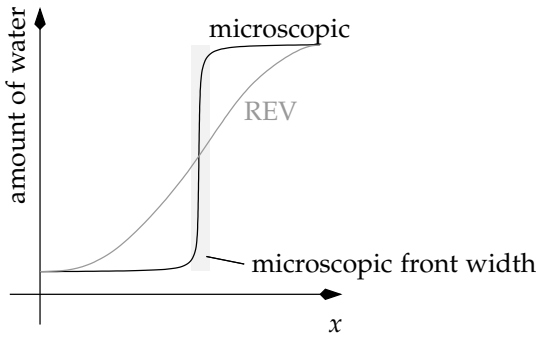


Figure 2.5: Microscopic and macroscopic view of a water front. Although the front may be very sharp at the microscopic scale, this is not resolved at the REV scale and the front is smeared out.

2.3 Scales

As shown above, water content θ is only defined macroscopically. At a scale below the REV scale, θ is undefined. The REV length scale is at least a few correlation lengths of the medium. An illustrative example for a measured 3D soil structure is given in Roth (2005, sect. 3.2.1). Hence, water content in a homogeneous medium is always smooth. However, a water front moving through the soil may be very sharp. Especially during infiltration, the infiltration front is self-sharpening, since water further away from the front is faster because conductivity is higher due to higher water content (see also Roth (2005, sect. 5.4)). Consequently the front on a sub-REV scale is different to the front as seen on a REV scale. Although Richards' equation is capable of describing this self-sharpening mechanism, the resulting dynamics may have length scales below the REV size and thereby implicitly produces results on a scale the equation is not capable to describe by definition. Accordingly sub-REV front widths as described by Richards' equation do not necessarily describe the real microscopic behaviour correctly. Macroscopic water content changes will always be much smoother and do not represent the front sharpness at the microscopic scale. Here, the front is only defined in the length scale of the REV and no finer features are resolved. This is sketched in figure 2.5. The discrepancy may be particularly large in coarse textured media and on fast movements where non-equilibrium processes are involved.

If the local equilibrium hypothesis is violated, the macroscopic variables are no well-defined quantities any more and REV-scale effective properties do not exist. An example is a rapid water movement which involves positive water pressure in the unsaturated zone and therefore reverse-curvature at water-air interfaces. Here, a sub-REV description must be used. Note that in such situations, Richards' equation is also not valid and it cannot be used to describe such phenomena.

The driving force of water movements is a gradient in the matric potential ψ_m . At the microscopic scale, it is defined by the curvature of the water-air interfaces according to Young-Laplace's equation (2.4). In the bulk of the water, ψ_m is per definition the energy density caused by capillarity which is necessary to move the water in from a reference state and thereby equal to the curvature of the nearby interfaces. In the air phase, the potential is defined by equation (2.26). Assuming local thermodynamic equilibrium, it is equal to the potential of the interfaces around it, which causes a very high water vapour saturation. Accordingly, the microscopic ψ_m is continuous across water/air phase boundaries. However, it is undefined in the matrix. At the macroscopic scale ψ_m denotes the average curvature in the REV and in equilibrium, it is always coupled with θ by the water characteristic $\theta(\psi_m)$. Gradients at scales smaller than the REV and especially dynamic positive curvature which is caused by forced imbibition at positive pressure into a low-conductive medium are not represented at the macroscopic scale and indicate that the macroscopic description is at its limit.

If the system response due to rapid changes is to be analysed, the contrast explained above

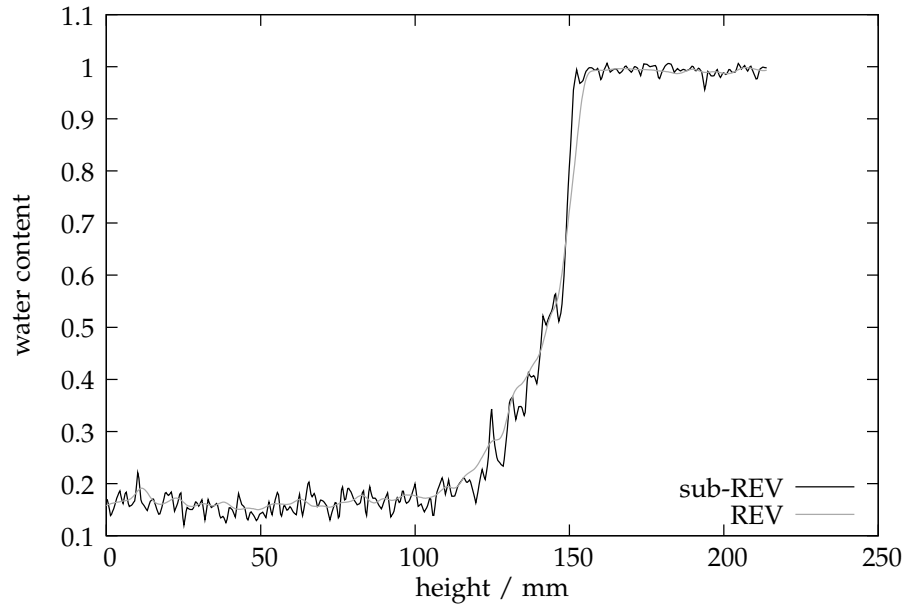


Figure 2.6: Vertical water content distribution profile in static equilibrium of the sand medium in the Hele-Shaw cell at different scales. Water content was observed with X-ray transmission at a sub-REV resolution of 1 mm/pixel, REV data was averaged with a Gaussian filter of $\sigma = 4$ mm radius, which yields an effective resolution of 8 mm. The profile was made at one vertical position.

has a crucial influence. On one hand, macroscopic water content is to be measured and the scale has to be selected accordingly. On the other hand, infiltration and drainage fronts microscopically are often sharper than the REV size, and if these fronts are to be analysed, sub-REV resolution is required for precise front characterisation. Of course, then the values for the amount of water do not represent a macroscopic water content any more.

In static equilibrium, the water content distribution along the height represents the water characteristic, $\theta = \theta(\rho_w g z)$. If the water characteristic is very sharp in contrast to the REV size, the transition will also be at a sub-REV scale. This may be the case with relatively coarse media. Here, we again have the problem of resolving the transition in the macroscopic description. This is illustrated in figure 2.6. As expected, the water content distribution is much smoother, but the transition is also blurred.

If the requirements for an REV would be weakened by allowing shape-dependence, the problem may be alleviated. When approaching the front, REV's would have to be stretched alongside the front and compressed perpendicular. This would allow to locate the front more sharply in the macroscopic description. However, it would have implications on the macroscopic properties, since they have other characteristics than with normal averaging.

2.4 Flow regimes in the subsurface

Due to the multiphase nature of the flow process, the dynamics of water in soils differs fundamentally in different regions of the soil with different water contents. This is schematically shown in figure 2.7 and described in the following.

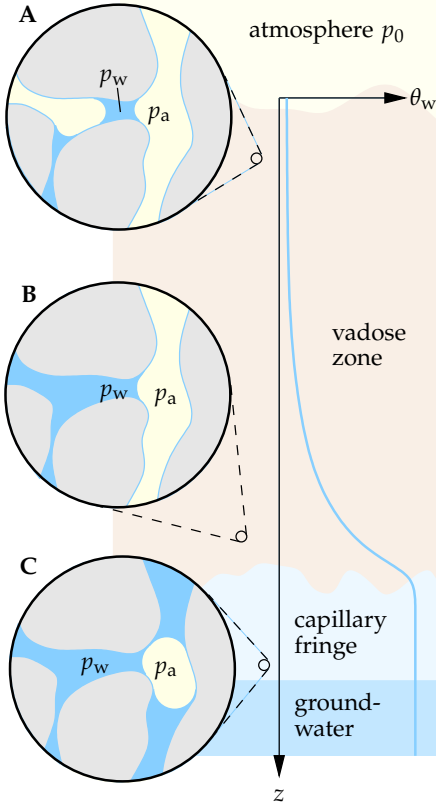


Figure 2.7: Flow regimes in soils. Far from the groundwater, in the *degenerate multiphase regime*, water content θ_w is so small that the air phase is continuous and connected to the atmosphere, gradients of p_a are negligible compared to those of p_w . The two phases are decoupled (A). Further down, nearer to the groundwater or with high infiltration, gradients of p_a cannot be neglected any more due to low θ_a , although the air phase is still continuous. Hence, the two phases become coupled. This regime is called *continuous multiphase regime* (B). With θ_w increasing even more, the air phase becomes discontinuous, leading to the *discontinuous multiphase regime* where entrapped air bubbles can be found (C). Even further down, below the groundwater where all gas is dissolved, only the water phase exists. This is the groundwater or *singlephase regime*. Image source: Roth (2005)

2.4.1 Degenerate multiphase regime

Where water content is low, far from the groundwater, the air content is high and thus the air phase continuous and connected to the atmosphere. Since the viscosity of air is much smaller than the one of water ($\mu_a/\mu_w = 0.0136$ at $T = 10^\circ\text{C}$), the pressure gradient necessary to move the air can be neglected in this situation (i. e. the air is assumed to be infinitely mobile, $\mu_a = 0$). Thus, the two phases decouple, a movement of the water phase is always followed by an instantaneous movement of the air phase. The volume change of air is buffered by the atmosphere which is a pressure reservoir, $p_a = p_0 = \text{const}$. The system can therefore be adequately described with only the water phase. The only contribution to the pressure potential is given by the interfacial forces, equation (2.4). In that case, ψ_p is called matric potential ψ_m , because it describes the energy density which is caused by the geometry of the soil matrix due to capillarity:

$$\psi_m = p_w - p_a. \quad (2.18)$$

In the convention used here, ψ_m is negative if the water is bound ($p_w < p_0$) and positive if it is free. This regime is called *degenerate multiphase regime*.

Inserting the flux law, equation (2.11), $j_w = -K(\theta_w)[\nabla\psi_m - \rho_w g]$, into the mass balance for water, $\frac{\partial}{\partial t}\theta_w + \text{div } j_w = 0$, yields an equation which describes the water movement in the degenerate multiphase regime:

$$\frac{\partial\theta_w(\psi_m)}{\partial\psi_m} \frac{\partial\psi_m}{\partial t} - \nabla \cdot [K(\psi_m) (\nabla\psi_m - \rho_w g)] = 0. \quad (2.19)$$

It is called *Richards' equation*, since it was proposed by Richards (1931). For a complete description, the material properties $\theta_w(\psi_m)$ and $K(\theta_w)$ (see section 2.2.3) must also be given. A

prerequisite is that the system is in a state of dynamic equilibrium such that these relationships exist. This is similar to thermodynamics where the equations of state are only defined in thermodynamic equilibrium. The estimation of the material properties is normally the most difficult problem when using Richards' equation.

2.4.2 Continuous multiphase regime

Both air and water phase are continuous, i. e. any two points in phase i can be connected by a path within that same phase, but due to higher water content than in the degenerate multiphase regime, gradients of p_a cannot be neglected. For example, if the air must flow through small channels, a non-neglectable pressure gradient is necessary to push the air through. Since only small pressure fluctuations are considered, the air is assumed to be incompressible. The flux of each phase is then described by the Buckingham-Darcy law (2.11) and the potential as defined in (2.8). One-dimensional horizontal (i. e. no gravity) flow is then described by (Roth, 2005)

$$j_a = -K_a \frac{\partial}{\partial x} \psi_a \quad (2.20)$$

$$j_w = -K_w \left[\frac{\partial}{\partial x} \psi_m + \frac{\partial}{\partial x} \psi_a \right] \quad (2.21)$$

since $\psi_w = [p_w - p_a] + [p_a - p_0] = \psi_m + \psi_a$, with $\psi_m = p_w - p_a$ as defined in equation (2.18). Sometimes ψ_m is also called *static capillary pressure* and $\psi_m + \psi_a$ *dynamic capillary pressure*. The conservation of mass and thus (since the fluids are considered incompressible) fluid volume is written as (Roth, 2005)

$$\frac{\partial}{\partial t} \theta_a - \frac{\partial}{\partial x} \left[K_a(\theta_a) \frac{\partial}{\partial x} \psi_a \right] = 0 \quad (2.22)$$

$$\frac{\partial}{\partial t} \theta_w - \frac{\partial}{\partial x} \left[K_w(\theta_w) \left\{ \frac{\partial}{\partial x} \psi_m + \frac{\partial}{\partial x} \psi_a \right\} \right] = 0 \quad (2.23)$$

$$\frac{\partial}{\partial t} \theta_a + \frac{\partial}{\partial t} \theta_w = 0, \quad (2.24)$$

where the last equation describes that pores are only filled with water or air. To describe the simultaneous flow of water and air in soils, these systems of coupled partial differential equations have to be solved.

2.4.3 Discontinuous multiphase regime

When air content becomes small enough, the air phase will not be continuous anymore, but discontinuous patches of air totally enclosed by water form which typically reside in large pores. Air may leave such regions as bubbles or by dissolution. As a consequence of bubble formation and dissolution, the air pressure in different isolated regions may be different. To solve the flow problem, the pore-scale processes of bubble formation, transport, and dissolution must be averaged, which has not been solved to date (Roth, 2005). Numerical codes for multiphase flow typically use macroscopic fluid contents and potentials for both fluids and couple these macroscopic variables with heuristic equations. However, physically the coupling is microscopical and averaging would have to be done *after* coupling the phases.

2.4.4 Singlephase regime

Only the water phase is remaining and saturation is constant, $\theta_a = 0$ and $\theta_s = \phi$. Since no fluid-fluid interfaces are present, the dynamics are linear, conductivity K is a constant which only depends on the porous material, i. e. the geometry of the soil matrix.

2.5 Water vapour transport

In the dry range, water vapour transport becomes a relevant transport process. To distinguish liquid and vapour, superscripts g (gas) and ℓ (liquid) are used in the following. The equivalent flux of liquid water transported by diffusion of water vapour j_w^g is given by the diffusion law

$$j_w^g = -V_{m,w} D_w^g \nabla \nu_w^g, \quad (2.25)$$

where $V_{m,w}$ is the molar volume of liquid water, D_w^g the diffusion coefficient of water vapour in air, and ν_w^g the molar density of water vapour. Assuming local thermodynamic equilibrium, ν_w^g is given by (Rawlins and Campbell, 1986)

$$\nu_w^g = \frac{p_{w,s}(T)}{RT} \exp \left[\frac{\psi_m V_{w,m}}{RT} \right], \quad (2.26)$$

where $p_{w,s}(T)$ is the partial pressure of water vapour over pure liquid water at temperature T . It can be described with Magnus' formula (Murray, 1967) as

$$p_{w,s}(T) = 610.78 \text{ Pa} \exp \left(\frac{17.2694(T - 273.16 \text{ K})}{T - 35.86 \text{ K}} \right). \quad (2.27)$$

These relations can also be used to calculate the equivalent matric potential from a given water vapour concentration.

Equation (2.25) can be reformulated, using the chain rule, as

$$j_w^g = -V_{m,w} D_w^g \left[\frac{\partial \nu_w^g}{\partial T} \nabla T + \frac{\partial \nu_w^g}{\partial \psi_m} \nabla \psi_m \right]. \quad (2.28)$$

By neglecting the temperature dependent part, approximating the vapour by an ideal gas, and using equation (2.26) one obtains

$$j_w^g = -D_w^g \frac{p_{w,s}(T) V_{m,w}^2 \exp \left(\frac{\psi_m V_{w,m}}{RT} \right)}{[RT]^2} \nabla \psi_m. \quad (2.29)$$

To account for tortuosity ζ , the model

$$\zeta(\theta^\ell) = \frac{\theta^{\ell^2}}{\theta_s^{2/3}} \quad (2.30)$$

is used (Jin and Jury, 1996). The diffusion of water vapour is not hindered as much by liquid films as that of other gases, because it can condensate on one side and re-evaporate on the other. While Philip and de Vries (1957) limit this effect to thermal induced vapour flow, it appears plausible that it also holds for isothermal vapour transport. Therefore the saturated water content θ_s can be used for θ^ℓ in the tortuosity model yielding the relation $D_w^g = \theta_s^{4/3} D_{w,\text{atm}}^g$ where $D_{w,\text{atm}}^g$ is the diffusion coefficient for water vapour in free air. Inserting the tortuosity model into equation (2.29) yields

$$j_w^g = -K^g(\psi_m) \nabla \psi_m \quad (2.31)$$

with

$$K^g(\psi_m) = D_{w,\text{atm}}^g \frac{\theta_s^{4/3} p_{w,s}(T) V_{m,w}^2 \exp \left(\frac{\psi_m V_{w,m}}{RT} \right)}{[RT]^2}. \quad (2.32)$$

Since equation (2.31) has the same form as the Buckingham-Darcy law, Richards' equation (2.19) can be extended to include water vapour transport as effective conductivity:

$$\begin{aligned}\frac{\partial \theta^\ell}{\partial t} &= \nabla \cdot \left[K^g(\psi_m) \nabla \psi_m + K^\ell(\theta^\ell) \nabla \left[\psi_m - \rho_w^\ell g z \right] \right] \\ &\approx \nabla \cdot \left[[K^g(\psi_m) + K^\ell(\theta^\ell)] \nabla \left[\psi_m - \rho_w^\ell g z \right] \right].\end{aligned}\quad (2.33)$$

The neglected term $-K^g(\psi_m) \nabla(\rho_w^\ell g z)$ is small since ...

2.6 Dynamic effects

The water content at a particular potential sometimes differs greatly from the retention curve, especially during rapid flow. The hydraulic relations not only depend on actual system state and the wetting/drying history, but also on rate of system change. A long time is required until the system re-enters equilibrium.

Various authors showed dynamic non-equilibrium in experiments with traditional probes like tensiometers and time domain reflectometry (TDR). Reviews are given by [Durner and Flühler \(2005\)](#); [Šimůnek et al. \(2003\)](#). The dynamic effect depends on the size of pressure changes, it is larger for big changes. It is also larger for soils with wide pore-size distribution. According to ([Durner and Flühler, 2005](#)), the problem of dynamic effects has not been treated adequately yet.

[Davidson et al. \(1966\)](#) made imbibition and drainage experiments, where more water was removed in one large step than in a sequence of smaller steps. [Topp et al. \(1967\)](#) compared $\theta(\psi_m)$ of vertical sand columns. During drainage, at a given ψ_m , θ was higher in transient experiments than in static equilibrium or steady-state conditions. [Smiles et al. \(1971\)](#) conducted a series of imbibition/drainage cycles with horizontal sand columns using stepwise pressure changes and found $\theta(\psi_m)$ to be unique during imbibition but not during drainage. Given an initial $\theta > \theta_{\text{crit}}$, where θ_{crit} is the critical water content, they found $\theta(\psi_m)$ to depend on the pressure step size and on the time required to reach an equilibrium state. No deviations of gas pressure from atmospheric pressure could be detected. [Vachaud et al. \(1972\)](#) made a similar experiment with the same results. In drainage experiments in vertical packed sand columns, [Stauffer \(1977\)](#) found that during dynamic flow, the water content θ was higher than in the static retention curve. He as well found no deviation of gas phase from atmospheric pressure. [Plagge \(1991\)](#); [Plagge et al. \(1999\)](#) conducted evaporation experiments with silty soils and measured ψ_m with micro-tensiometers and θ with TDR in different heights. They found that local $\theta(\psi_m)$ curves, depending on height, differed considerably. The differences were systematic and reproducible. [Lennartz \(1992\)](#) made a systematic investigation with evaporation experiments. He found clear differences, but no clear trends. [Wildenschild et al. \(2001\)](#), in one- and multistep-outflow-experiments, measured $\theta(\psi_m)$ using tensiometers and average water content θ . In coarse-textured media, they found that with a given ψ_m , more water was retained with greater pressure steps, while they found no effect for fine-textured media. In contrary, [Constantz \(1993\)](#), in one- and multistep outflow experiments, found more water with lower pressure steps. The origin of this discrepancy is not clear.

Mechanisms for dynamic effects

One mechanism for dynamic effects is the dynamic contact angle ([Stauffer, 1977](#)). An experiment by Aribert (1970), as described by [Stauffer \(1977\)](#), shows that for two immiscible fluids in a tube, the contact angle depends on velocity. However, [Wildenschild et al. \(2001\)](#) reported that the contribution of dynamic contact angle to dynamic effects is minor, the contact angle change due to velocity was found to be smaller than 30° in any case, sometimes it does not change at

all. The following additional mechanisms were reported: (i) entrapment of water at high flow rates leads to hydraulic isolation of water-filled pores, (ii) pore water blockage: after a pressure step, the lower sample part is drained faster, isolating flow paths, and (iii) air entrapment.

Schultze et al. (1999) found that a considerable portion of the dynamic effects were caused by a non-negligible air permeability and concluded that a two-phase model is needed to reasonably describe dynamic effects. They also pointed out that measuring air pressure is difficult and tensiometer no longer measure ψ_m if the air pressure is not constant and equal to ambient pressure (section 2.4). They stated that a quantification and isolation of two-phase flow, hysteresis, and true dynamic effects is needed for further process.

Models

Stauffer (1977) used an empirical extension of the retention characteristic which depended on the change of water content: $\Delta\psi_m = \psi_{m,dyn} - \psi_{m,stat} = -\alpha\rho_w \frac{\partial\theta}{\partial t}$. The free parameter α was estimated by plotting $\frac{\partial\theta}{\partial t}$ against $\Delta\psi_m/\rho_w$ and fitting a straight line. His data fitted relatively well to that model.

Friedman (1999) developed a model for the dynamic contact angle effect on the hydraulic functions, depending on capillary number $Ca = \frac{\mu v}{\sigma}$. The model showed that for higher Ca , the water content at given ψ_m was lower. However, for reasonable capillary numbers, the effect due to dynamic contact angle was minor. It began to be significant at $Ca \approx 10^{-4}$. This fits with the conclusions of Wildenschild et al. (2001) who reported that the influence of the dynamic contact angle was minor.

The Ross and Smetten (2000) model combined Richards' equation with a kinetic description. They decoupled θ from ψ_m , obtaining two independent variables. The coupling was done with the additional requirement $\frac{\partial\theta}{\partial t} = f(\theta, \theta_{equil})$. f is a known function which describes the equilibration of dynamic features. A linear driving was assumed by choosing $f(\theta, \theta_{equil}) = \frac{\theta_{equil} - \theta}{\tau}$, where τ is the equilibration time. The dynamic water content was numerically approximated using the iterative procedure $\theta^{j+1} = \theta^j + [\theta_{equil}^{j+1} - \theta^j] [1 - e^{-\Delta t/\tau}]$. When decoupling θ and ψ_m , it must be decided whether $K(\theta_{equil})$ or $K(\theta)$ is to be used for conductivity estimation. Ross and Smetten (2000) chose $K = K(\psi_m) = K(\theta_{equil}(\psi_m))$. According to Šimůnek et al. (2003) this means to fill large pores first. Using $K(\theta)$ reduces the non-equilibrium features.

3 Hysteresis

3.1 Definition and characteristics of hysteresis in general

Consider a thermodynamic system with the independent state variables x_i and dependent variables y_i . If the system is taken from a state A to a state B within an equilibrium process, i. e. infinitely slow, along a given path, the dependent variables y_i will run through certain values. If the process is reversed and the system is brought back to A , normally the y_i pass reversely the same values as on the way $A \rightarrow B$ (figure 3.1(a)).

For a certain set of processes however, even if the process is exceedingly slow, the path taken by the y_i from B back to A is different from that on the way $A \rightarrow B$. Two types of such processes can be distinguished: *metastability* and *hysteresis* (Everett and Whitton, 1952). If any part of the process involves *non-stationarity*, it is called metastable. An example for metastability is melting and then freezing a pure substance which exhibits super-cooling. Here, x is temperature and y may be entropy. On the reverse way, the still liquid substance has a temperature lower than the melting point and solidifies instantaneously on small disturbances (figure 3.1(b): when decreasing x , the system jumps from C to D , but on the way back it continues on the upper line until point E , before it jumps down to F). On the other hand if both paths correspond to *stable, reproducible* values of the dependent variables y_i , the process is called hysteretic (figure 3.1(c)).

Hysteresis can be produced if a system consists of a large number N of small sub-systems which exhibit metastability. Then the superposition of the many sub-systems leads to a smooth path. Equivalently, the corresponding thermodynamic potential of the system has several local minima such that, at some points, the state can jump from one minimum to another. This is indicated in figure 3.2. A system features irreversible processes until the thermodynamic potential reaches a minimum (or the entropy a maximum). Thus, if the system is brought to point C on the red line by increasing x , the system jumps to point D and thereby changes its state to a stable minimum. Going back by again decreasing x , the system now follows the blue line.

Figure 3.1(c) shows the so-called “main hysteresis loop”. This loop is encountered if the

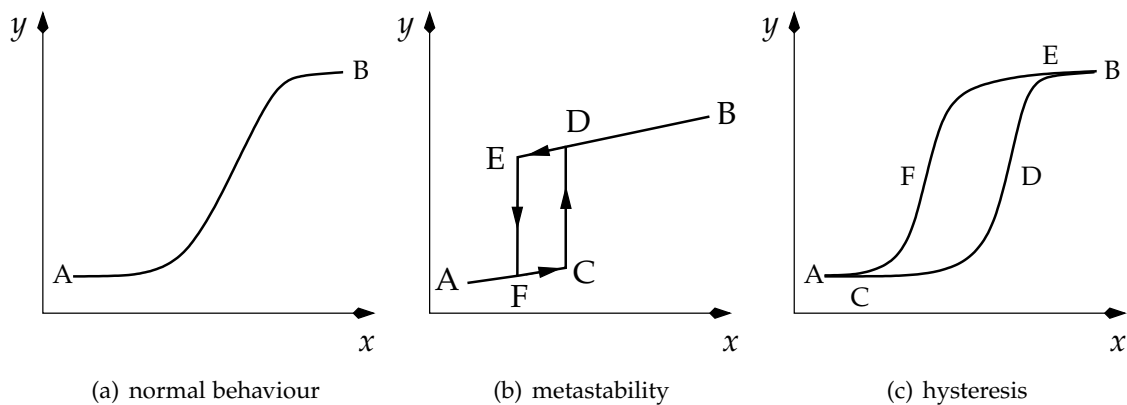


Figure 3.1: Example system loops for a normal system (a), a system exhibiting metastability (b) and hysteresis (c).

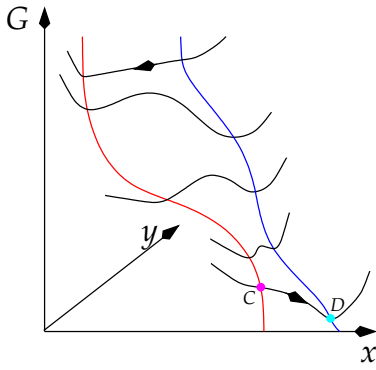


Figure 3.2: Hysteresis caused by non-uniqueness of the thermodynamic potential. If the system is on the red curve and x increased, at point C it jumps to point D and thereby switches to the blue line, where it reaches a new stable minimum.

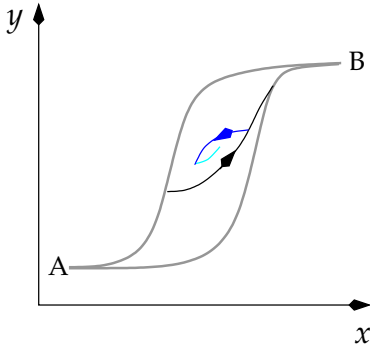


Figure 3.3: First-order (black) and higher-order (blue, cyan) scanning curves of a hysteretic process. The main curve is shown in gray.

system is brought through the whole range of possible values until the branches of the the hysteresis loop again meet each other. If the direction is reverted while the system is still on a branch, the system exhibits so-called *scanning loops* as shown in figure 3.3. These curves are always located inside the main loop. Every point inside the loop can be reached by a quasi-infinite number of possible ways. The simplest form of a scanning curve is a return curve or first-order scanning curve, which is obtained by starting from one end and reversing on a main branch, e. g. at point F in figure 3.1(c). These curves are typically leading towards the other branch (but not necessarily reaching them). If the direction is reverted again somewhere on the scanning curve, higher-order scanning curves are obtained, shown in blue and cyan in figure 3.3.

3.2 Mechanisms for hysteresis in porous media

In general, in a porous medium $\theta(\psi_m)$ will not be a unique function even if the saturation-drainage process is infinitely slow and thus in theory reversible. Indeed, this has already been found by the first experiments in the 1930s. Figure 3.4 shows an example of wetting and drying loops as reported by Topp and Miller (1966).

There are several mechanisms at the micro scale which lead to hysteresis. These are explained in the following.

3.2.1 Inkbottle effect

Look at a single pore in pore space (figure 3.5). Initially, the pore is saturated with water (figure 3.5 (a)). Now, pressure on the left end of the pore is decreased. The water-air interface moves continuously into the pore up to the first minimal radius (b), which corresponds to a certain pressure p_1 according to equation (2.4). Reducing the pressure any further empties the whole cavity, until the interface can form an even smaller radius: the possible radii in the

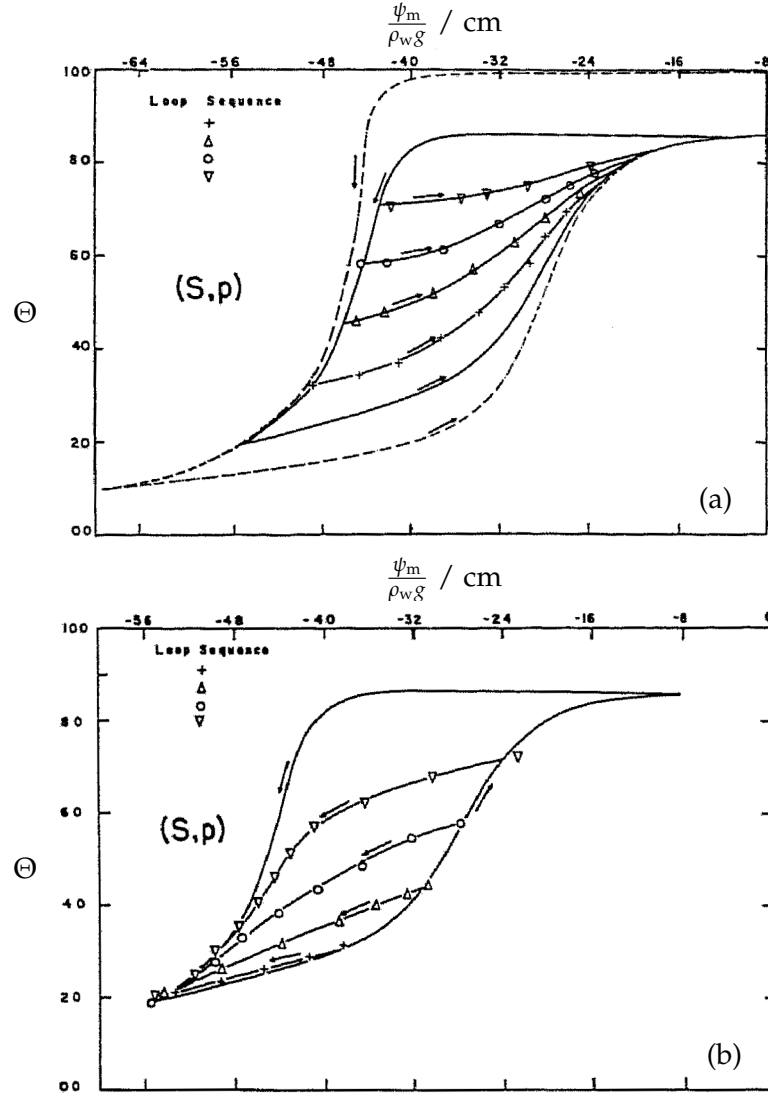


Figure 3.4: Pressure-saturation relation reported by [Topp and Miller \(1966\)](#) for water in thin packing of glass beads with mean diameter of $180 \mu\text{m}$. Water content was measured by γ -absorption and pressure with transducers. (a) shows wetting loops, (b) shows drying loops. This figure was extracted from figures 2 and 3 of the original publication.

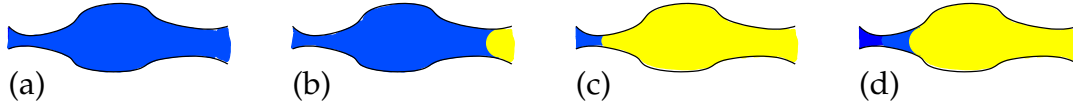


Figure 3.5: Illustration of hysteresis in a porous medium caused by the ink-bottle effect. Water is displayed in blue, air in yellow. Sub-frame (a) shows a completely saturated pore, (b) that pore at the air entry point p_1 , sub-frame (c) at pressure $p_2 < p_1$ and sub-frame (d) at p_1 after re-irrigation, with the same curvature radius as (b), but different water content.

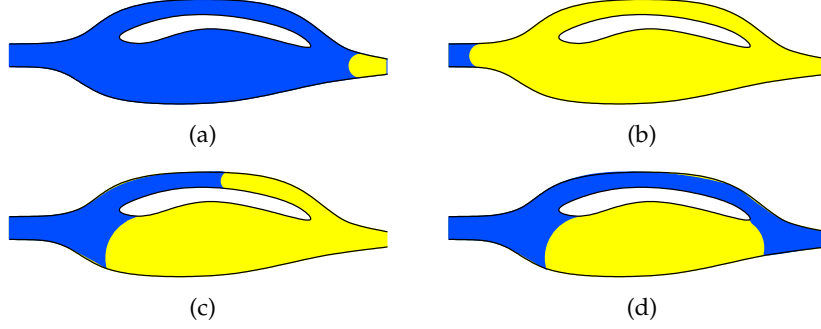


Figure 3.6: Illustration of hysteresis in a porous medium caused by entrapped air. Water is displayed in blue, air in yellow. (a) Both pores are saturated, (b) during drainage the water flows out of all pores, (c) on imbibition water flows faster in the small pore due to higher capillary forces, (d) this leads to air entrapment in the large pore.

cavity are too large to sustain the interface as the pressure difference is too high. This leads to figure 3.5 (c).

At this point the process is reverted and the pressure is increased again. This does not lead back to (b); instead, at p_1 , the configuration (d) will be found, which has the same curvature radius as (b) but a different water content θ . This effect is called the *ink-bottle effect* and is caused by the connection of small and large cavities in pore space.

3.2.2 Entrapped air

On imbibition, smaller pores are invaded first due to higher capillary forces. Therefore, in a system consisting of a large and a small pore (figure 3.6) the water in the small pore reaches the end junction first (sub-frame (c)) and air is entrapped in the large pore (d). The result is entrapped air that is completely enclosed by water and thus does not form a connected phase any more. Thereby, the system has another water content than the fully water-filled one at the same potential (a). Thus, after drying and re-wetting, even if the pressure is set to the same value as before, the water content θ does not reach its initial value again. As a result, the main hysteresis loops change. Draining once more hence results in another $\theta(\psi_m)$ curve, the “secondary drainage branch”, and so on.

On drainage, the entrapped air can hinder the water in the small pore to flow out. When the water recedes, due to the much lower pressure inside the air phase in the bubble, the pressure gradient may not be large enough to move the minisci and the air bubble stays at its place while the water in the large pore recedes. Thereby, parts of the pore space could be “locked up” by air bubbles.

Note that when entrapped air is present, the Richards’ formulation may not be valid any more if the non-continuous air phase drastically changes the flow behaviour such that Darcy’s

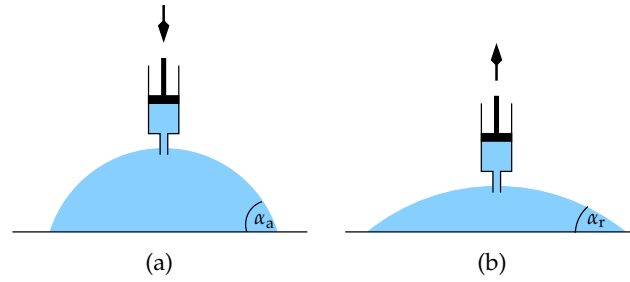


Figure 3.7: Contact angle hysteresis: (a) contact angle when the drop is inflated, (b) contact angle when the drop is deflated.

law is not valid. On the other hand, if the amount of entrapped air is small, the system may still be described with Richard's equation with an effectively smaller pore space available for water transport and therefore slightly modified parameters. Thus, in that case, macroscopically the entrapped air only has an indirect influence.

3.2.3 Contact angle hysteresis

The surface of the solid phase is not completely smooth; there are always disturbances, at least in some molecular level. Hence, the contact angle is not unique any more, the receding contact angle is smaller than the advancing contact angle (de Gennes et al., 2004). For example, if a drop is inflated (figure 3.7(a)), the contact angle α can exceed the equilibrium contact angle, but the line of contact still does not move. Not before the contact angle exceeds a certain threshold α_a , the advancing contact angle, the line of contact begins to move. Similar, in deflating the drop, the line of contact does not move before the contact angle drops below α_r (receding contact angle). This leads to a hysteretic behaviour of the line of contact.

On a “good” surface, $\alpha_a - \alpha_r$ is small ($< 5^\circ$), while on rough and/or dirty surfaces, it can exceed 50° (de Gennes et al., 2004).

3.2.4 Pinning and snap-off

Often a local increase in the wettability is found, i. e. there are spots with a higher wettability than the rest of the surface (e. g. a sand grain). The hysteretic effect of such spots (de Gennes et al., 2004) is shown in figure 3.8. The receding line of contact is moved to the right (a). On encountering the disturbance, it is pinned locally by the higher wettability and stretched (b). Finally, at a certain distance d when the force is high enough, the line breaks off (c). This break-off dissipates energy.

On the way back however, the line of contact moves very close to the disturbance without anything happening (figure 3.8(d)). At some small distance d^* which is much smaller than d , the water nearly instantaneously covers the disturbance (e). This is a very quick process which happens in less than 0.1 s.

3.2.5 Surface roughness

The wettability also depends on surface roughness. As roughness increases, a hydrophilic substance becomes even more hydrophilic, while a hydrophobic one can become literally “super-hydrophobic” (de Gennes et al., 2004). Ustohal et al. (1998) demonstrated that the effective properties of a medium with microscopically heterogeneous wettabilities may be obtained from

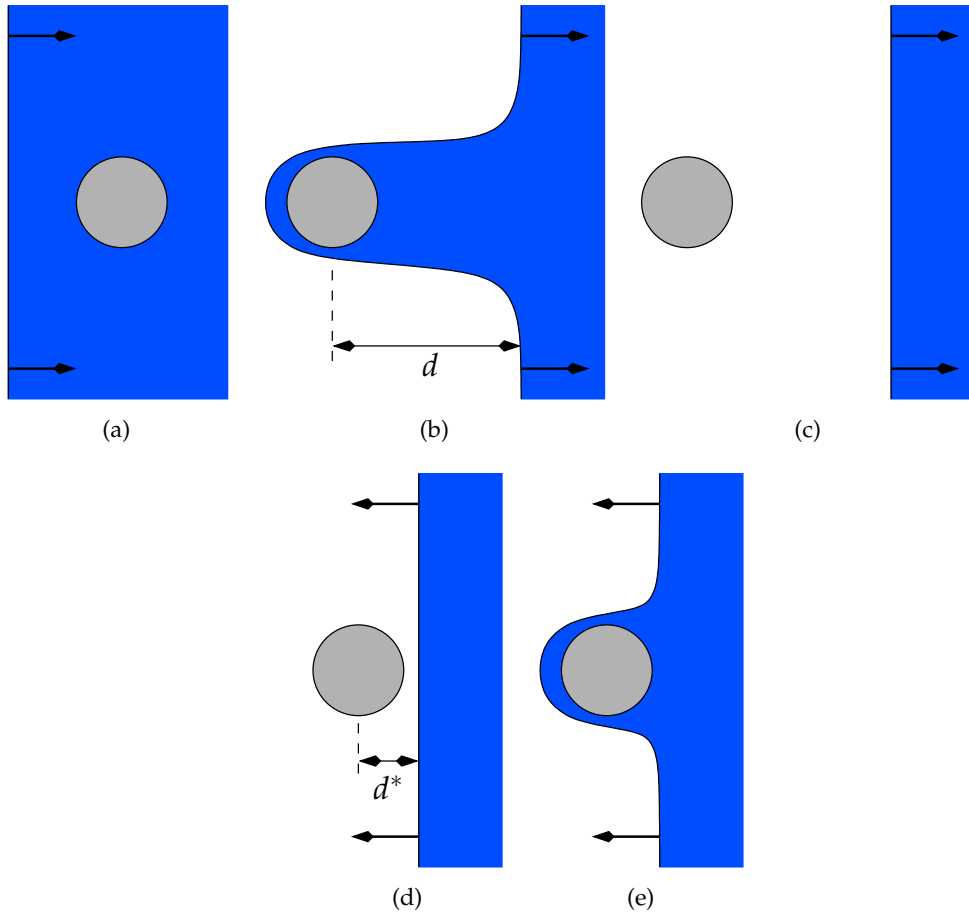


Figure 3.8: Hysteresis caused by pinning of the receding triple line. The top row shows the receding line of contact, the bottom row the advancing one.

a weighted average of the constituents. If the roughness changes on a macroscopic level, different regions of the medium have different properties which might lead to different flow paths on wetting and drainage.

3.3 Modelling hysteresis

Understanding the mechanisms on the microscopic level, hysteresis for $\theta(\psi_m)$ will be expected.

The macroscopic water content is made up of a complex microscopic water configuration. There are many different possibilities how a certain curvature (determined by the matric head which corresponds to a pressure jump) can be realised, corresponding to different water contents θ . At a given value of ψ_m , the change of water content $\frac{\partial \theta}{\partial \psi_m}$ depends on that complicated microscopic configuration. In the non-hysteretic case, the system averages itself when the soil hydraulic properties are estimated experimentally and nothing needs to be known about the microscopic water configuration at any point. This situation changes completely when hysteresis is encountered. At a certain point (ψ_m, θ) where the direction of flow is inverted, the microscopic water configuration at that precise point determines $\frac{\partial \theta}{\partial \psi_m}$. But, in difference to the non-hysteretic case, it may not have been estimated experimentally because it is impossible to run experiments for every single possible hysteresis loop since the number is infinite. Thereby, when modelling the soil, one runs into a fundamental problem. There are several ways to solve

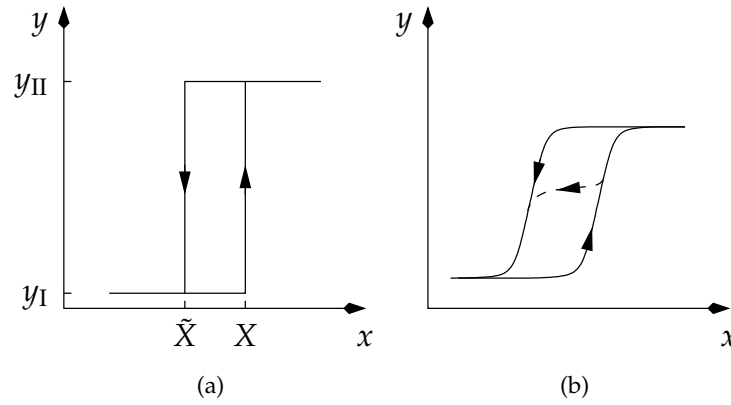


Figure 3.9: Illustration of the independent domain model: (a) The characteristics of an atomic domain and (b) the resulting characteristics of a large number N of independent domains with distributed characteristics. The model can also reproduce scanning curves, shown as the dotted line.

it:

1. Something must be said about the microscopic behaviour of the system. A physical model for the hysteresis process is made and calibrated with experimental data. The major existing models are reviewed below.
2. Empirical parameterisations are used to estimate scanning curves by measured main hysteresis loops, often by scaling and shifting. There is no physical insight obtained from this strategy, so only the results of a selected number of models is compared to experimental data here.
3. The system is explicitly simulated at a microscopic scale and the macroscopic properties are estimated using this simulation. This presupposes that the microscopic structure is measured e. g. using micro-tomography. It is also very costly in terms of computational power and, at the time of writing, no sufficiently functional microscopic model is available.
4. It is claimed that the hysteretic process becomes non-hysteretic again if an additional variable of state, e. g. the area of the interface, is included in the consideration.

3.3.1 The independent domain model

The independent domain model proposed by [Everett and Whitton \(1952\)](#) is a general model for hysteresis and was later adopted to capillary hysteresis by [Poulovassilis \(1962\)](#) and modified by many others, e. g. [Mualem \(1973\)](#). It is based on the assumption that the system consists of many independent binary atomic sub-systems which exhibit metastability, the so-called “domains”, and that the total effect is made up of the sum of the sub-systems. The domains are assumed to have two possible states I and II. It is supposed that the characteristics of the individual sub-systems vary such that the threshold X for the switch from state I to II on the path $A \rightarrow B$ is distributed according to the random distribution p . The threshold \tilde{X} on the reverse path is distributed according to $\tilde{p}(x) = p(x + \xi)$. The offset $\xi = \tilde{X} - X$ is also distributed according to a random distribution p_ξ .

Figure 3.9 illustrates this approach. The atomic domains have a rectangular path. Due to the many domains with different characteristics, every domain switches at a slightly different

value, and a smooth hysteretic curve forms whose shape depends on the distribution functions p and p_ξ .

Applied to a porous medium, this leads to the assumption that the medium consists of a large number of independent pores which switch instantaneously from a complete water-filled to a complete dry state at a potential ψ_d and vice versa at $\psi_i \geq \psi_d$ (“d” stands for drainage, “i” for imbibition). The difference $\psi_i - \psi_d$ is assumed to be independent of ψ_m . ψ_d and ψ_i are distributed according to a distribution function $p(\psi_d, \psi_i)$. p corresponds to the relative water-filled volume. After a sequence of imbibition and drainage processes, the actual water content can be calculated by integrating the distribution function p . An implicit assumption is that the water content only depends on the *last* drainage/imbibition processes and the previous cycles are irrelevant.

Mualem (1973) made the simplifying assumption that p is separable:

$$p(\psi_d, \psi_i) = l(\psi_d)h(\psi_i) . \quad (3.1)$$

This assumption greatly simplifies the calculation. The water content on a wetting curve then becomes

$$\theta_w(\psi_m) = \int_{\psi_{\min}}^{\psi_m} h(\psi_i) \left[\int_{\psi_{\min}}^{\psi_i} l(\psi_d) d\psi_d \right] d\psi_i \quad (3.2)$$

where ψ_{\min} is the minimal possible water content. The drainage curve is calculated similarly. l and h can be estimated using measured primary imbibition and drainage data.

Mualem (1977) simplified the model even more by assuming

$$l(\psi) = h(\psi) , \quad (3.3)$$

thus $p(\psi_d, \psi_i) = h(\psi_d)h(\psi_i)$. This additional simplification allows to use only either the main drainage or the main imbibition loop to calibrate the model, but means that the distribution of bottlenecks is equal to the distribution of cavities, which is not obvious.

3.3.2 The dependent domain model

The dependent domain theory is an extension of the independent domain theory presented in section 3.3.1 above. It was proposed by Enderby (1956) and accounts for the fact that if a domain i changes its state from (say) 1 to 2, it may either aid or hinder the transition of neighbouring domains. For example, in magnetic hysteresis, if several domains are already aligned according to the external field and exhibit an additional magnetic moment, it exerts an aligning tendency on the neighbouring domains. In capillary hysteresis, if one pore has been drained, it can block the neighbouring pore because its water cannot flow out through the drained pore.

As in the independent domain model, the domain i changes from state I to II at $x = X_i$ and from II to I at $x = \tilde{X}_i$. However, for a given domain, X and \tilde{X} will not be fixed but will depend on the state of the neighbouring domains. They can either increase or decrease on a switch of a neighbour, depending on the interaction type (ease or hindrance).

Consider a one-dimensional array of domains (figure 3.10). Suppose that a switch of a domain eases the switch of a neighbouring domain. When x is increased (path $A \rightarrow B$), X_i is reduced if domain $i - 1$ switches from I to II and reduced even more if domain $i + 1$ switches as well. Thus the transition of the neighbours might reduce X_i below the current x and domain i may switch too (figure 3.10(d)). This may transform another neighbour and so on, such that

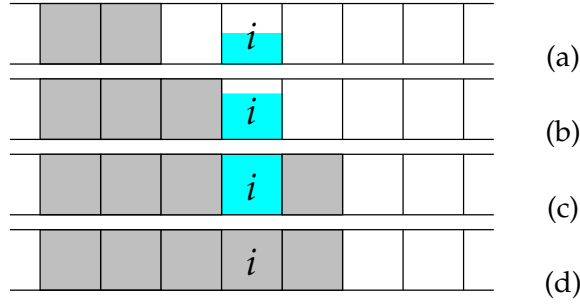


Figure 3.10: Illustration of the dependent domain theory with a linear chain. Gray fields are supposed to be in state II. The cyan filling in domain i represents how far it is from switching, i.e. $X_i - x$. On increasing x , when the neighbouring domains switch from I to II, X_i is reduced, the cyan area increases.

a chain of transformations might be started. It ends when a domain is reached which does not switch even if the neighbour has already transformed.

A dependent domain model for capillary hysteresis was introduced by [Topp \(1971\)](#) and changed and improved by many other researchers (e. g. [Poulovassilis and Childs, 1971](#); [Mualem and Dagan, 1975](#); [Mualem, 1984](#)). The base idea is to introduce weight factors, $q_d \leq 1$ which denotes the volume ratio of actually drained pores and the drained pores in the independent domain model, and similar q_i for the imbibed pores. They account for blockage by water and air. Thus, the model is not strictly a dependent domain model in the way suggested by [Enderby \(1956\)](#). A matrix-like notation is introduced which denotes the history of imbibition (from bottom to top) and drainage (from top to bottom) processes. $\begin{pmatrix} \psi_1 & \psi_2 \end{pmatrix}$ thus denotes an imbibition from ψ_1 to ψ_2 . The water content calculation now becomes

$$\theta(\psi_m) = \int \int p(\psi_i, \psi_d) q_i(\psi_i, \psi_d, \theta, \psi_m) q_d(\psi_i, \psi_d, \theta, \psi_m) d\psi_i d\psi_d \quad (3.4)$$

which is an integral equation for θ . As this is difficult to calculate, averages

$$\bar{q}_i(\theta) = \frac{\theta \begin{pmatrix} \cdots & \psi_m \\ & \psi_1 \end{pmatrix} - \theta \begin{pmatrix} \cdots & \psi_1 \end{pmatrix}}{\theta_0 \begin{pmatrix} \cdots & \psi_m \\ & \psi_1 \end{pmatrix} - \theta_0 \begin{pmatrix} \cdots & \psi_1 \end{pmatrix}} \quad (3.5)$$

and $\bar{q}_d(\theta)$ (similar) are used. The dots \dots denote any number of previous processes, θ_0 is the water content without blockage, as calculated by the independent domain model. Equation (3.5) therefore is the ratio of the real water content change introduced by the last imbibition process and the water content change of the same process as given by the independent model. The \bar{q}_j are assumed to depend on θ only. As a further simplification, blockage against air entry is assumed to be negligible, $q_d = 1$.

In the [Mualem and Dagan \(1975\)](#) model, experimental data of a scanning curve is needed in addition to the main wetting and drainage curves to obtain q_i . The [Mualem \(1984\)](#) model is based on the extended similarity hypothesis equation (3.3) and only needs the primary hysteresis curves for calibration.

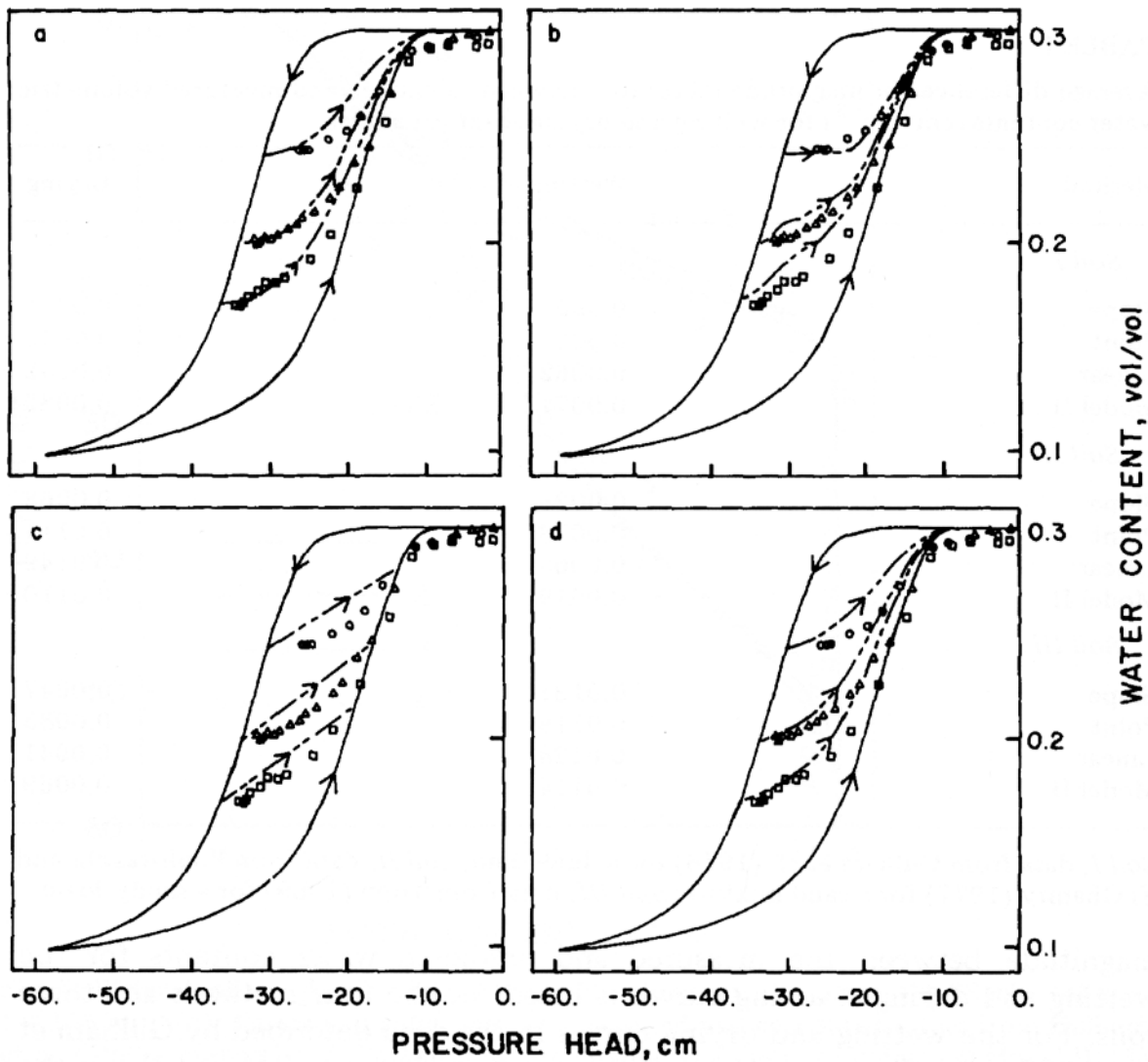


Figure 3.11: Comparison of the performance of four hysteresis models in reproducing wetting scanning curves by Jaynes (1984). Solid lines (main loop) and points are measured, broken lines are fitted values. (a) slope, (b) point, (c) linear method, (d) Mualem and Miller (1979) model. This figure was extracted from figure 4 of the original publication.

3.4 Model performance

Although the hysteresis algorithms mostly were tested in the context of the flow problems for which they were developed, systematic comparisons of several models with the same data are rare. Jaynes (1984) compared four different hysteresis models using experimental data by Gillham et al. (1976). Three of them are empirical models: (a) a point method (Dane and Wierenga, 1975), (b) a linear method (Hanks et al., 1969), (c) slope method by the author of the comparison himself. The fourth model (d) is the domain model of Mualem and Miller (1979).

Figure 3.11 shows the ability of the four models to reproduce wetting scanning curves. All models were calibrated with the main loops. All models predict the measured data poorly. Parts of the lines are far off the measured values and partly even do not get the shape correctly.

Lehmann et al. (1998) measured the response of a sand column to a fluctuating water table. Water content was measured with TDR probes and the matric potential with tensiometers. They

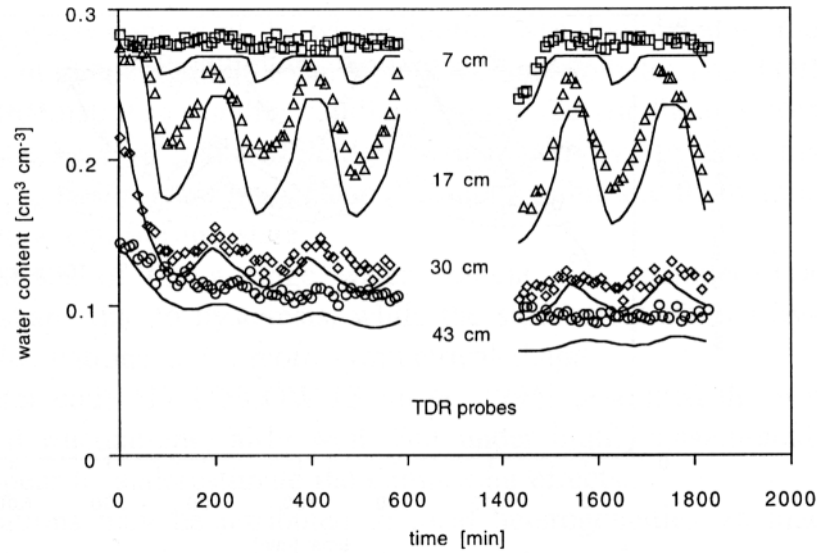


Figure 3.12: Simulated (lines) and experimental (symbols) data at four locations of flow with cyclic water table fluctuations by [Lehmann et al. \(1998\)](#). This figure was extracted from figure 12 of the original publication.

compared the results with simulations using the [Mualem \(1984\)](#) model. The main hysteresis loop was measured directly with the same experimental setup using quasi-static conditions. Figure 3.12 shows the result of one of the experiments. The agreement between experiment and simulation is poor and partly not even qualitatively correct. For example, the curve at 7 cm has large asymmetric fluctuations which are not seen in the data, the 43 cm simulation is even anti-cyclic to the measurement. [Lehmann et al. \(1998\)](#) attribute parts of the discrepancy to uncertainty in the estimation of the main loop. This is certainly correct, as the direct method applied there is not very precise. All the same the aberration seems too large to be explained by deviations of the main loop only.

[Jaynes \(1984\)](#) also investigated the so-called “capillary pumping” of the four models. If ψ_m cycles between two fixed values, some algorithms predict progressively decreasing water content, until a stable loop is found. This is claimed to be an artefact and no real property of porous media ([Klute and Heermann, 1974](#)). However, entrapped air would have exactly such an effect, which is also indicated by newer experiments.

Models (a) and (b) (as defined at the beginning of the section) exhibit distinct pumping behaviour and also partially crossing scanning curves which is definitely incorrect. Model (c) was explicitly developed to avoid the pumping effect which was achieved fairly well. However it seems doubtful if in reality the scanning curves would stay on a nearly straight line. Model (d) shows a more realistic behaviour, as a parting of the paths in the two directions would be expected.

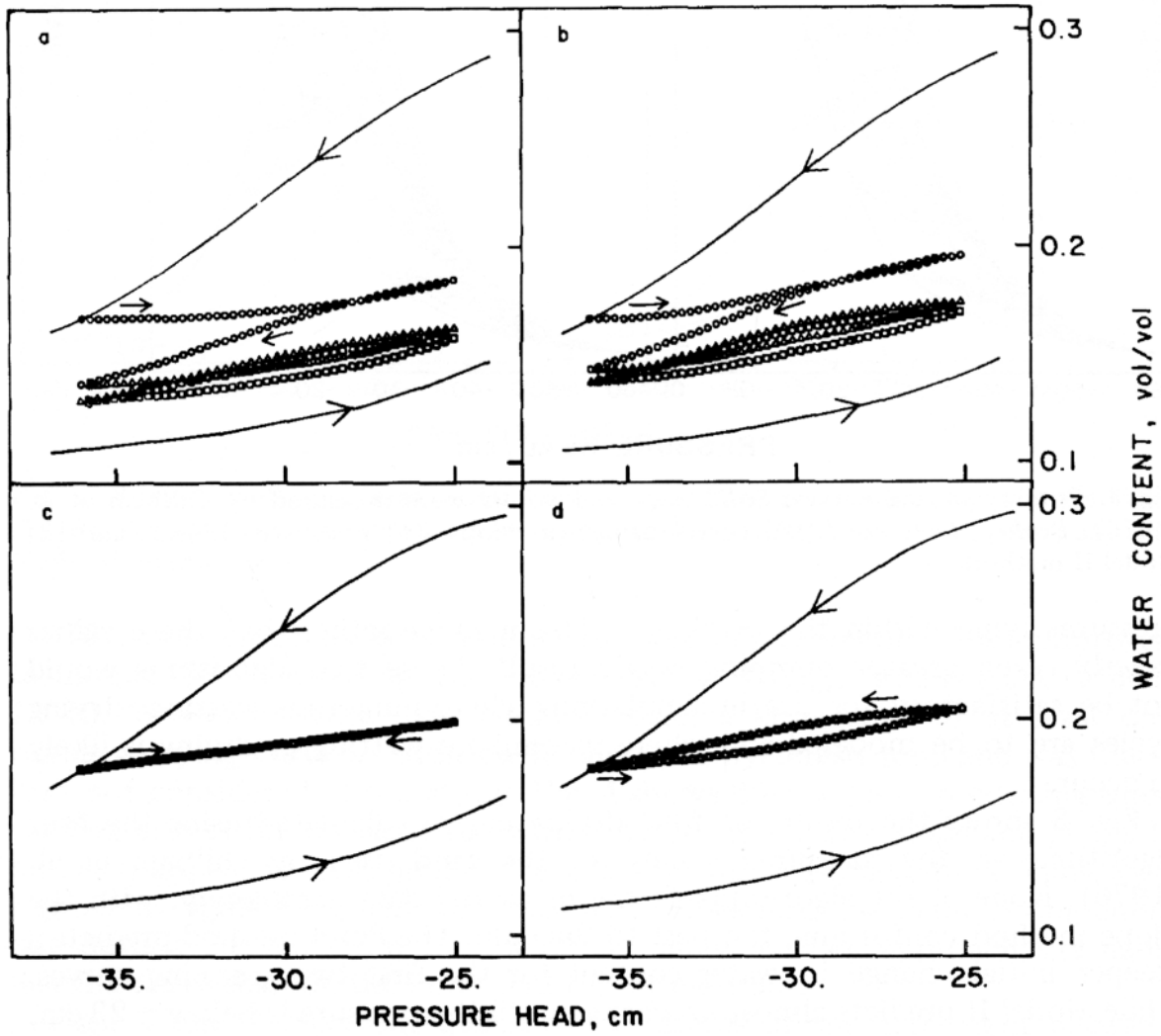


Figure 3.13: Comparison of the capillary pumping behaviour of four hysteresis models by Jaynes (1984). Solid lines are the measured main loop, symbols represent the models in the first (\circ), second (\triangle) and third (\square) cycle. This figure was extracted from figure 2 of the original publication.

4 Image analysis

This is a short introduction into the methods of image analysis used in this thesis. A comprehensive treatment can be found e. g. in [Jähne \(2005\)](#).

4.1 Image representation

Images constitute a spatial distribution of the irradiance. In the real physical world, at a macroscopic scale this is a continuous function $E(x)$. However, as computers can only handle finite arrays of discrete numbers, the irradiation must be *discretised* to an array of points. This process is called *sampling* and is described below. A digital camera for example samples the data in its photo-sensor, which has a discrete number of photo diodes. A discrete 2D image is represented by an $M \times N$ matrix, a point in that 2D grid is called pixel. Its position is given in the matrix notation: g_{ij} denotes the pixel in the i th row and the j th column. Note that images or signals are indexed beginning at 0, the first entry of a 2D image is g_{00} . In contrast, dimensions are indexed beginning at 1, e. g. in a 2D space x_1 would be the x-axis and x_2 the y-axis.

$M \times N$ dimensional 2D images can be considered as an $M \times N$ dimensional vector space with a vector product

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i,j=0}^{N-1,M-1} x_{ij} y_{ij} . \quad (4.1)$$

A natural basis of this vector space is

$$({}^m B)_{ij} = \delta_{ni} \delta_{mj} \quad (4.2)$$

where δ_{ij} denotes Kronecker's delta. An image can be transformed into another basis without loss of information. Another important vector space for images is Fourier space, which is introduced in section [B](#).

4.1.1 Homogeneous coordinates

For any N -dimensional vector space, affine transformations can be written as

$$\boldsymbol{\xi} = A\mathbf{x} + \mathbf{b} . \quad (4.3)$$

However, it would be favourable to write this transformation as a simple matrix multiplication. This would also allow to concatenate several transformations into one matrix. It is achieved by introducing homogeneous coordinates. For that, the dimension of the vector space is increased by 1 as follows:

$$\mathbf{x} = (x_1, x_2, \dots, x_N) \mapsto \tilde{\mathbf{x}} = (x_1, x_2, \dots, x_N, 1) . \quad (4.4)$$

The additional component may only have the values one or zero. If after a transformation the additional vector element is not equal to one or zero, the vector is re-scaled such that that element again equals one:

$$(x_1, \dots, x_N, w) \mapsto \left(\frac{x_1}{w}, \dots, \frac{x_N}{w}, 1 \right) . \quad (4.5)$$

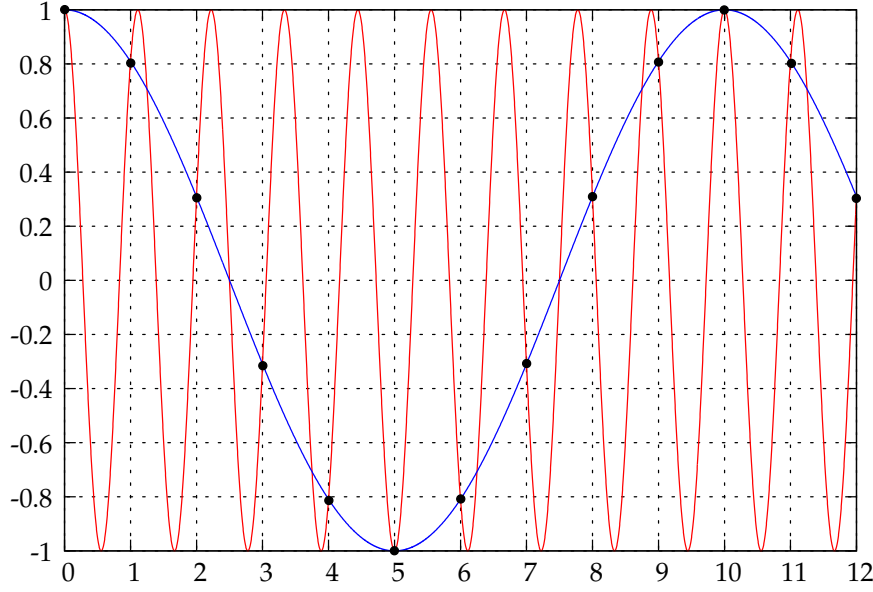


Figure 4.1: Demonstration of the aliasing effect: a 1D signal is sampled with a sampling distance Δx of $9/10$ of the wavelength of the original (red) signal. The result is an aliased signal with an apparent wavelength of 10 times the sampling distance (blue).

With this new vector space, affine transformations are written as

$$\tilde{\xi} = \tilde{A}\tilde{x}, \quad (4.6)$$

where the $N + 1 \times N + 1$ matrix \tilde{A} implicitly contains b .

4.2 Sampling and sampling theorem

It is intuitively clear that sampling leads to a reduction in resolution. Structures equal or finer than about the scale of the sampling distance is lost. Considerable distortions may occur when the sampled object contains finer structures. This is demonstrated in figure 4.1 which shows a sinusoidal signal which is sampled with a sampling distance Δx slightly smaller than the wavelength. As a result a signal with a much larger wavelength is observed. This effect is called aliasing or Moiré effect.

Thus it is important to understand the sampling process and how to avoid the loss of information. Since the aliasing effect was demonstrated with a periodic signal, it is clear that the key is to analyse the sampling process in Fourier space.

Consider the digitisation of the continuous irradiance field $E(x)$ to a discrete grid

$$r_{ij} = (i\Delta x_1, j\Delta x_2) \quad \text{with } i, j \in \mathbb{Z}, \quad (4.7)$$

where Δx_i is the spacing in i -direction of the rectangular grid. Each pixel in (for example) the photo-sensor has a finite sensitive area which averages the radiation according to its specific point spread function $h_p(x)$. Assuming that h_p is equal for all pixels one obtains

$$g(r_{ij}) = \int_{-\infty}^{\infty} E(\xi) h_p(r_{ij} - \xi) d\xi. \quad (4.8)$$

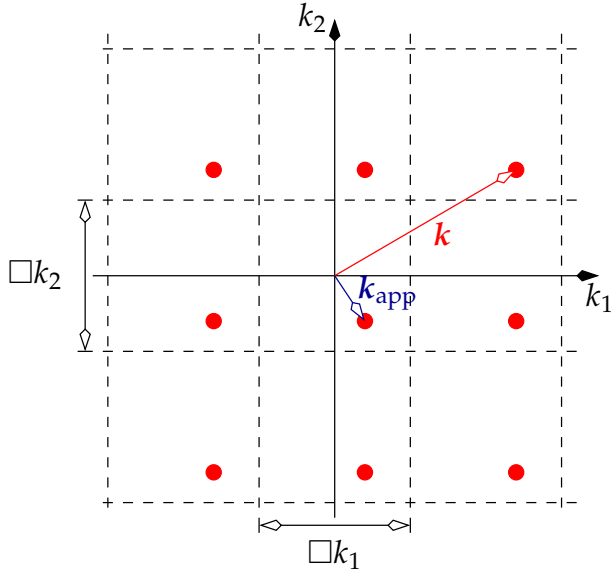


Figure 4.2: Explanation of the Moiré effect with a periodic structure that does not satisfy the sampling condition. The original spectrum contains a single peak marked with the vector k . Because k is too large, in the central cell only a periodic copy is found (wave vector k_{app}), which has another length and direction than the original vector.

Thus, E is convolved with the PSF of the pixels and then sampled at the grid points. Accordingly, these two processes can be treated separately. As convolved is an associative process, we can combine the PSF of the optical system h_o (lens, shutter, etc.) with the PSF of the pixels h_p to a total PSF h .

Sampling means that all information which is not on a grid point is lost, i. e. the function g is multiplied with the “ δ -comb”, which has δ -functions on the grid points and is zero on all other points:

$$g_s(x) = g(x) \sum_{i,j} \delta(x - r_{ij}) \quad \circ \bullet \quad \hat{g}_s(k) = \sum_{i,j} \hat{g}(k - \hat{r}_{ij}), \quad (4.9)$$

where

$$\hat{r}_{ij} = (i\Delta k_1, i\Delta k_2) \quad \text{with } i, j \in \mathbb{Z} \text{ and } \Delta k_l = \frac{1}{\Delta x_l} \quad (4.10)$$

is the reciprocal grid. In Fourier space, g is convolved with another δ -comb whose grid constants are reciprocal to the grid constants in position space. A dense mesh in position space therefore yields a coarse mesh in Fourier space and vice versa.

As mentioned in Appendix B, a result of the Fourier transform is that images are periodic. If the image spectrum is so large that parts of it overlap with the periodically repeated copies, it cannot be distinguished whether the spectral amplitude comes from the original spectrum or from its copy. This is illustrated with the periodic structure in figure 4.2. The spectrum contains one peak. As the wave number is too large for the sampling grid, the central cell only contains a periodic copy of the peak. Thus, the observed apparent wave vector k_{app} differs from the true one by a grid translation vector \hat{r}_{ij} on the reciprocal grid. It has both a wrong length and direction. If there were other structures, they would overlap with k_{app} and thereby create a so-called Moiré or aliasing effect. In the one-dimensional example in figure 4.1, one obtains an aliased wave number of $k_{\text{app}} = k - \Delta k = 9/10\Delta k - \Delta k = -1/10k$.

Therefore, overlapping must be avoided. This can be achieved by limiting the spectrum to the area that extends up to one element of the reciprocal grid (the first dashed lines around a grid point in figure 4.2), i. e. to the first Brillouin zone. Thus, a periodic structure will only be obtained correctly if at least two samples per wavelength are taken (Jähne, 2005):

Theorem 4.1 (Sampling theorem). *If the spectrum $\hat{g}(\mathbf{k})$ of a continuous function $g(\mathbf{x})$ is band-limited such that*

$$\forall i : \forall |\mathbf{k}_i| \geq \square \mathbf{k}_i / 2 : \hat{g}(\mathbf{k}) = 0 , \quad (4.11)$$

then it can be reconstructed exactly from samples with the distance $\Delta x_i = 1 / \square \mathbf{k}_i$.

The maximum wavenumber which can be sampled without errors with a given grid is called *Nyquist* wave number. It is often convenient to use dimensionless wave numbers which are scaled to the limiting wave number:

$$\tilde{\mathbf{k}} = \frac{\mathbf{k}}{\square \mathbf{k} / 2} = 2\mathbf{k}\delta\mathbf{x} \quad (4.12)$$

Reconstruction from samples

The sampling theorem states the conditions under which a continuous signal can be reconstructed from sampled points. The reconstruction is performed by *interpolation* of the sampled points. The interpolated points are calculated from the weighted sampled values $g(\mathbf{r}_{ij})$:

$$\begin{aligned} g_r(\mathbf{x}) &= \sum_{i,j} h(\mathbf{x} - \mathbf{r}_{ij}) g_s(\mathbf{r}_{ij}) \\ &= \sum_{i,j} \int_{-\infty}^{\infty} h(\mathbf{x} - \boldsymbol{\xi}) g_s(\boldsymbol{\xi}) \delta(\mathbf{r}_{ij} - \boldsymbol{\xi}) d\boldsymbol{\xi} \\ &= \int_{-\infty}^{\infty} h(\mathbf{x} - \boldsymbol{\xi}) \sum_{ij} \delta(\mathbf{r}_{ij} - \boldsymbol{\xi}) g_s(\boldsymbol{\xi}) d\boldsymbol{\xi} \rightsquigarrow \hat{g}_r(\mathbf{k}) = \hat{h}(\mathbf{k}) \sum_{\mu,\nu} \hat{g}_s(\mathbf{k} - \hat{\mathbf{r}}_{\mu\nu}) \end{aligned} \quad (4.13)$$

The interpolated function is only equal to the original one if the weighting function h is a box function with the size of the elementary cell of the reciprocal grid. Then all replicated and shifted spectra are eliminated and the original band-limited function is reproduced.

The inverse Fourier transform of the box function is the *sinc* function. It only decreases with $1/x$, therefore a correct interpolation requires an infinitely large signal. This condition can be weakened if the sampling theorem is “more than fulfilled”, i.e. $\hat{g}(\mathbf{k})$ becomes zero before the Nyquist zone is reached. Then, according to equation (4.13), $\hat{h}(\mathbf{k})$ can be chosen arbitrarily in the region where \hat{g} is zero. This freedom can be used to construct an interpolation function that has a minimum-length interpolation mask.

4.3 Neighbourhood operations

Neighbourhood operations relate pixel values in a small region. They are used for feature extraction.

4.3.1 Averaging

Gaussian averaging

In the most simple case averaging filter can be used to identify regions with constant illumination. Noise or inhomogeneous background can be detected. Convolution is the basis for a large number of average filters. These filters should be shift invariant, preserve the mean

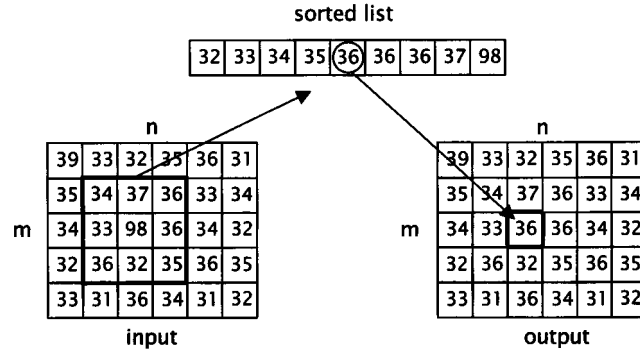


Figure 4.3: Illustration of the operation of the median filter with $N = 3$. All values in the 3×3 neighbourhood of the pixel to be processed are sorted. The middle value is taken as new pixel value. Image source: Jähne (2005)

value and (in most cases) should be isotropic. The transfer function is expected to decrease monotonously.

A reasonable averaging filter is the (linear) binomial filter. Its transfer function is a cosine to the power of R , where R is the order of the binomial used as the discretisation for the Gaussian. For $R = 2$ the 1D filter mask is $b = [1/4, 1/2, 1/4]$ and the transfer function $\hat{b} = \cos^2(\pi\tilde{k}/2)$, where \tilde{k} is the normalised wave number. Since the Gaussian is separable, multi-dimensional filters result from convolution of horizontal and vertical 1D filter. For broad averaging, a multi-scale representation is useful (see Jähne (2005, sect. 11.5, sect. 5)).

Median filter

Linear filters blur edges and mix objects and background. They effectively eliminate Gaussian noise, but they are not efficient for binary (“salt and pepper”) noise. This is eliminated by the median filter. It is invariant for constant neighbourhood and edges, but eliminates impulse-like pixels (outliers). The filter works in the following way: All pixel values in the N neighbourhood are sorted linearly and the middle value (the so-called median) is then used as new pixel value. This process is illustrated in figure 4.3. The median filter is a non-linear filter.

4.3.2 Edge detection and derivatives

Edge detection requires neighbourhood operations which detect gray value changes while ignoring constant gray values. Thereby an image is generated which marks changes, while all other regions are dark. Ideally, an edge is a spatial discontinuity of the gray value. In practice, e.g. noise smears out the edges. Edges can be detected by derivatives. Since images are discrete, the derivative has to be approximated by finite differences. For not breaking symmetries, a symmetric difference approach is best suited. It is defined by the filter mask $[1/2, 0, -1/2]$. When using two element masks, the grid is shifted by half a pixel. Due to the discrete nature the gradient filter exhibits anisotropy of direction and an error in the direction of the gradient which must be kept in mind when using those filters.

4.4 Noise model for image sensors

Digital image sensors are semiconductor solid state devices. In each pixel, the photons interact with the semiconductor. Electrons are released due to the photo effect. Let N_{ph} be the number of photons which hit one pixel. Due to the statistical nature of light, the number of photons

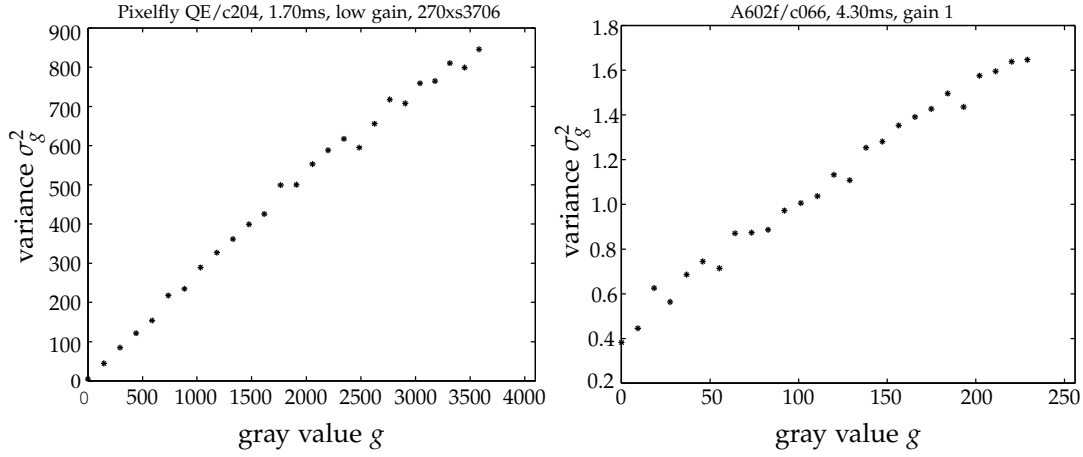


Figure 4.4: Variance versus gray value of two scientific cameras. The measured data fits well to the simple noise model. *This figure was extracted from Jähne (2005).*

is Poisson distributed. For large N_{ph} , the Poisson distribution converges towards a Gaussian distribution with mean N_{ph} and standard deviation $\sigma_{\text{ph}} = \sqrt{N_{\text{ph}}}$.

Let η be the quantity yield, i. e. on average one photon releases η electrons ($\eta < 1$). Then the number of electrons, N_e , is $N_e = \eta N_{\text{ph}}$ and the standard deviation is $\sigma_{N_e} = \sqrt{N_e}$. A typical CCD pixel element collects about 10000 or a bit more electrons. Thus, even with ideal electronics the error is already 1 %.

The read-out electronics, amplification etc. add additional sources of noise. It can be assumed that all these are Gaussian distributed independent random variables. Thus, the signal Q is

$$Q = N_e + Q_0 \quad (4.14)$$

with variance

$$\sigma_Q^2 = \sigma_{N_e}^2 + \sigma_{Q_0}^2. \quad (4.15)$$

The amplifier is assumed to be linear. Thus, the gray value g is

$$g = KQ \quad (4.16)$$

and its variance

$$\sigma_g^2 = K^2 \sigma_Q^2 = K^2 \sigma_{Q_0}^2 + K^2 N_e = K^2 (\sigma_{Q_0}^2 - Q_0) + K^2 Q = \sigma_0^2 + Kg. \quad (4.17)$$

This equation predicts a linear increase of the variation with the measured signal. Normally, image sensors fit quite good to that simple noise model. Figure 4.4 shows experimental data of two scientific cameras (Jähne, 2005).

4.5 Deconvolution

As mentioned above, during image acquisition the true irradiation is convolved with the PSF of the acquisition system. Additionally, there might be convolution in the physical system being investigated, for example according to multiple scattering. As convolution is linear, these can be put into one single PSF p . The spreading of an initially infinitesimally small light beam by multiple scattering in a porous medium is shown schematically in figure 4.5.

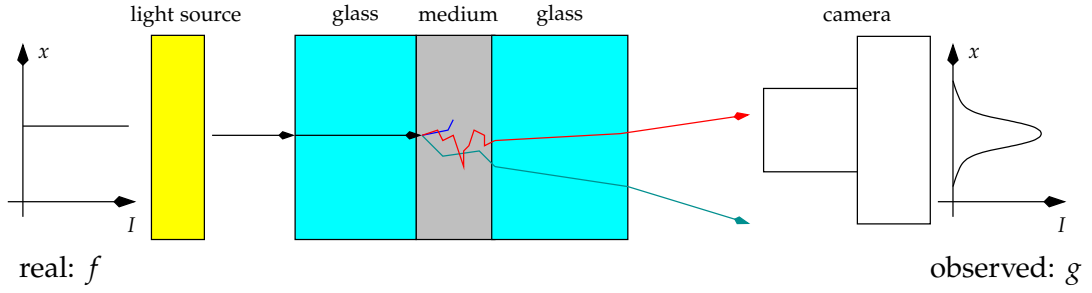


Figure 4.5: Schematic illustration of the spreading of a point source by multiple scattering in a porous medium.

p can also be interpreted as a PDF: $p(x - \xi)$ is the probability that, presumed that a photon was emitted at location ξ , it will be detected at location x . Accordingly, it is a conditional probability, which can be written $p(x|\xi)$ (read: the probability that $X = x$ provided that $\Xi = \xi$, where X and Ξ are random variables with samples x and ξ).

The problem is the restoration of the original (or true) image f from the observed blurred image g with the known PSF p . This approach is called deconvolution or inverse filtering. At first glance, it might seem easy: Using the convolution theorem, equation (B.13), the convolution $\hat{g}(\mathbf{k}) = \hat{f}(\mathbf{k})\hat{p}(\mathbf{k})$ can be readily inverted, obtaining

$$\hat{f}(\mathbf{k}) = \frac{\hat{g}(\mathbf{k})}{\hat{p}(\mathbf{k})} = \hat{g}(\mathbf{k})\hat{p}^{-1}(\mathbf{k}) \quad (4.18)$$

meaning the observed image is simply multiplied by the inverse of the optical transfer function (OTF), \hat{p}^{-1} . However, additional noise n during the measurement process will also affect the observed image. As the absolute value of the optical transfer function is very small in large regions, the numerical division leads to very large numbers at high wave numbers, which amplifies noise.

For a quantitative analysis the simple image formation model

$$g(x) = [f * p](x) + n(x) \quad (4.19)$$

can be used which states that the noise is added to the image *after* the convolution: The noise largely stems from the sensor and the Poisson nature of the collected photons. In the following, this model is referred to as “model I”. In real images, as described in section 4.4, the variance depends on the gray value. Therefore, to make $n(x)$ an independent noise term which does not depend on the actual image, the image must be transformed with a bijective gray value transformation to obtain constant, gray value independent noise:

$$h(g) = \frac{K}{2} \sqrt{Kg + \sigma_0^2}, \quad (4.20)$$

where K and σ_0 are the slope and offset obtained from the camera calibration, as defined in equation (4.17). Using error propagation results in

$$\sigma_h^2 = \left[\frac{\partial h}{\partial g} \right]^2 \sigma_g^2 = \left[-\frac{1}{\sqrt{Kg + \sigma_0^2}} \right]^2 [Kg + \sigma_0^2] = 1, \quad (4.21)$$

thus after the transformation the noise is constant with variance 1 and the image fits to the noise model (4.19). After deconvolution the image is transformed with the inverse transformation h^{-1} to re-obtain the original values.

As an alternative to model (4.19) where the noise is included explicitly,

$$g(x) = [f * p](x) \quad (4.22)$$

might be interpreted as one sample of the random variable G . Here, the noise is implicitly included in the stochastic nature of g . Knowing the PDF of G , we can obtain the image f with maximal likelihood. This approach, in the following referred to as “model II”, is used in the Richardson-Lucy deconvolution, section 4.5.3

Being an inverse problem, deconvolution is also an ill-posed problem, and no unique solution can be found, especially in the presence of noise (Magain et al., 1998). This is due to the fact that many light distributions are, after convolution, compatible with the observed image. Therefore, regularisation techniques have commonly used to put constraints on the reconstructed image, e. g. smoothness constraints. Care must also be taken that the sampling theorem is still fulfilled in the deconvolved data (Magain et al., 1998).

4.5.1 Determination of the PSF

The most straight-forward way to determine the PSF would be to measure the response to a point light source. This method is frequently used in astronomy; due to their large distance, single objects have in good approximation an angular extend of zero. However, in light transmission experiments infinitely small light sources are difficult to achieve. Another possibility is a line source to obtain the PSF in one dimension. Measurements along all dimensions are necessary to obtain the full PSF if no symmetries can be exploited. But again, infinitely small light sources are experimentally difficult.

To overcome the experimental limitations of δ light sources, edges can be used. Mathematically, an edge can be described with Heaviside’s function,

$$h(x) = \begin{cases} 1 & , x > 0 \\ 0 & , x \leq 0 \end{cases} . \quad (4.23)$$

If a slit is positioned vertically between light source and object, the resulting irradiance is

$$f(x, y_0) = f_0[h(x - x_1) - h(x - x_2)] \quad (4.24)$$

assuming a homogeneous light source with irradiance f_0 , where x_1 and x_2 are the x coordinates of the edges of the slit. The observed image is then $g(x) = (f * p)(x)$. By derivating g in respect to x at a fixed y_0 , one obtains

$$\begin{aligned} \frac{\partial g(x)}{\partial x} &= \frac{\partial}{\partial x} \int_{-\infty}^{\infty} f(\xi) p(x - \xi) d\xi = \int_{-\infty}^{\infty} \left[0 + f(\xi) \frac{\partial p(x - \xi)}{\partial x} \right] d\xi \\ &= - \int_{-\infty}^{\infty} f(\xi) \frac{\partial p(x - \xi)}{\partial \xi} d\xi \\ &= \int_{-\infty}^{\infty} \underbrace{\frac{\partial f(\xi)}{\partial \xi}}_{=\delta(x-x_1)-\delta(x-x_2)} p(x - \xi) d\xi - \underbrace{[f(\xi)p(x - \xi)]_{\xi=-\infty}^{\infty}}_{=0} \\ &= p(x - x_1) - p(x - x_2) \end{aligned} \quad (4.25)$$

Thus, the PSF can be determined by derivating the measured images of slits positioned in all directions. When some knowledge about the expected shape of the PSF is available, an

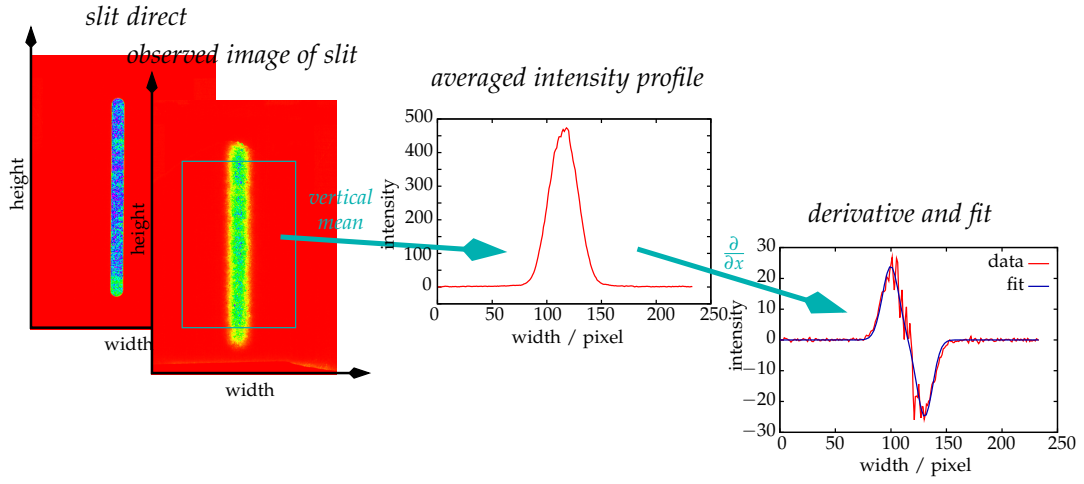


Figure 4.6: Sketch of the determination of the horizontal component of the PSF using a slit, assuming a Gaussian shaped PSF.

analytical function can be fitted to the derived data to obtain an analytical expression of the PSF. The advantage is that noise which is always present in the derived image does not disturb the PSF as much as in a sampled representation. Figure 4.6 shows this approach schematically.

4.5.2 Band-limiting deconvolution

The easiest way to circumvent the amplification of noise at high wave numbers is to band-limit the inverse OTF to some constant α : The effective inverse OTF is set to

$$\hat{q}(\mathbf{k}) := \begin{cases} \frac{1}{\hat{p}(\mathbf{k})} & , |\mathbf{k}| < \alpha \\ 1 & , \text{else} \end{cases} \quad (4.26)$$

assuming that wave numbers with a norm larger than α primarily contain noise. α must be tuned to the noise level of the image; this is a tradeoff between loosing details and noise in the deblurred image.

The lower the noise level of the observed image, the better this approach works. When a certain level is exceeded, a reasonable deconvolution is not possible: α must be set so small for a reasonable noise level that too much information is lost. If, in such a case, α is chosen large enough to retain the details, the image is much too noisy.

The constant in (4.26) for $|\mathbf{k}| \geq \alpha$ could also be set to zero; however, the effect is that too much details are lost and the image is more smoothened than sharpened.

As the jump in the inverse OTF introduces artificial discontinuities, another approach is to continue the inverse OTF with a polynomial \hat{r} to assert smooth values and first derivative. A third-order polynomial has four degrees of freedom, which are defined by the following conditions:

$$\hat{r}(\mathbf{k}_b) = 1/\hat{p}(\mathbf{k}_b) \quad (4.27)$$

$$\frac{\partial \hat{r}(\mathbf{k}_b)}{\partial \mathbf{k}} = \frac{\partial}{\partial \mathbf{k}} 1/\hat{p}(\mathbf{k}_b) \quad (4.28)$$

$$\hat{r}(\mathbf{k}_c) = 1 \quad (4.29)$$

$$\frac{\partial \hat{r}(\mathbf{k}_c)}{\partial \mathbf{k}} = 0 \quad (4.30)$$

where $k_b = \{k \mid |k| = \alpha\}$ is the boundary wave number and k_c with $|k_c| > |k_b|$ the wave number where \hat{q} begins to be constant. Note that k_b and k_c are sets of points. k_c can be set to $(1 + \beta)k_b$, where β is an additional free (scalar) parameter. This approach for defining k_c conserves the symmetries of the optical transfer function. The conditions (4.27)–(4.30) define a set of linear equations for each coordinate which can be solved for the unknown coefficients of the polynomial \hat{r} .

4.5.3 Richardson-Lucy deconvolution

The Richardson-Lucy (R-L) deconvolution algorithm is an iterative procedure proposed by Richardson (1972) and Lucy (1974) to obtain the most likely true image using the model (4.22). It is useful if the PSF is known, but nothing special about the noise. A review can be found in Hanisch et al. (1997). An effective implementation to accelerate the procedure was given by Biggs and Andrews (1997).

As mentioned above, g is assumed to be one sample of the random variable G . G is assumed to be Poisson distributed, according to the process of photon counting. Then, the probability to measure a value $g(x)$ at location x is

$$\mathcal{P}(G(x) = g(x)) = \frac{e^{-\gamma(x)} \gamma(x)^{g(x)}}{g(x)!}, \quad (4.31)$$

where $\gamma(x)$ is the expected value of G at location x .

For simplicity, in the following all $K := M \cdot N$ pixel at locations x_{ij} are labelled linearly with one index k , $k = 1, \dots, K$. The probability that all measured pixel of the image have the value g_k is the product of the probabilities of each single pixel (4.31),

$$\mathcal{P}(G = g) = \prod_{k=1}^K \frac{e^{-\gamma_k} \gamma_k^{g_k}}{g_k!}. \quad (4.32)$$

The true image, i.e. the expected value of f , is unknown. The idea now is to estimate a solution φ , using to the model (4.22) (i.e. $\gamma = \varphi * p$), such that the probability $\mathcal{P}(G = g)$ for obtaining the actually measured image g is maximised. In other words, the PDF of G is adjusted such that it is most likely to measure g . This is done by adjusting φ (and thereby the expected value γ of the Poisson distribution). Accordingly, we interpret \mathcal{P} as a function of $\gamma(\varphi)$ (recall that $\gamma_i = \sum_{k=1}^K \varphi_k p_{ik}$, where $p_{ik} = p_{i-k} = p(x_i - \xi_k)$ is the discrete PSF), g is a known parameter.

Maximising \mathcal{P} is equivalent to maximising $\ln \mathcal{P}$. The maximum is found by setting the derivative in respect to the unknown φ_i to zero:

$$\frac{\partial[\ln \mathcal{P}]}{\partial \varphi_i} = \sum_{k=1}^K \left[-\frac{\partial}{\partial \varphi_i} \gamma_k + g_k \frac{\partial}{\partial \varphi_i} \ln \gamma_k - 0 \right] = \sum_{k=1}^K p_{ki} \left[\frac{g_k}{\gamma_k} - 1 \right] = 0. \quad (4.33)$$

A solution of equation (4.33) is $\gamma = g$.

Still the question of estimating φ remains. To obtain an iterative procedure, f and g are regarded as PDFs. This can always be done if they are normalised. They would then express the probability to have a photon detected in a certain pixel. As explained above, p can be seen as conditional probability. Then, Bayes' law (Bronstein et al., 2001, sect. 16.2.1.3, p. 771) can be used to estimate the (inverse) probability $q(\xi|x)$ that $\Xi = \xi$ presumed that $X = x$:

$$q(\xi|x) = \frac{f(\xi)p(x|\xi)}{g(x)} = \frac{f(\xi)p(x|\xi)}{[f * p](x)} \quad (4.34)$$

q is the inverse PSF, $q(\xi) = p^{-1}(\xi)$:

$$f(\xi) = [g(x) * q(\xi|x)](\xi) \quad (4.35)$$

Inserting (4.34) into (4.35) yields

$$\begin{aligned} f(\xi) &= \left[g(x) * \frac{[f(\xi)p(x|\xi)]}{[f(\chi) * p(x|\chi)]} \right](\xi) \\ &= f(\xi) \left[g(x) * \frac{p(x|\xi)}{[f(\chi) * p(x|\chi)]} \right](\xi) \end{aligned} \quad (4.36)$$

Although this equation, when directly evaluated using $g(x) = (f * p)(x)$, results in $1 = 1$, it can be used to motivate an iterative procedure. f on the left side is interpreted as an updated approximation to φ , while f on the right as an old approximation. $g(x)$ is the measured data. The term $f * p$ in the denominator is thus an approximation to γ . Then we have

$$\varphi^{m+1}(\xi_i) = \varphi^m(\xi_i) \sum_{k=1}^K \frac{g(x_k)}{\gamma(x_k)} p(x_k|\xi_i), \quad (4.37)$$

where φ^m denotes the m th iteration of φ . Obviously, this iteration scheme converges when $g = \gamma$: In that case, the fraction becomes one and $\varphi^{m+1} = \varphi^m$. Equation (4.33) told us that $g = \gamma$ is also the solution with maximum likelihood. Therefore, if the iteration (4.37) converges, it converges to the most likely solution.

Note that the sum in equation (4.37) is not a convolution of the fraction g/γ with p , as we sum over k but the indices of p are ki , not ik .

Recall that all images are positive, as no negative numbers are possible when counting photons. Of course the PSF p is also positive (remember that it can be interpreted as a PDF). Thus, provided an initial positive guess φ^0 , equation (4.37) shows that φ^m will also be positive for all m .

In realistic cases, due to measurement noise, γ will not fit the data perfectly. Then the procedure will converge to the most likely solution. This can be realised as follows: One iteration step changes φ_i by

$$\Delta\varphi_i^m = \varphi_i^{m+1} - \varphi_i^m = \varphi_i^m \left[\sum_{k=1}^K \frac{g_k p_{ki}}{\gamma_k} - 1 \right] \quad (4.38)$$

For small $\Delta\varphi$, the change of the logarithm of the likelihood, $\Delta[\ln \mathcal{P}]$, can be approximated by a first-order Taylor expansion. Noting that $\sum_{k=1}^K p_{ki} = 1$, as p is a PDF, one obtains

$$\Delta[\ln \mathcal{P}]^m \approx \frac{\partial[\ln \mathcal{P}]}{\partial \varphi_i^m} \Delta\varphi_i^m = \varphi_i^m \left[\sum_{k=1}^K \frac{g_k p_{ki}}{\gamma_k} - 1 \right]^2 \geq 0. \quad (4.39)$$

As noted above, all φ^m are positive if φ^0 is positive. Therefore $\Delta[\ln \mathcal{P}]$ is also always positive, the likelihood increases with every iteration step.

Suppressing noise

A problem of R-L is that it may yield noise amplification. This is a generic problem of maximum likelihood algorithms, which try to fit the data as closely as possible. With too many iterations, the deconvolved image sometimes shows “speckles” (figure 4.7) which are not representative of a real structure but the result of fitting the noise too closely (Hanisch et al., 1997). In order to reproduce a small noise bump in the observed data, the reconstructed data must have a

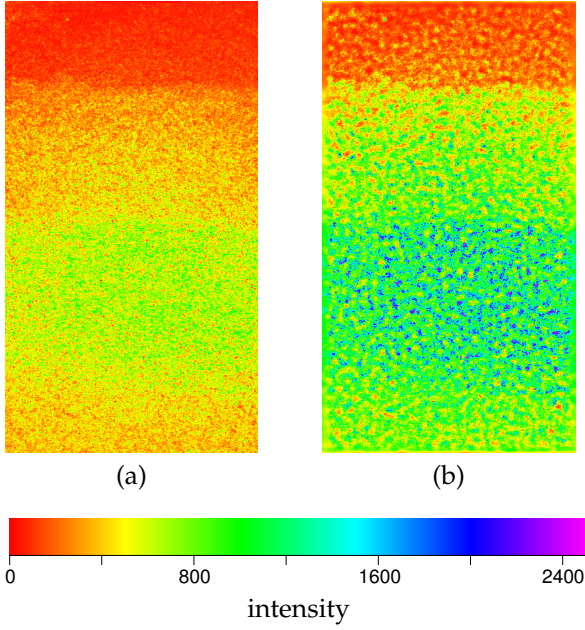


Figure 4.7: “Speckle” artefacts generated by noise amplification by the Richardson-Lucy deconvolution: (a) original image, (b) de-convolved image.

large noise peak. To conserve the optical flux, pixels near the bright spike must then be very dark. In the R-L method, only the non-negativity requirement limits noise amplification: If the compensating dark areas have decreased to zero, nearby spikes cannot grow any further.

A practical approach to limiting noise amplification is simply to stop the iteration in time. However, it is difficult to say when exactly to stop. There is no criterion for R-L how close the fit should be. Additionally, the required number of iterations may be different in different image areas. Smooth, extended areas may be fitted well after only a few iterations, while peaks or abrupt changes need much more iterations.

An effective approach for an adaptive stopping is to modify the maximum likelihood function equation (4.32) such that it becomes flatter in the vicinity of a good fit by introducing a damping function δ with parameter η , as described in Hanisch et al. (1997). Instead of $\ln \mathcal{P}$, $\delta(\ln \mathcal{P})$ is minimised, with

$$\delta(y) = \begin{cases} \frac{\eta-1}{\eta+1} (1 - y^{\eta+1} + y^\eta), & y < 1 \\ y, & y \geq 1 \end{cases}. \quad (4.40)$$

δ is chosen to be linearly proportional to y for $y > 0$, constant for $y \approx 0$ and has continuous first and second derivatives at $y = 1$. η determines how sudden δ becomes flat for $y < 1$. For $\eta = 0$ there is no flattening at all. The larger η , the flatter becomes δ .

To determine the threshold at which damping occurs, a new variable t is introduced into $\ln \mathcal{P}$ and the constants in are chosen such that in the modified $\ln \mathcal{P}'$, the expected value in the presence of Poisson noise is 1 if $t = 1$. t then determines the number of standard deviations σ of G when damping occurs. Using Stirling’s formula, $\ln n! \approx n \ln n - n$, one obtains

$$\ln \mathcal{P}'(G = g) \approx \sum_{k=1}^K -\frac{2}{t^2} \left[g_k \ln \frac{\gamma_k}{g_k} - \gamma_k + g_k \right] \quad (4.41)$$

4.5.4 Wiener filter

Wiener’s algorithm (Lückenhaus, 1997; Gonzalez and Woods, 1993, sect. 5.5) is a linear deconvolution filter derived from a least-squares relation for noise model I. The idea is to minimise

the mean error,

$$E([f(x) - \varphi(x)]^2) \rightarrow \min , \quad (4.42)$$

where E denotes the expected value (equation (A.4)), and φ the reconstructed image. The goal is to obtain the best linear estimate, i. e. find the inverse PSF q with $\varphi(x) = [q * g](x)$ which solves (4.42). The minimum with respect to q is calculated by setting the derivative to zero. Since E is linear, derivative and expectation can be interchanged and we obtain

$$E\left(\frac{\partial}{\partial q}[f - q * g]^2\right) = E(2[f - q * g][-g]) = 0 \quad (4.43)$$

$$E(fg) = q * E(gg) , \quad (4.44)$$

because q is not a random variable and thus can be taken out of the expected value. $E(fg)$ is the cross-correlation of f and g and $E(gg)$ the autocorrelation of g . Fourier-transforming (4.44) allows solving for \hat{q} yields

$$\hat{q} = \frac{\hat{S}_{fg}}{\hat{S}_{gg}} , \quad (4.45)$$

where $\hat{S}_{gg} = \hat{g}^* \hat{g} \bullet \bullet E(gg)$, the Fourier transform of the autocorrelation, is the spectral power density of g , $\hat{S}_{fg} = \hat{g}^* \hat{f} \bullet \bullet E(fg)$ and a^* denotes the conjugate complex of a . Using the acquisition model, (4.19), and considering that the noise n and the original image f are uncorrelated, equation (4.45) can be rewritten in terms of the OTF \hat{p} , the spectral power density of the noise S_{nn} and the spectral power density of the original image S_{ff} :

$$\hat{q}(k) = \frac{\hat{p}^*(k)S_{ff}(k)}{|\hat{p}(k)|^2 S_{ff}(k) + S_{nn}(k)} = \frac{\hat{p}^*(k)}{|\hat{p}(k)|^2 + \frac{S_{nn}(k)}{S_{ff}(k)}} \quad (4.46)$$

This is the OTF of the Wiener filter.

Without noise, the Wiener filter reduces to the simple deconvolution, $\hat{q} = 1/\hat{p}$. For large noise, the noise term S_{nn}/S_{ff} gets large and the frequency response goes to zero. This leads to a smoothing of noisy parts of the image. Thus, the Wiener filter is a combination of an inverse filter and a noise-dependent smoothing.

In practice, the spectral power densities of noise and original image are unknown. Thus, a reasonable estimate is necessary. If the properties of the noise are known, e. g. via a calibration measurement of the image acquisition sensor, S_{nn} can be reasonably estimated. Another possibility is extracting the noise from the measured image g using a noise elimination filter, e. g. a median filter, and obtaining a noise estimate using the difference between measured and denoised image. S_{ff} normally has to be estimated from the measured image. In the absence of further information, $S_{nn}(k)/S_{ff}(k)$ is often assumed to be constant for all frequencies. Then the required information reduces to the noise-to-signal power ratio.

The Wiener filter preserves the optical flux. However it sometimes generates spherical artefacts around peaks, or line-shaped artefacts alongside edges.

4.5.5 Ringing

The appearance of spurious high-frequency structures in the reconstructed image is called “ringing”. It can also occur in perfectly noiseless data because the high-frequency structures vanish in the blurred original data. These correspond to frequencies where the OTF has zeros. Ringing can only be eliminated by placing constraints on the object.

5 Light transmission through porous media

With modern cameras, optical measurements provide a high spacial and temporal resolution, which is very attractive and allows assessing phenomena which cannot be investigated with traditional sensors. At any rate, the measured irradiance must be a proxy for the quantities of interest, thus some processes must exist which depend on the quantity of interest and have an influence on the optical properties of the investigated object. This influence can then be exploited to optically measure the quantity of interest. In the context of this thesis, water in porous media is to be measured optically. In the following, the theoretical foundation of this measurement is presented.

5.1 Properties of light

Light, or more generally electromagnetic radiation, and its interactions with matter are described by quantum theory. However, in the limiting case of large quantum numbers, i. e. for macroscopic and also some atomic phenomena, the discrete nature of radiation can be neglected and it can also be described with classical electrodynamical methods (Jackson, 1999, I1). For example, in 1 m distance from a 100 W light bulb, there are about 10^{15} photons $\text{cm}^{-2} \text{s}$. In terms of the classical theory, radiation is a wave solution of Maxwell's equations. Assuming non-magnetic, non-conductive media, the wave equations

$$\nabla^2 \mathbf{E} - \epsilon \epsilon_0 \mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0 \quad (5.1)$$

$$\nabla^2 \mathbf{B} - \epsilon \epsilon_0 \mu_0 \frac{\partial^2 \mathbf{B}}{\partial t^2} = 0 \quad (5.2)$$

can be derived from Maxwell's equations, where \mathbf{E} is the electric field vector, \mathbf{B} the magnetic field vector, ϵ_0 the electric field constant, ϵ the relative electric permittivity which describes the medium, μ_0 the magnetic field constant, and $c = 1/\sqrt{\epsilon_0 \mu_0}$ the vacuum light velocity (Zinth and Zinth, 2009). The wave velocity, or phase velocity, in the medium is $u = c/\sqrt{\epsilon}$. The influence of the medium on wave propagation is thus described by the relative electric permittivity ϵ . $n := \sqrt{\epsilon}$ is called the refractive index of the medium. n is in general a complex number and depends on wavelength.

Solutions of the wave equation are plain waves,

$$\mathbf{E}_{\pm}(\mathbf{x}, t) = \mathbf{E}_0 \exp i(\mathbf{k} \cdot \mathbf{x} \pm \omega t) \quad (5.3)$$

and any linear combinations thereof, with the additional requirement

$$k^2 u^2 = k^2 \frac{c^2}{n^2} = \omega^2, \quad (5.4)$$

where \mathbf{k} is the wave vector, ω the angular frequency, and E_0 the amplitude of the electric field (Jackson, 1999, sect. 7). Equation (5.4) is called dispersion relation and relates the absolute value of the wave vector with the angular frequency. The wave length is defined as $\lambda = \frac{2\pi}{k} = \frac{2\pi u}{\omega}$ and describes the length between two wave minima or maxima. The velocity and

therefore the wavelength may change if the wave comes into another medium with a different permittivity. The frequency $\nu = \frac{\omega}{2\pi}$ describes the number of periods per time. Energy transport by electromagnetic waves, i. e. its energy flux density, is described by the Poynting vector

$$\mathbf{S} = \epsilon_0 c^2 (\mathbf{E} \times \mathbf{B}) . \quad (5.5)$$

The temporal average of \mathbf{S} during one oscillation period results in the radiation intensity I (Zinth and Zinth, 2009):

$$I = \overline{|\mathbf{S}|} = \epsilon_0 n c \overline{|\mathbf{E}|^2} . \quad (5.6)$$

5.1.1 Refraction and Reflection

At the interface between two materials with different permittivities ϵ_1 and ϵ_2 , light can be refracted, reflected, and/or diffracted. If the thickness of the bulk of the two materials 1 and 2 is large compared to the wavelength, $\ell \gg \lambda$, i. e. additional interfaces to other materials are far off such that the radiation does not “see” them when it is located at the boundary (but the wavelength is still not short enough that the radiation resolves the atomic scale), it is sufficient to consider the isolated interface of two semi-infinite media (figure 5.1). By considering the boundary conditions for the phase factors at the boundary, Snellius’ refraction law,

$$\frac{\sin \alpha}{\sin \alpha'} = \frac{|\mathbf{k}'|}{|\mathbf{k}|} = \frac{n_2}{n_1} , \quad (5.7)$$

and $\alpha'' = \alpha$ for reflection can be deduced (Jackson, 1999, sect. 7.3). By exploiting the continuity conditions at the interface, Fresnel’s formulas for the amplitude of the refracted and reflected waves can be derived (Jackson, 1999, sect. 7.3):

$$t_{\perp} := \frac{E'_{0\perp}}{E_{0\perp}} = \frac{2n_1 \cos \alpha}{n_1 \cos \alpha + \sqrt{n_2^2 - n_1^2 \sin^2 \alpha}} \quad (5.8)$$

$$r_{\perp} := \frac{E''_{0\perp}}{E_{0\perp}} = \frac{n_1 \cos \alpha - \sqrt{n_2^2 - n_1^2 \sin^2 \alpha}}{n_1 \cos \alpha + \sqrt{n_2^2 - n_1^2 \sin^2 \alpha}} \quad (5.9)$$

$$t_{\parallel} := \frac{E'_{0\parallel}}{E_{0\parallel}} = \frac{2n_1 n_2 \cos \alpha}{n_2^2 \cos \alpha + n_1 \sqrt{n_2^2 - n_1^2 \sin^2 \alpha}} \quad (5.10)$$

$$r_{\parallel} := \frac{E''_{0\parallel}}{E_{0\parallel}} = \frac{n_2^2 \cos \alpha - n_1 \sqrt{n_2^2 - n_1^2 \sin^2 \alpha}}{n_2^2 \cos \alpha + n_1 \sqrt{n_2^2 - n_1^2 \sin^2 \alpha}} , \quad (5.11)$$

where E_{\perp} and E_{\parallel} denote the electric field components tangential and parallel to the interface, respectively. The intensity reflection coefficient is then

$$R = \frac{I''}{I} = |\mathbf{r}|^2 , \quad (5.12)$$

where $\mathbf{r} = \mathbf{r}_{\perp} + \mathbf{r}_{\parallel}$ is the vector consisting of the two components r_{\perp} and r_{\parallel} . When calculating the intensity transmission coefficient it must be regarded that for radiation beams with finite extent the area of the light beam changes for $\alpha \neq 0$ since $\alpha' \neq \alpha$ (Zinth and Zinth, 2009):

$$T = \frac{I'}{I} = |\mathbf{t}|^2 \frac{\cos \alpha}{\cos \alpha'} = |\mathbf{t}|^2 \frac{n_2}{n_1} . \quad (5.13)$$

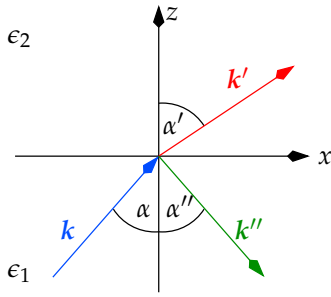


Figure 5.1: Geometry of refraction and reflection at the interface of two media with different permittivities ϵ_1 and ϵ_2 : A wave with wave vector k hits the interface and leads to a refracted wave with k' and a reflected wave with k'' .

5.1.2 Absorption

Inside one medium, the wave propagates linearly in direction of the wave vector. The imaginary part of n describes absorption of the medium and leads to an exponential decay of the intensity, which can be seen when inserting the dispersion relation with $n = n' + in''$ explicitly into the wave equation (5.3):

$$E = E_0 \exp \left(i \left[\frac{n' \omega}{c} \frac{k}{|k|} \cdot x - \omega t \right] \right) \exp \left(- \frac{n'' \omega}{c} \frac{k}{|k|} \cdot x \right). \quad (5.14)$$

By using equation (5.6) the intensity decrease due to absorption can be calculated. The resulting relation is called Lambert-Beers law and can be written as

$$I(L) = I_0(\lambda) e^{-\int_L \sigma(\lambda) \rho dl}, \quad (5.15)$$

where L denotes the optical path inside the medium, I_0 is the intensity at the start of the path, I the intensity after passing the path L , σ the cross section and ρ density. Note that cross section and density may change along the path which is not explicitly noted in equation (5.15), and also depend e.g. on pressure. The product $\mu = \rho\sigma$ is called absorption coefficient with unit inverse length. In a homogeneous medium equation (5.15) simplifies to $I = I_0 \exp(-\sigma\rho L)$. The absorption coefficient is related to the complex refractive index by $\mu = 2\omega n''/c$, as can be seen by comparing with equation (5.14).

Absorption is caused by the interaction of the radiation with the medium, which is described by quantum theory. Photons can be absorbed if their energy $h\nu$ matches to put the atom or molecule into another quantum-mechanical state. This causes the discrete nature of absorption spectra. If the energy of the photon is large enough to overcome the binding energy of an electron, that electron can be removed from the atom (photo effect). The molecule or atom is also needed for momentum conservation since it takes the recoil, absorption is not possible for free electrons. Thereby, a free electron with energy $h\nu - E_b$ is produced, where E_b is the binding energy. Here, a continuous absorption spectrum is encountered since the free electron may have an arbitrary amount of kinetic energy. The excited atom relaxes by filling the hole caused by the missing electron with one from a higher shell. During this process a characteristic photon is emitted, or a secondary (Auger) electron may be produced. The photo effect dominates for X-ray photons with $h\nu < 200$ keV. Its cross section depends on the energy of the photon, the cross section is $\sigma \propto (h\nu)^{-7/2}$ for non-relativistic rays (Heitler, 1984). This leads to a change of the spectrum towards higher photon energies by absorption, the so-called beam hardening. Photons with lower energy and thus higher cross section are absorbed first and the photons with higher energies remain.

For small molecules like water, N_2 , or O_2 , at room temperature, only rotational states are excited thermally. These are occupied according to the Boltzmann distribution. Vibrational and electronic transitions have much higher energies, thus these molecules are in the electronic

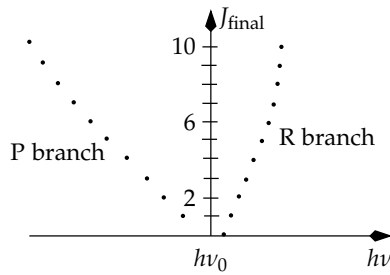


Figure 5.2: Fortrat diagram of a vibration-rotation transition. Points mark possible transitions with the corresponding energy $h\nu$ according to the rotation quantum number J of the final state. (According to [Demtröder \(2000\)](#))

and vibrational ground state. Typically vibrational transitions have energies corresponding to the infrared and electronic transitions energies corresponding to visible to ultra-violet.

Homo-nuclear diatomic molecules like N_2 and O_2 do not have a dipole moment. Since therefore no dipole interactions are possible, vibration-rotation transitions do not take place. Thus, these molecules cannot absorb thermal radiation. (Higher order moments and changing polarisability allow transitions with very low intensity.) For other molecules, since the photon has an angular momentum of $1\hbar$, due to angular momentum conservation, changes of the vibrational state must include a change of the rotational state of $\Delta J = \pm 1$, where J is the rotation quantum number. This leads to vibration-rotation absorption bands. Depending on the initial value of J in the absorbing molecule, different energies are found for the vibration-rotation transition (figure 5.2). In the approximation of the harmonic oscillator, the change of the vibration state must be $\Delta\nu = \pm 1$, where ν denotes the vibration quantum number. For the anharmonic oscillator, the selection rules also allow $\Delta\nu = \{\pm 2, \pm 3, \dots\}$, so-called overtones, with much lower intensity than $\Delta\nu = \pm 1$ ([Haken and Wolf, 2005](#)).

The excited state typically has a short lifetime. It may return by emitting a photon, or by inner conversion to another rotational state and subsequent relaxation. With relaxation, the energy is emitted non-radiative as heat by collisions with other molecules.

For higher photon energies like X-rays, Compton scattering becomes relevant. This is an inelastic scattering process where the energy loss depends on the incident angle between the photon before and after the scattering process. If the energy of the photon is larger than the rest energy of an electron-positron pair, $h\nu > 2m_e c^2 = 2 \cdot 511 \text{ keV}$, a pair can be produced if another particle is present which is needed for momentum conservation. Due to the high energy needed, this effect is only relevant for gamma rays.

5.1.3 Geometric optics

For more complex problems than simple refraction and reflection, e. g. imaging, the application of Maxwell's equations would be relatively complicated. However in the case $\ell \gg \lambda$ considered above where all dimensions of the problem are much larger than the wavelength and therefore all interfaces can be considered as interfaces of semi-infinite media, only the linear propagation of radiation inside one medium and the reflection and refraction laws are needed. This limit is called geometric optics. Formally, Maxwell's theory simplifies to geometric optics in the limit $\lambda \rightarrow 0$. However, the wavelength must not be so short that the radiation resolves the atomic scale, since then no macroscopic properties are encountered and the interactions are totally governed by quantum effects.

If the wavelength of the radiation becomes comparable to the dimensions of objects with which it interacts, $\lambda \approx \ell$, the assumption of two semi-infinite media and thus Fresnel's equations are not valid anymore. Here, Maxwell's equations must be solved explicitly for the special boundary conditions ([Zinth and Zinth, 2009](#)). For example, spherical particles with $d \approx \lambda$ exhibit Mie scattering (e. g. droplets in clouds). Interactions where $\ell \ll \lambda$ are again different (Rayleigh scattering).

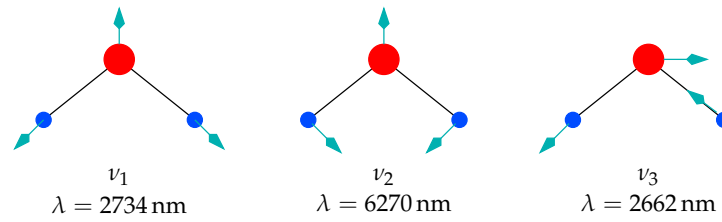


Figure 5.3: Normal oscillations of the water molecule. (According to [Haken and Wolf \(2005\)](#).)

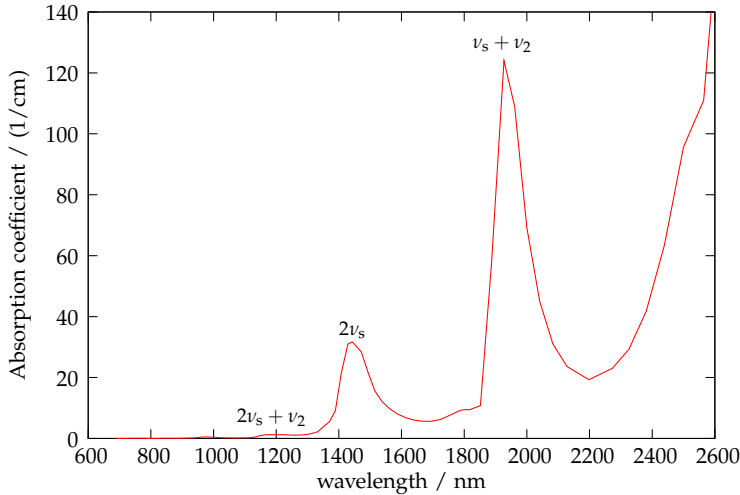


Figure 5.4: Absorption coefficient of water depending on wavelength as measured by [Palmer and Williams \(1974\)](#). Absorption bands were labelled according to [Chaplin \(2009\)](#).

5.2 Optical properties of water

The three-atomic water molecule (H_2O) is special compared to other natural oxides. It is liquid at ambient temperature and shows an anormal density behaviour with the highest density at 4°C . This is due to the molecule structure. The two hydrogen atoms have a binding angle of 104.5° and due to the higher electron density at the oxygen atom the molecule is polar. Therefore it facilitates H-bonds with neighbouring H_2O molecules with a binding energy between covalent and van der Waals bonds, about 5 times larger than thermal energy ([Chaplin, 2009](#)). Therefore, in the liquid phase, water molecules form $(\text{H}_2\text{O})_n$ clusters. These clusters fluctuate on a time scale of 10^{-12} s.

A system of N mass points has $3N$ degrees of freedom, thereof (for $N > 2$) 3 of translation, 3 of rotation and $3N - 6$ of inner motion. Therefore as a three-atomic molecule water has three possible normal oscillations which are sketched in figure 5.3. All possible vibrational states are linear combinations of these normal oscillations. All three exhibit periodic changes of the dipole moment. In liquid water, due to the molecule clusters and the corresponding interaction between molecules, absorption bands become broader. The main stretching bands ν_1 and ν_3 in liquid water are shifted to a lower frequency and they are close to each other ([Eisenberg and Kauzmann, 1969](#)). Therefore the stretching band is often considered as one single band (2865 nm) and denoted ν_s . The bending frequency ν_2 is increased (6079 nm) and the $2\nu_2$ band is very close to the ν_s band ([Eisenberg and Kauzmann, 1969](#)). Figure 5.4 shows the absorption of liquid water measured by [Palmer and Williams \(1974\)](#) with added labels of absorption bands according to [Chaplin \(2009\)](#).

Figure 5.5 shows the wavelength-dependent real refractive indices of water and ice. The real part of the refractive index is determined by the atomic composition of a material. The incoming wave induces dipole vibrations of the electron shell, which in turn couple with the

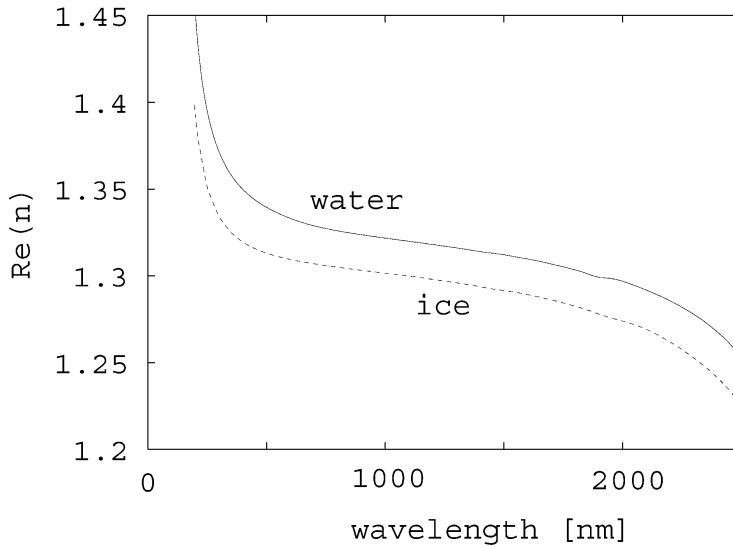


Figure 5.5: Real part of the refractive index n for water (solid line) and ice at 266 K (dashed line). Data were taken from Segelstein (1981) (water) and Warren (1984) (ice). Image source: Bänninger et al. (2005)

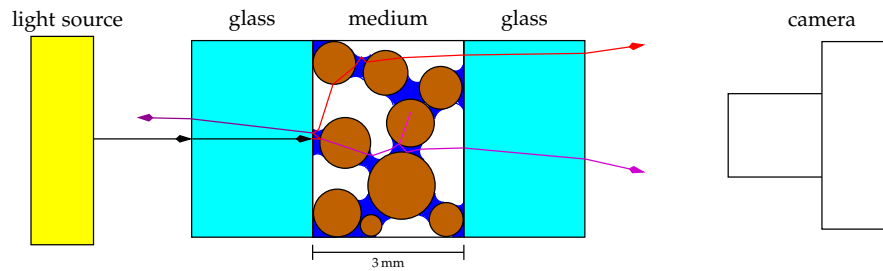


Figure 5.6: Propagation of radiation in a porous medium. Incoming radiation is reflected and refracted on the water-matrix, water-air, and air-matrix interfaces according to the actual geometry and the refractive indices n_i . In the bulk of the materials, absorption according to Lambert-Beers law takes place. Note that this is a cross-section of a 3D medium, which leads e. g. to apparently floating grains in the figure.

electromagnetic wave.

The optical properties of free water differs from that of water bound to a surface of another medium due to the different arrangement of water molecules (Bänninger, 2004, A.4.3), since clusters cannot form as inside the bulk of free water. Due to the changed hydrogen-bonding network, also the optical properties differ slightly.

5.3 Radiative transport through thin porous media

5.3.1 Visible and infrared light

Typical dimensions of soil grains are in the order of several hundred micrometers. The sands used in the Hele-Shaw-cell experiments carried out in this thesis had a grain sizes of more than $250 \mu\text{m}$. Even if pores can be of smaller size than the grain size by a factor of 2 or 3, for the wavelengths considered in this thesis, $\lambda < 1700 \text{ nm}$ and thus $\ell \gg \lambda$ is a reasonable assumption. Hence geometric optics can be used to describe the propagation of radiation.

Radiation entering the porous medium is reflected and refracted according to the actual geometry of the water-air, matrix-water and matrix-air interfaces and their respective refractive indices it encounters (figure 5.6). In the bulk it is absorbed according to Lambert-Beers law. Table 5.1 shows mean real refractive indices for the constituents of the soil for visible light. The

Table 5.1: Refractive indices of sand, water, and air for visible light (Lide, 1995; Hoa, 1981).

material	refractive index
sand	1.6
water	1.33
air	1.0

refractive index of water is much nearer to that of sand than the refractive index of air. According to Fresnel's formulas, equations (5.8)–(5.11), radiation is scattered in forward direction more often if the difference of refractive indices at the interface is smaller. Therefore, if more water is in the medium, there are more water-sand interfaces and less air-sand interfaces and more radiation will be transmitted and less reflected. This effect can be seen when walking on the beach, where the wet sand directly beside the shore is dark while the dry sand further away from the sea is bright. If there is no considerable absorption in water or air, this effect dominates radiative transport through the porous medium. However, if there are strong absorption bands at the wavelength considered (e.g. the water absorption at 1455 nm), the intensity increase at higher water contents can be masked by the absorption due to a longer path in water, and the intensity may actually decrease when the water content increases. In that case, wavelength dependent information can be used to separate the processes.

Let the radiation intensity in front of the sample be $I_0(x)$, where x denotes the (2D) spatial position on the surface of the medium. The light beam then propagates through the medium. Parts of the beam are eventually transmitted through the cell. Let the transmitted light intensity be $I(x)$. After passing the complicated path through the medium with refraction, reflection, and absorption, the intensity of one particular beam can be written as

$$I = I_0 \prod_{i=1}^{N_t} T_i \prod_{i=1}^{N_r} R_i \exp \left(- \int_L \rho(\xi) \sigma(\xi) d\xi \right), \quad (5.16)$$

where N_t is the number of transmissions and N_r the number of reflections along the light path, T_i and R_i are the corresponding transmission and reflection coefficient, respectively, for each incident, and L parametrises the light path. The T_i and R_i depend on the materials on each side of the interface and the incident angle, therefore on the actual matrix/water/air configuration in the medium and thus also on the water content. The optical properties T_i , R_i , and σ depend on the wavelength. To obtain the total transmitted intensity at one particular location, the intensity contributions by all the light paths to that point must be added:

$$I(x) = \sum_{j \in \text{paths}} I_{0,j} \prod_{i=1}^{N_{t,j}} T_{ij} \prod_{i=1}^{N_{r,j}} R_{ij} \exp \left(- \int_{L_j} \rho(\xi) \sigma(\xi) d\xi \right). \quad (5.17)$$

Consider a porous medium of fixed thickness d with the constituents matrix (sand), water, and air. It is assumed that each of the three phases is optically homogeneous, i.e. ρ and σ are constant within the phase (but still change with wavelength). In such a medium, equation (5.16) can be rewritten as

$$I = I_0 \prod_{i=1}^{N_t} T_i \prod_{i=1}^{N_r} R_i \exp \left(- \sum_{i \in \text{phases}} \rho_i \sigma_i d_i \right), \quad (5.18)$$

where d_i denotes the total length of the light beam in phase i . Absorption in the air phase is very small and can be neglected. The transmitted light intensity at x can then be written as

$$I(x) = I_0(x)f(\theta(x)) \exp(-\rho_w \sigma_w d_w(x)) , \quad (5.19)$$

where f is a function which implicitly contains the T_i and R_i along all light paths and the contribution of absorption by the matrix. It is a priori not known, because it depends on the complicated matrix/water/air configuration. Note that although the light paths start from a small region around x , in (5.19) the incoming intensity of all light paths has been assumed to be $I_0(x)$, because I_0 is typically varying slowly in space.

For wavelengths where water is transparent, $\sigma_w \approx 0$, θ can be estimated from I/I_0 . By using the standard two-point calibration (Jähne, 2005, sect. 10.3.3), the measured intensity can be normalised such that also fixed-pattern perturbations such as the grain structure of the matrix are removed. With an image of the completely saturated and completely dry medium, one obtains

$$\frac{I - I_{\text{dry}}}{I_{\text{wet}} - I_{\text{dry}}} = \frac{f(\theta) - f(\theta_r)}{f(\theta_s) - f(\theta_r)} =: F(\Theta) . \quad (5.20)$$

This normalisation has the advantage that not only inhomogeneous illumination is corrected (note that I_0 cancels in (5.20)), but also the fixed grain structure. Therefore one can concentrate on the property of interest. The relation $F(\Theta)$ is to be estimated separately.

Since the real part of the refractive index varies slowly with wavelength, it can be assumed constant for two adjacent wavelengths. By choosing these wavelengths such that for λ_1 the absorption of water is significant and for λ_2 it is approximately zero, division of equation (5.19) for these two wavelengths yields

$$\frac{I(x, \lambda_1)}{I(x, \lambda_2)} = \frac{I_0(x, \lambda_1)}{I_0(x, \lambda_2)} \exp(-\rho_w [\sigma_w(\lambda_1) - \sigma_w(\lambda_2)] d_w(x)) . \quad (5.21)$$

Accordingly, the mean path length in water d_w which is a measure for water content can be determined without knowledge of f if the difference of $\rho_w \sigma_w$ of both wavelengths and the relative spectral intensity $I_0(\lambda_2)/I_0(\lambda_1)$ are known. The latter can be estimated using a measurement of the completely dry sample.

Since the precise path of radiation through the porous medium and thus the change of light intensity strongly depends on the actual geometry, for a quantitative calculation of radiative transport in porous media the precise geometry of all phases must be known. Bänninger et al. (2005) presented a two-dimensional beam tracing model to calculate the reflectance and transmittance in a small soil sample. Two-dimensional cross sections of a pore network were set up, using artificial data or a sand cube measured with X-ray tomography. Water configuration was calculated with four different models, by adding water films of increasing thickness around the grains, by using a pore-network model, or a Lattice-Boltzmann simulation. The beams of a virtual diffuse light source were traced with a numerical model using a wavelength-dependent complex refractive index for water. Wet porous media contain bound water, which has a different molecular arrangement than bulk water, and thus also the refractive index is different. Therefore Bänninger et al. (2005) used the refractive index of ice for bound water. Using the model, reflection and transmission coefficients were calculated depending on the water content of the medium.

As expected, the simulations show an increasing transmittance with increasing water content (figure 5.7). The decrease in transmittance at very low water saturation is due to changing optical properties when the very thin water film is added to the grains after a state with no

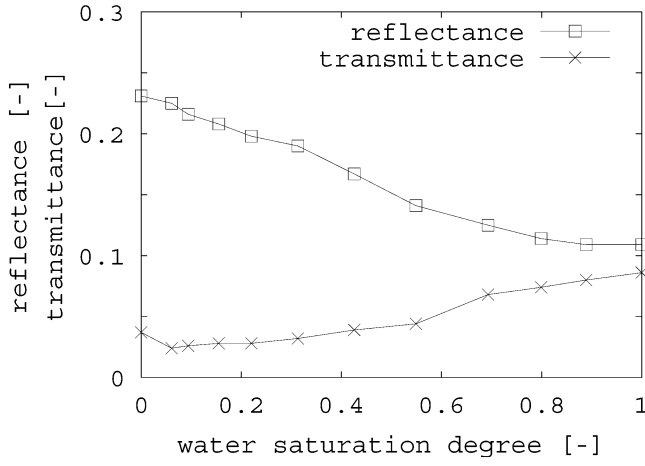


Figure 5.7: Dependence of reflectance and transmittance on water saturation computed by Bänninger et al. (2005). The water distribution was calculated with a pore network model.

water. Here, the light absorption is more significant than the scattering behaviour. Results also show that the spatial arrangement of the water is relevant for radiative transfer. Due to the changing light paths in the complicated geometry, changing the spatial location of the water while retaining the same water content may lead to different optical properties. Bänninger et al. (2005) found that especially for small samples, arrangement may be “more dominant than the amount of water”, which was caused by too small a cross-sectional image used for the 2D radiative transfer model, and they concluded that the cross section should contain a significant number of grains to represent the scattering medium. Therefore, it is also likely that if the sample thickness is too small, experimental water content determination may run into such problems.

5.3.2 X-rays

X-ray radiation has such a short wavelength that it resolves the atomic scale (X-rays with an energy of 100 keV have a wavelength of 0.12 Å). Therefore, the radiation encounters single atoms and does not recognise macroscopic material properties or interfaces. Thus, X-rays do not encounter multiple scattering in porous media. If the radiation passes different materials, density and atomic weight and therefore the attenuation changes. Processes leading to attenuation of photons are described in section 5.1.2. The attenuation is specified by Lambert-Beers law, equation (5.15). For a medium of constant thickness d consisting of sand, water, and air inside a sample cell, the attenuation reads

$$\begin{aligned} I &= I_0 \exp(-\{\mu_{\text{sand}}d_{\text{sand}} + \mu_{\text{water}}d_{\text{water}} + \mu_{\text{air}}d_{\text{air}} + \mu_{\text{cell}}d_{\text{cell}}\}) \\ &= I_0 \exp(-\{[(1 - \phi)\mu_{\text{sand}} + \theta\mu_{\text{water}} + (\phi - \theta)\mu_{\text{air}}]d + \mu_{\text{cell}}d_{\text{cell}}\}) , \end{aligned} \quad (5.22)$$

where $\mu_i = \sigma_i \rho_i$ are the attenuation coefficients of the different materials and d_i is the corresponding optical path length. With $A := -\ln I/I_0$ one obtains

$$\frac{A - A_d}{A_w - A_d} = \frac{\theta - \theta_r}{\theta_s - \theta_r} = \Theta , \quad (5.23)$$

where A_w and A_d refer to the fully saturated ($\theta = \theta_s$) and dry ($\theta = \theta_r$) medium, respectively (Bayer, 2005). Thus, the water saturation can directly be measured with X-ray attenuation.

6 Numerical simulation and inverse modelling

In many cases, the partial differential equations (PDEs) describing the dynamics of the system of interest cannot be solved analytically. This is also true for Richards' equation (2.19). Thus, the equation must be solved numerically to obtain an approximate solution.

Recall that for physically meaningful problems initial and boundary conditions in connection with the PDE represent a unique solution of the evolution of the system. Thus, using a numerical simulation, starting from an initial condition, one can simulate the evolution of a system, using its PDE and boundary conditions, to an arbitrary point in the future.

Consider the one-dimensional Richards equation:

$$C(\psi_m) \frac{\partial \psi_m}{\partial t} - \frac{\partial}{\partial z} K(\psi_m) \left(\frac{\partial \psi_m}{\partial z} - \rho_w g \right) = 0 \quad (6.1)$$

A solution $\psi_m(z, t)$ satisfies the PDE in the domain $\Omega = \{(z, t) \mid z \in [z_0, z_1], t \in [t_0, t_1]\}$, the initial condition $\psi_m(z, t_0) = \psi_{m0}(z)$ and the boundary conditions $\psi_m(z_0, t) = B_0(t)$, $\psi_m(z_1, t) = B_1(t)$.

The fundamental problem when numerically solving such problems is that PDEs are continuous functions and thus have an infinite amount of information, while computers can only handle a limited amount of information. The solution to this is discretisation, typically in space and time, but sometimes also in frequency. The unknown function is approximated on a grid consisting of a finite number of points or elements inside the domain Ω . This can be done in several ways. The discretisation must be sufficiently fine to capture all significant changes of the unknown function. The shape of the grid may be adopted to the special geometry of the problem. The time step is often adopted during the simulation depending on how fast the solution changes in a particular time period.

Only a short overview of possible discretisations is given here. Extensive information on the numerical treatment of partial differential equations can be found in textbooks as [Großmann and Roos \(2005\)](#).

6.1 Finite differences

The starting point for finite differences consists of approximating the unknown function on a finite number of points on a grid and replacing derivatives by differences on that grid (Taylor series expansion). Assume that space was discretised homogeneously with grid constant Δz in space and Δt in time. Quantities are only defined on the points $z_i = i\Delta z \in \mathbb{R}$ and $t_j = j\Delta t \in \mathbb{R}$, $i \in \{1, \dots, N\} \subset \mathbb{N}$, $j \in \{1, \dots, M\} \subset \mathbb{N}$. To approximate the derivatives of a function f , differences between adjacent grid points are used. Adding and subtracting Taylor series of $f(z + \Delta z)$ and $f(z - \Delta z)$ results in ([Bastian, 2008](#))

$$\frac{\partial f}{\partial z}(z) \approx \frac{f(z + \Delta z) - f(z - \Delta z)}{2\Delta z} \quad (6.2)$$

$$\frac{\partial^2 f}{\partial z^2}(z) \approx \frac{f(z + \Delta z) - 2f(z) + f(z - \Delta z)}{(\Delta z)^2}. \quad (6.3)$$

A first order (forward differences) approximation in time leads to

$$\frac{\partial f}{\partial t}(t) \approx \frac{h(t + \Delta t) - h(t)}{\Delta t}. \quad (6.4)$$

With $\psi_i^j := \psi_m(z_i, t_j)$, $C_i^j := C(\psi_i^j)$ and $K_i^j := K(\psi_i^j)$, the terms of equation (6.1) can be approximated as (Roth, 2005)

$$C(\psi_m(z_i, t_{j+1})) \frac{\partial \psi_m}{\partial t}(z_i, t_{j+1}) \approx C_i^{j+1} \frac{\psi_i^{j+1} - \psi_i^j}{\Delta t} \quad (6.5)$$

$$\begin{aligned} \frac{\partial}{\partial z} \left\{ K(\psi_m(z_i, t_j)) \left[\frac{\partial \psi_m}{\partial z}(z_i, t_j) - \rho_w g \right] \right\} \approx \\ \frac{K_{i+\frac{1}{2}}^j [\psi_{i+1}^j - \psi_i^j] - K_{i-\frac{1}{2}}^j [\psi_i^j - \psi_{i-1}^j]}{(\Delta z)^2} - \rho_w g \frac{K_{i+\frac{1}{2}}^j - K_{i-\frac{1}{2}}^j}{\Delta z}. \end{aligned} \quad (6.6)$$

Evaluating equation (6.6) at t_{j+1} and inserting it and (6.5) into (6.1) results in

$$C_i^{j+1} \frac{\psi_i^{j+1} - \psi_i^j}{\Delta t} - \frac{K_{i+\frac{1}{2}}^{j+1} [\psi_{i+1}^{j+1} - \psi_i^{j+1}] - K_{i-\frac{1}{2}}^{j+1} [\psi_i^{j+1} - \psi_{i-1}^{j+1}]}{(\Delta z)^2} + \rho_w g \frac{K_{i+\frac{1}{2}}^{j+1} - K_{i-\frac{1}{2}}^{j+1}}{\Delta z} = 0. \quad (6.7)$$

The time discretisation chosen above is named implicit Euler. Here, evaluation at t_{j+1} leads to the so-called implicit formulation. A system of equations must be solved for each time step, one for ψ_i^{j+1} at every node in space. For example, for Dirichlet boundary conditions, the number of equations is $N - 2$, since at the boundary nodes the boundary conditions come in. The advantage of this formulation is that it is more stable, however at the price of lower accuracy. The resulting equations are non-linear since C and K depend on ψ_m in a non-linear way. A set of non-linear equations is difficult to solve, but efficient methods to solve a set of linear equations are available. Therefore the equations are linearised. A common solution is to guess the ψ_i^{j+1} necessary to evaluate C and K based on ψ_i^j . With fixed C and K , the system is linear. Then, K and C are iteratively improved with the new ψ_i^{j+1} .

In equation (6.7), values of K between nodes have to be approximated (e.g. for $K_{i-\frac{1}{2}}$). How this is done, and where C and K are evaluated, is crucial for the performance of the solver. Details are discussed in textbooks, e.g. Ippisch (2008).

6.2 Finite volumes

As seen previously, the finite differences approach requires the unknown function to be differentiable twice and the function K once. With another formulation the differentiability requirements can be weakened. Also, for flux equations the mass balance is an important issue. The mass conservation can be implicitly put into the discretisation. The method also has some further advantages like higher flexibility in the discretisation of irregularly shaped domains.

The domain is partitioned into small sub-regions (control volumes, typically low-order simplices). The spacial discretisation is performed by requiring the partial differential equation to be satisfied in average on each control volume. By using integration by parts, portions of the volume integrals are converted into surface integrals which then couple neighbouring volumes. A finite volume scheme can also be interpreted as a special kind of finite element discretisation.

Typically, the differential equation is discretised with an Euler or Runge-Kutta type integrator in time, and for the resulting spacial equation finite volumes are used. For a mass conservative solution, Richards' equation is written in mixed form (Ippisch, 2008):

$$\frac{\partial \theta}{\partial t} + \underbrace{\frac{\partial}{\partial z} \left\{ -K(\psi_m) \left[\frac{\partial \psi_m}{\partial z} - \rho_w g \right] \right\}}_{=:q} = 0 \quad \text{in } \Omega \subset \mathbb{R}^1, \quad (6.8)$$

with boundary conditions

$$\psi_m = \psi_0 \quad \text{in } \Gamma_D \subseteq \partial\Omega \quad (\text{Dirichlet-type}) \quad (6.9)$$

$$q \cdot \nu = \phi \quad \text{in } \Gamma_N = \partial\Omega - \Gamma_D \quad (\text{flux or Neumann-type}), \quad (6.10)$$

where ν is the normal vector.

Let C_i be the control volume around grid point i , γ_{ij} the common surface between volumes i and j , ν_{ij} the normal vector on γ_{ij} , n_i the index set of neighbour volumes of control volume i and b_i the index set of boundary neighbours. The latter can be split into $b_{i,D}$ and $b_{i,N}$ for Dirichlet and Neumann boundaries, respectively. On the 1-dimensional space used here $\int_{C_i} 1 \, dz = \Delta z$ and $\int_{\partial C_i} 1 \, ds = 1$. Time is discretised as in the example of the previous section above with $t_k = k\Delta t$, $k \in \{1, \dots, N\} \subset \mathbb{N}$. Again abbreviations $\psi_i^k := \psi_m(z_i, t_k)$ and similar for θ and K are used. Integrating over any cell similar to Bastian (2008, sect. 7.1) leads to

$$\begin{aligned} 0 &= \int_{C_i} \frac{\partial \theta}{\partial t} \, dz + \int_{C_i} \frac{\partial q}{\partial z} \, dz = \frac{\partial}{\partial t} \int_{C_i} \theta + \int_{\partial C_i} q \cdot \nu_{ij} \, ds \\ &= \frac{\partial}{\partial t} \int_{C_i} \theta + \sum_{j \in n_i} \int_{\gamma_{ij}} q \cdot \nu \, ds + \sum_{j \in b_i} q \cdot \nu_{ij} \, ds \\ &\approx \frac{\partial}{\partial t} \theta(z_i) \Delta z + \sum_{j \in n_i} 1 q \cdot \nu_{ij} + \sum_{j \in b_i} 1 q \cdot \nu_{ij} \\ &\approx \frac{\theta_i^{k+1} - \theta_i^k}{\Delta t} \Delta z - \sum_{j \in n_i} \left\{ K_{ij} \left[\frac{\psi_j^{k+1} - \psi_i^{k+1}}{\Delta z} + \rho_w g \right] \right\} \\ &\quad - \sum_{j \in b_{i,D}} \left\{ K_j \left[\frac{\psi_j^{k+1} - \psi_i^{k+1}}{\Delta z/2} + \rho_w g \right] \right\} - \sum_{j \in b_{i,N}} \phi(z_j), \end{aligned} \quad (6.11)$$

where Gauss' theorem was used in the first line, the boundary integral was split into the different faces of the cell in the second line, the midpoint rule was applied in the third line. In the last line, the remaining differentiations were approximated by finite differences and the boundary faces were split into Dirichlet type and Neumann type boundaries. Equation (6.11) is a set of non-linear equations which must then be linearised and solved as described above in the finite differences section.

For faces between volumes i and j inside Ω , K may be evaluated either on the face corresponding to control volume i or on the face corresponding to volume j . These values may differ due to the discrete character of the solution. Therefore, some decision must be made about where to evaluate. Here, typically the volume where the velocity points outwards, the so-called upwind volume, is chosen. This method is called "upwinding".

6.3 Inversion

Simulation is a flexible tool for a forward problem, i. e. solving Richards' equation with given initial and boundary conditions. However, the hydraulic parameters are not known a priori,

thus a simulation offhand is not possible.

To obtain the parameters, an inverse modelling approach is used: initially, a parameter set is guessed and the forward problem is numerically solved. Then, the deviation of the modelled and the experimental data is analysed using an objective function. A natural choice of the objective function for uncorrelated normal distributed variables is the root mean square of deviations, that is

$$\chi = \sqrt{\frac{1}{N} \sum_{i=0}^N \left(\frac{x_i^{\text{model}} - x_i^{\text{exp}}}{\sigma_i^{\text{exp}}} \right)^2} \quad (6.12)$$

where the x_i denote the data points, and N is the number of points.

An iterative generic minimisation method is used to adjust the parameter set in order to minimise these deviations. Every iteration results in a new, corrected parameter set, which is then put into the model and the forward problem is again solved numerically, until a reasonable agreement between the measured and the modelled data is achieved, i. e. until χ is minimal. In practise, several conditions are often used to detect convergence, e. g. if χ has reached the value theoretically predicted as minimal value, or the decrease of χ has become smaller than a particular (small) value.

Part II

The soil-atmosphere boundary during evaporation

7 Introduction

The upper boundary of the vadose zone is the interface to the atmosphere. Interactions include precipitation, evaporation, transpiration, and the exchange of gases. The boundary conditions imposed by the atmosphere force the soil system and to a relatively large extent determine its state. On the other hand, the soil-atmosphere interaction is also important for atmospheric cycles and of significance for the climate. Thus, understanding the processes at the soil-atmosphere boundary is also of relevance for climate models, and the comprehension of the water and carbon cycle and of trace gases. In the following, evaporation is considered in more detail. It also plays a major role in practical applications like agriculture, in particular in semi-arid or arid regions with irrigation. Furthermore, it is not only important for the soil-atmosphere boundary, but also for many industrial and engineering applications like drying.

Many investigations are reported in the literature which assess the evaporation process. Evaporation involves coupled mass and heat transfer and depends in a complicated way on the atmospheric boundary conditions as well as on the properties of the porous medium. Evaporation from an initially saturated porous medium typically begins with a relatively high drying rate determined primarily by the external forcing. This phase continues as long as the medium can sustain the evaporative flow. Then it changes to a stage with falling drying rates. Researchers as early as [Sherwood \(1930\)](#) reported these two different regimes which he called “constant-rate period” and “falling-rate period” in experimental observations of paper drying. He divided the falling-rate period into two sub-periods, one directly following the constant-rate period where the decrease was attributed to a decrease in wetted surface area, and a second one where internal liquid diffusion controls the evaporation rate. In his review [Scherer \(1990\)](#) focused on drying in gels and ceramic. For the constant rate period, the evaporation rate was given with a diffusion-like law. They attribute the end of the constant rate period to the end of shrinkage of the sample, when the menisci are driven into the medium. The difference between the first and second falling rate period was explained by available film-like liquid water transport to the surface and totally diffusion controlled vapour flow.

Extensive work has been put into pore-scale modelling of the drying process; a review is given by [Prat \(2002\)](#). The main objective of REV-scale pore models is the determination of parameters like the permeability or the effective diffusion coefficient as a function of the saturation Θ in the limit of slow drying, while product-scale models are used to determine the liquid phase distribution and the evolution of drying rates ([Prat, 2002](#)). An explicit network of pores connected with bonds is solved numerically on a lattice. Inversion percolation theory is applied for modelling. The pore-scale analysis is valuable for the understanding of the detailed pore-scale processes. It shows that the drying front consists of a fractal region whose width is limited by capillary forces, since capillary pumping disrupts the inversion percolation pattern ([Tsimpanogiannis et al., 1999](#)). However these models typically cannot be directly applied for macroscopic problems, since the actual geometry of the medium is usually unknown.

[Shokri et al. \(2008\)](#) confirmed the fractal geometry of the drying front with drying experiments of sand in Hele-Shaw cells. They measured water content distribution with neutron absorption and the drying rate by mass loss recorded by digital balances. Their experiments focused on drying behaviour during the first stage. Within the unsaturated zone hydraulically connected pathways from the drying front to the evaporating surface were found. The front

morphology and dynamics were reported not to be affected by the evaporation rate.

On medium and large scales, several semi-empirical models exist for soil-atmosphere coupling. An overview is given in [Foken \(2003\)](#). These include energy balance models which normally use meteorological variables like temperature, wind speed and radiation along with some geophysical measurements to calculate cross-interface fluxes (e.g. Penman-Monteith), multi-layer models which parametrise energy balance for different layers separately, and resistance models. [Allen et al. \(1998\)](#) describe recommended field-scale evapo-transpiration models in detail.

The novel evaporation setup presented by [Schneider \(2005\)](#) allows laboratory measurements of bare soil with controlled atmospheric boundary conditions. A model for the soil-atmosphere boundary was also proposed in that thesis. It used a diffusive boundary layer approach coupled with a Richards' pore space model. It is reasonably simple but still provides a sufficient macroscopic description. For a deeper understanding of the processes described by that model numeric experiments are helpful, because detailed potential and water content profiles, sensitivity studies etc. are easily possible, which is not feasible in the laboratory. One aim of this study was to investigate the physical processes at the soil-atmosphere boundary as represented by that numerical evaporation model in more detail.

Evaporation experiments can also be used to determine hydraulic properties of soils. Movement of soil water is usually described by Richards' equation ([Jury et al., 1991](#)). A crucial part of this are the hydraulic material properties, in particular the soil water characteristic $\theta_\ell(\psi_m)$ and the hydraulic conductivity function $K(\theta_\ell)$. These properties are difficult to measure directly ([Topp and Miller, 1966](#)) which led to the development of inverse methods where a given parametrisation is adjusted such that modelled results are in optimal agreement with the corresponding measurement. Comprehensive reviews are provided e.g. by [Hopmans and Šimůnek \(1999\)](#) and [Vrugt et al. \(2008\)](#). Most popular today is the multi-step outflow (MSO) method ([Eching et al., 1994](#); [van Dam et al., 1994](#)) which evolved from one-step outflow ([Parker et al., 1985](#)). However the method is only applicable for the range of moderately negative potentials, hence to the rather wet range of soils.¹ Evaporation experiments allow virtually unlimited values of the matric potential (by making the air above the surface dry). More precisely, they are a natural complement to MSO: They are sensitive for strongly negative potentials but their accuracy deteriorates rapidly as potentials approach 0 kPa, the range where MSO works best. [Gardner and Miklich \(1962\)](#) were the first to propose such experiments. Their setup was basically retained in later studies ([Wendroth et al., 1993](#); [Tamari et al., 1993](#); [Romano and Santini, 1999](#); [Peters and Durner, 2008](#)). Here, a saturated soil sample is placed on a balance and exposed to free air while the matric potential in several depths is measured by tensiometers. However, a number of fundamental difficulties arise with this approach ([Schneider, 2005](#)). The limitations of this setup are overcome by the setup presented by [Schneider \(2005\)](#).

Virtual experiments allow the investigation of experimental design issues like the choice and placement of measurement devices and the choice of boundary conditions to optimally determine the hydraulic properties. The second aim of this study was to investigate and optimise

1 In outflow methods, gas pressure p_g in the soil sample equals ambient atmospheric pressure, as is the case in soils, while the pressure p_ℓ in the liquid phase at the lower end of the sample is reduced in one or more steps. Correspondingly, the matric potential $\psi_m = p_\ell - p_g$ in the sample is reduced and the resulting flow of water is recorded. Obviously, these methods are fundamentally limited to $\psi_m > -100$ kPa since p_ℓ must be larger than the vapour pressure of water. Practical limitations like the permeability of the phase separator at the lower boundary are more strict and typically lead to $\psi_m > -20$ kPa. A seemingly simple way to circumvent the fundamental limitation of multi-step outflow measurements would be to keep p_ℓ constant at ambient atmospheric pressure and to increase p_g . Since there is no fundamental limit to increasing p_g , ψ_m can be made arbitrarily negative. However, the water phase now is in a state completely different from that in a natural soil with the same value of ψ_m . Since the relation between the water contents of these two states is unknown, the parameters are not transferable.

the determination of hydraulic properties using the Schneider (2005) experiment. Specifically, (i) the sensitivity of the measurands to parameter changes to optimise the boundary conditions of the experiment was studied, (ii) it was explored if adding more observables into the inversion process yields significantly more information about the system, and (iii) the identifiability and uniqueness of the solution were analysed. The analysis also revealed valuable information about the physical processes. Multi-dimensional non-linear optimisation problems often have more than one minimum and the minima are often not well-localised, leading to ambiguous or contradictory solutions. Thus it is important to preclude such a behaviour with a detailed analysis.

The parameters of a sandy loam sample of a test site near Heidelberg, Germany were estimated by inverting the data of an evaporation experiment. This showed that the model is capable of describing the involved physical processes reasonably and that the new setup is suitable for parameter estimation.

8 Materials and Methods

8.1 Setup of the Evaporation Experiment

The soil sample is contained in a PVC cylinder. The bottom of the column is closed, the top of the soil column is closed by a gas-tight head space (evaporation chamber) (figure 8.1). A constant flow of air is established through the head space to remove the water vapour. The water vapour partial pressure p_w and temperature T of the incoming air are controlled and thereby the boundary condition at the upper boundary is set. The water flux is quantified by the difference of water vapour content before and after the evaporation chamber and the prescribed air flow through the head space.

The incoming air is conditioned with a cold trap, where the air is cooled to a defined temperature and abundant water is frozen out. The remaining water vapour corresponds to the saturation water vapour at the temperature of the cold trap. Afterwards the air is passively warmed back to ambient temperature in a copper calorifier. Water vapour concentration ν is measured by an infrared absorption gas analyser simultaneously before and after the head space. Inside the head space temperature and total pressure are measured. A fan in the head space mixes the air and ensures a homogeneous potential. This is required for the measurement used here to work, since otherwise a water vapour gradient would evolve in the head space and the water vapour content of the incoming air would not necessarily be the same as directly above the boundary layer. The total pressure is needed to calculate the water vapour partial pressure p_w from water vapour concentration ν . Air flow through the system is controlled with an adjustable vacuum pump and measured using the pressure drop in a capillary with known conductivity or using a flowmeter. Additionally, water content was measured with a time domain reflectometry (TDR) probe. It could be inserted either horizontally, or vertically to measure the total water content. Experimental details are described in [Schneider \(2005\)](#). In contrast to [Schneider \(2005\)](#), the bottom of the sample was totally sealed.

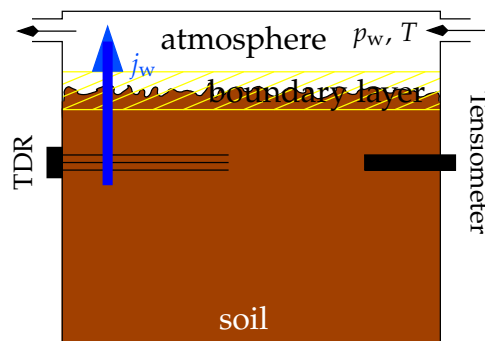


Figure 8.1: Sketch of the experimental setup. Evaporation takes place into a gas-tight head space above the soil surface. Air is flowing through it to take away the water. Water vapour molar fraction and temperature in the head space is controlled to define the boundary condition. The water flux is measured by vapour difference of incoming and outgoing air and the controlled gas flow.

8.2 Numerical Model

The numerical model $\mu\phi^1$ was used for simulations. Within a collaboration with its developer, it was extended to include the soil-atmosphere boundary and vapour transport. The soil column is modelled as a uniform one-dimensional medium with Richards' equation (2.33) including water vapour transport as effective conductivity (section 2.5). It is assumed that the soil water characteristic may be described by the van Genuchten or Brooks-Corey parametrisation, respectively, and the hydraulic conductivity function by the corresponding Mualem parametrisation (section 2.2.3).

A crucial step in the modelling is the representation of the upper boundary. At the transition between soil and air, many effects will affect the potential, resulting in a potential in the upper soil lower than the effective potential of the air. The air in the head space is turbolently mixed. The eddies cannot penetrate the boundary, thus eddy size will decrease when approaching the surface. In a thin layer around the surface, diffusive transport will dominate. Since turbulent mixing is much more efficient than diffusive transport, the vapour flow is controlled by a thin diffusive layer around the soil-atmosphere boundary. Because the soil is rigid, the boundary layer thickness is not likely to change. Therefore the boundary is modelled as a diffusive layer of constant thickness r_b assuming that the time scale of diffusion across this layer is much smaller than the time scale on which the boundary condition changes. This appears reasonable since the time scale of diffusion, given by $r_b^2/[2D_w]$, where D_w is the diffusion coefficient of vapour in air, is some 0.1 s for a layer thickness of $r_b = 2$ mm. The vapour flux across such a layer is given by the diffusion law as

$$\begin{aligned} j_{w,bnd} &= -D_w V_{w,m} \frac{\partial v_w}{\partial z} \\ &\approx -D_w V_{w,m} \frac{v_{w,air} - v_{w,soil}}{r_b} \\ &= -\frac{D_w V_{w,m}}{RT} \frac{p_{w,exp} - p_{w,s}(T) \exp\left(\frac{\psi_m V_{w,m}}{RT}\right)}{r_b} \end{aligned} \quad (8.1)$$

where v_w is the molar density of water vapour, $V_{w,m}$ the molar volume of liquid water, $p_{w,exp}$ the partial pressure of water vapour and T the temperature in the well-mixed head space above the soil column, and r_b the effective thickness of the boundary layer. In the last step, molar density was replaced by partial pressure using the ideal gas law, and equation (2.26) on page 13 was used to describe the partial pressure inside the soil using the matric potential. By definition, the processes in this layer are not resolved well. The real soil surface is rough. The physical location of the boundary layer is not well defined, i. e. the fraction of the layer that is within the soil column, and the porosity of the respective parts of the soil. Also the transition from liquid to vapour occurs in the boundary layer. However, in the thin layer considered here, all these effects are linear and only change the effective thickness of the boundary layer r_b . Therefore r_b becomes an effective fitting parameter without direct physical interpretation.

Hydraulic parameters were estimated from the measurements using inverse modelling. The numerical forward model together with the Levenberg-Marquardt algorithm was used, where the residuum is calculated by the squared sum of normalised deviations:

$$\chi^2 = \sum_i \left[\frac{y_i^{\text{model}} - y_i^{\text{measured}}}{\sigma_i} \right]^2. \quad (8.2)$$

¹ Developed by Olaf Ippisch, Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University

Table 8.1: Parameters used for the synthetic data sets: The van Genuchten parameters α and n , the saturated hydraulic conductivity K_s , residual water content θ_r , the saturated water content θ_s , and the resistance of the boundary layer r_b .

parameter	sand	sandy loam	silt
α / m^{-1}	5	10	0.5
n	42	2	2
$K_s / \text{cm h}^{-1}$	2	0.1	0.1
$\theta_s / \text{m}^3 \text{m}^{-3}$	0.3	0.3	0.3
$\theta_r / \text{m}^3 \text{m}^{-3}$	0	0	0
r_b / mm	3	3	3

Table 8.2: The different boundary condition scenarios used in the simulations. All simulations were simulated isothermally at $T = 293 \text{ K}$ and carried out until $t = 550 \text{ h}$.

scenario	p_w / kPa
“onestep”	1
“twostep”	$\begin{cases} 0.25 & , t \leq 28.8 \text{ h} \\ 2 & , t > 28.8 \text{ h} \end{cases}$
“threestep”	$\begin{cases} 0.25 & , t \leq 62.5 \text{ h} \\ 2 & , t > 62.5 \text{ h} \wedge t \leq 196.3 \text{ h} \\ 0.25 & , t > 196.3 \text{ h} \end{cases}$

The $\mu\phi$ forward model integrates Richards’ equation using a cell-centred finite-volume scheme with full-up-winding in space and an implicit Euler scheme in time. Linearisation of the non-linear equations is done by an inexact Newton method with line search. The linear equations are solved with a direct solver. For the time solver the time step is adapted automatically. A no-flux condition was used for the lower boundary. At the upper boundary the evaporation was calculated by equation (8.1).

Energy loss due to the latent heat of evaporation and the heat transfer in the soil sample was not simulated, assuming that the heat exchange between the sample and its environment is fast enough to compensate for the heat loss by evaporation. As shown in section 9.2 this assumption is violated in the initial phase of an experiment with a sandy loam. While no principle differences in the system behaviour are expected, the consequences of this simplification still have to be studied in the future.

The sensitivities required by the Levenberg-Marquardt algorithm were derived by external numerical differentiation.

8.3 Numerical Experiments and Analysis

Virtual experiments (by running the numerical model) were conducted where the true parameters are known and therefore the performance of the inversion process as well as the parameter space can easily be analysed. In this study, three parameter sets were used: a sand, a silt and a sandy loam. The corresponding parameters are given in table 8.1. All experiments were simulated isothermally with $T = 293 \text{ K}$. Boundary conditions which were finally used are given in table 8.2.

To test if additional measurements can improve the quality of the parameter estimation besides the evaporation flux j_w two additional (virtual) measurements are considered: (i) the

Table 8.3: Uncertainties assumed for the virtual measurement devices.

device	measurand	uncertainty
flux upper boundary	j_w	5 % of value
tensiometer	ψ_m	0.1 kPa
permittivity	ϵ_c	2 % of value

matric potential ψ_m measured by a tensiometer, and (ii) the water content θ measured by dielectric (compound) permittivity ϵ_c . Both probes are assumed to be installed 2 cm below the surface. The influence of the installation depth is analysed below. The measurement uncertainty assumed for each of the virtual devices is given in table 8.3. Note that, in contrast to real measurements, the virtual instruments provided point measurements. For the permittivity measurement, θ_s was used as porosity and ϵ_{soil} was assumed to be known.

If the potential falls below the air-entry value of a tensiometer or below the vapour pressure of water, whichever is higher, the tensiometer releases water to the sample. To prevent this disturbance, the tensiometer is removed at -30 kPa. To analyse the impact of this removal also simulations where the tensiometer was removed at -70 kPa, and simulations with a (hypothetical) unlimited measurement range were conducted.

Measurements were made after logarithmically growing time intervals $\Delta t_i = 300 \text{ s} + 2000 \text{ s} \times \log(i)$, starting again with $i = 1$ after each change of the boundary condition.

As large gradients are encountered in the simulated soil, especially at the drying front, a fine grid and small time steps are needed to avoid numeric noise. On the other hand, to keep the runtime reasonable, the spatial grid should be as coarse as possible. A grid convergence study was conducted showing that a reasonable grid convergence was obtained with a 1000 point non-regular grid with exponentially decreasing cell heights towards the soil surface. The uppermost cell had a height of 1×10^{-9} m. Of course, this exceedingly small size is not related to the real physics at that scale. However, one has to bear in mind that the necessary grid resolution can also depend on the hydraulic parameters used in the simulation. The time step was adopted automatically by the model.

Forward simulations of the evaporation experiment were used to study the physical dynamics of the system. To obtain the maximum amount of information about the unknown parameters to be optimised, a sensitivity analysis was performed for the experiment. Relative sensitivity coefficients were calculated according to

$$s_i(t, z, p_j) := \frac{\frac{\partial m_i}{\partial p_j}(t, z, p_j)}{\frac{m_i(t, z, p_j)}{p_j}} \approx \frac{\frac{m_i(t, z, p_j + \Delta p_j) - m_i(t, z, p_j)}{\Delta p_j}}{\frac{m_i(t, z, p_j)}{p_j}}, \quad (8.3)$$

where m_i denotes measurand quantity i (e. g. the water flux at the upper boundary j_w) and p_j is the j th parameter. s_i is a dimensionless quantity normalised by the measurand and parameter value which allows to compare the sensitivities of different measurands and different parameters, and is (except for numeric noise caused by the numeric differentiation) independent of the step size Δp_j . The results of the sensitivity analysis were used to optimise the boundary conditions of the experiment.

To check whether the data measured in the experiment is sufficient to identify a unique set of soil hydraulic parameters, response surfaces were calculated for all scenarios (similar to Toorman et al. (1992) for the onestep outflow experiment and Šimůnek et al. (1998) for traditional evaporation experiments). Two parameters were varied independently while all other parameters were kept at their true values. The χ^2 surface, defined by equation (8.2), is then displayed in contour plots. While this only shows a subset of the true five-dimensional

parameter space along parameter planes and new features might occur in the intermediate space, it is nevertheless a good indicator whether one unique and identifiable minimum exists. In the simulations, parameter i was multiplied by a factor γ_i , with γ_i varying from 0.1 to 2 in steps of 0.1 and from 2.2 to 3.8 in steps of 0.4, respectively.

Response surfaces were also used to investigate if adding more observables yields substantially more information for the inversion process. The χ^2 sum visualises the power of the experiment to identify the parameters. If the minimum is more localised, the parameters are identified easier and noise on the experimental data is less severe.

Finally, it was checked how well the inverse model converges to the real parameters given the measurement error and the cross correlation between the model parameters. The forward model was used to generate synthetic data for the best combination of observations as determined from the response surfaces. Random noise normally distributed with a standard deviation of σ_i was added to the data, where σ_i is the measurement uncertainty in the evaporation experiment (table 8.3) as determined by Schneider (2005). Five different data sets were generated for each scenario to account for the random influence of measurement noise. From each of this data sets, the parameters α , n , K_s and θ_s of the van Genuchten/Mualem model and the resistance of the boundary layer r_b were estimated with a variety of initial conditions. 10 parameter sets were randomly created in the range reasonable for the soil under examination. A logarithmic random distribution between the upper and the lower limit was used. These sets were used as start parameters for the gradient based inversion process resulting in 50 sets of estimated parameters for each soil. This approach combines a Monte-Carlo method with the Levenberg-Marquardt minimisation (Monte-Carlo Levenberg-Marquardt, MCLM). The inverse solutions were then compared with the real parameters and the resulting hydraulic functions with the true functions. A deviation coefficient was defined according to

$$d := \frac{p_{\text{inv}} - p_{\text{true}}}{p_{\text{true}}} \quad (8.4)$$

where p_{inv} denotes the inverted and p_{true} the true parameter.

9 Results

9.1 Numerical study of the evaporation process and parameter estimation analysis of the evaporation experiment

As the results of the numerical experiments are quite similar for the three different soil types under examination, only the results for the silt are discussed in detail. The results of the inverse modelling are given for all three soil types.

9.1.1 Onestep experiment

Physics of the process

The most simple scenario for an evaporation experiment which is also used in classical evaporation experiments is a onestep experiment as shown in figure 9.1. After saturation the sample is exposed to a constant vapour pressure at the upper boundary resulting in a progressive drying of the sample.

In this scenario two different regimes are distinguishable: Regime I, where the outflow is limited by the resistance of the boundary layer r_b , and regime II, where it is limited by the soil hydraulic properties. These two regimes are in accordance with experimental findings in the literature. Regime I leads to a constant value of j_w which only depends on r_b . Therefore, in this regime hydraulic properties cannot be determined with only the outflux as measurand. A sketch of the potential profile near the surface for three times is shown in figure 9.2. The major part of the potential drop is caused by the resistance of the boundary layer. Due to the much higher conductivity in the soil, water is delivered to the evaporating surface with a minimal gradient, which can also be seen in the simulation results (figure 9.3.(a)). Hence, the hydraulic properties have only a minimal influence on the flux. However, the potential of the sample changes with time due to the successive drying of the soil. Measurements of potential could therefore give information about the hydraulic properties of the sample. For $t \lesssim 20$ h, the potential in the soil is above -10 kPa and therefore can be measured easily with a tensiometer. However, due to the small deviations from the linear decrease (which would be expected for a

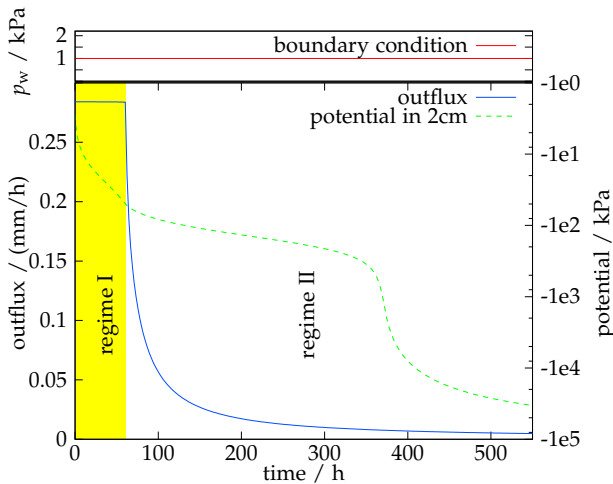


Figure 9.1: Water flux at the upper boundary j_w and potential in 2 cm depth of the one-step measurement.

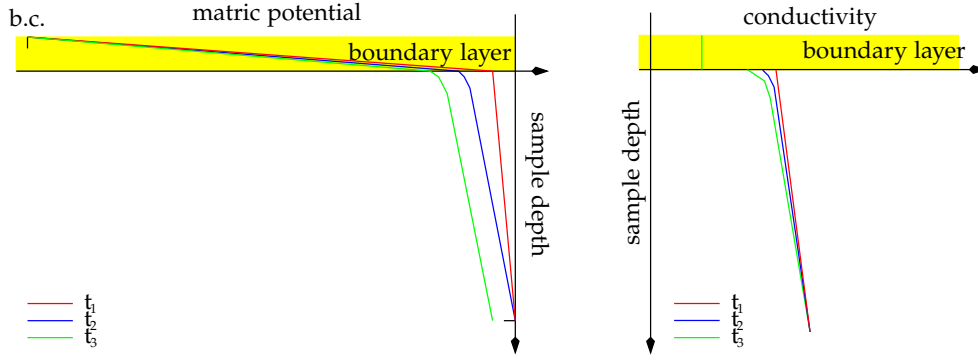


Figure 9.2: Sketch of the potential (left) and conductivity (right) development in regime I.

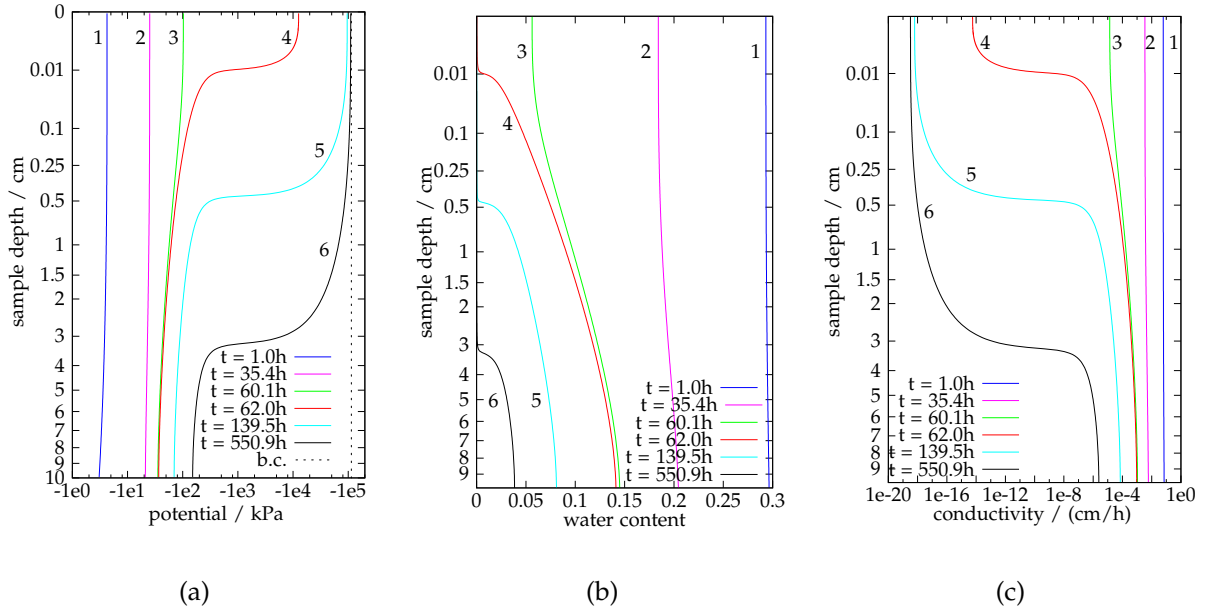


Figure 9.3: Potential $\psi_m(z)$ (a), water content $\theta(z)$ (b) and hydraulic conductivity $K(\theta(z))$ (c) distributions for the onestep experiment at different times t_i : (1) at an early time in regime I, (3) directly before the transition to regime II, (4) a short time later after the transition, (6) at the end of the experiment. Notice the non-linear scaling of the depth axis.

hydraulic conductivity which is constant over the whole sample), one would need a very high accuracy to obtain information about the hydraulic conductivity.

With continuing evaporation the potential and the water content decrease (figure 9.3), most rapidly near the surface. Eventually the conductivity of the soil becomes limiting. The system enters regime II and j_w starts to decrease rapidly. This transition is quite abrupt because (i) the function $K(\theta)$ is very steep in the relevant range and (ii) the effective hydraulic conductivity is dominated by the dry low-conductive layers. Therefore the hydraulic properties seen in this regime are the properties of the dry region. During this transition a drying front forms at the surface and then moves into the soil (figure 9.3).

Regime I occurs only if the saturated hydraulic resistance is lower than the resistance of the boundary layer. This is illustrated by reducing K_s by a factor of 20 (figure 9.4, blue and dashed cyan curves). K_s is now lower than the flux which can be evaporated through the

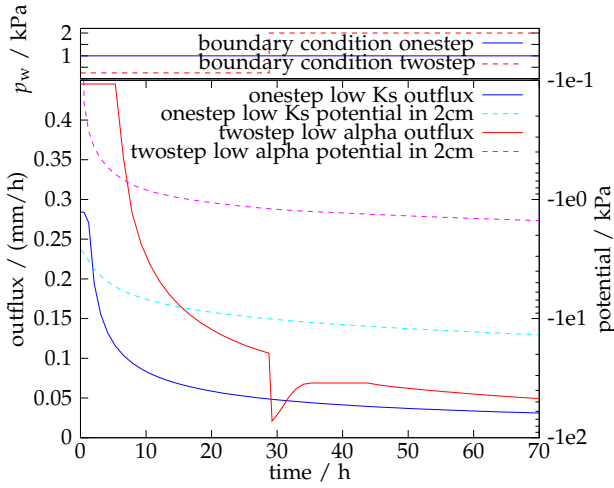


Figure 9.4: Water flux at the upper boundary j_w and potential in 2 cm depth for the “onestep” experiment with 20 times lower K_s , as well as for the “twostep” experiment with 20 times higher α . The former stays virtually no time in regime I due to the very low conductivity. The latter shows a more distinct switch-back to regime I but structural it is identical to the original soil.

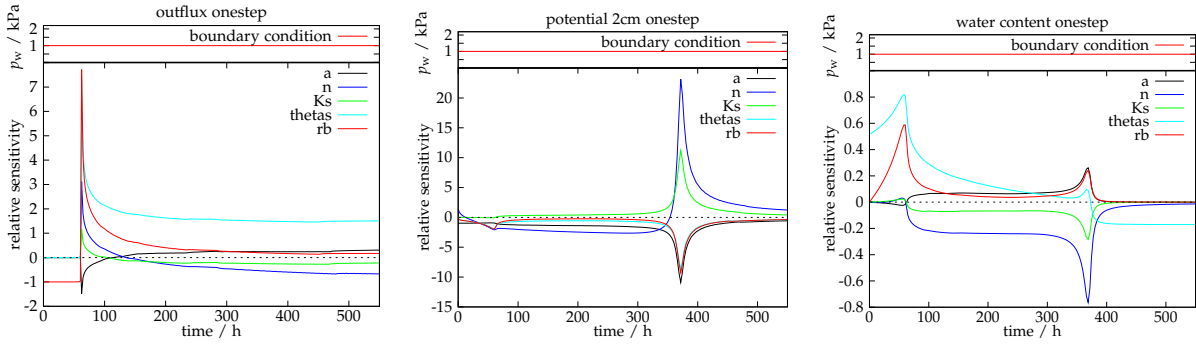


Figure 9.5: Sensitivity coefficients $s(t)$ for the onestep experiment for the outflux (left), the potential in 2 cm (middle) and the water content in 2 cm (right).

soil-atmosphere boundary layer. Thus, the sample directly enters regime II.

Sensitivity analysis

The relative sensitivity coefficients according to equation (8.3) were calculated for each measurement type for all hydraulic parameters (figure 9.5).

As the system is in regime I at the start of the experiment, the water flux is limited by the resistance of the soil-atmosphere boundary layer alone. At early times, the outflux j_w is therefore most sensitive to r_b while the sensitivity to all other parameters is very small.

When the topmost layer of the soil has dried out, the soil hydraulic properties, in particular the hydraulic conductivity, become limiting for the evaporation rate. The evaporation flux is most sensitive to all parameters exactly at the bend point where the system enters regime II and the outflow starts to decrease after the plateau. The sensitivity on r_b decreases continuously because the drier the sample, the less important is the resistance of the boundary layer. As θ_s scales the amount of available water when θ_r is held constant, the sensitivity of θ_s stays more or less constant after a quick decay. For all other parameters the sensitivity decreases after the maximum and after a zero-crossing eventually increases again with opposite sign. The zero-crossing can be explained by mass conservation. As the total water content of the sample is constant, a higher evaporation at earlier times has to be compensated by lower evaporation toward the end of the experiment and vice versa.

In contrast to the evaporation flux, the potential ψ_m is at the beginning of the experiment most sensitive to α and n which control the shape of the soil water capacity curve. K_s is the

only parameter for which the sensitivity is nearly zero during regime I. The sensitivity on r_b and θ_s are less important at the beginning, but increase during regime I and reach a maximum at the transition to regime II as well as the sensitivity on n .

r_b determines the speed of drainage in regime I, it is clear that the potential, which is connected to the water content by the water characteristic, is also dependent on r_b . This effect will become more pronounced as time passes because the longer a different outflux caused by a different r_b is retained, the higher are also the differences in the potential. θ_s determines the amount of available water. With a constant evaporation rate, the more water is available, the less is the relative change of water content and therefore the change of potential when all other parameters are kept constant. Thus, the behaviour of θ_s is analogous to the one of r_b .

The maxima are much less pronounced than for the evaporation flux. For all parameters the sensitivity is more or less constant or increases slowly in regime II and reaches a large peak at $t \approx 370$ h which is caused by the passing of the drying front at the tensiometer position. This peak is discussed in more detail in section 9.1.2. A zero-crossing only exists for n as only n influences the shape of the soil water capacity curve. The generally higher sensitivity in the dry range is in accordance with the result of Šimůnek et al. (1998) for traditional evaporation experiments. As they pointed out, the water characteristic becomes steeper for more negative potentials and thus parameter changes have more influence at lower potentials.

The water content is less sensitive to parameter changes than the evaporation flux and the potential. During regime I the only sensitive parameters are θ_s and r_b , which both increase with time. The maximum is again at the transition to regime II. For all other parameters there is no pronounced maximum at the transition point, but all sensitivity curves show a sensitivity maximum at the passing of the drying front. The water content is most sensitive to n and K_s during the early stage of regime II and to the available water θ_s towards the end of the experiment.

The sensitivity to changes of the saturated hydraulic conductivity is rather low for all types of measurements and reaches significant values only for the evaporation at the transition point and at the passing of the drying front.

9.1.2 Multistep experiments

Physics of the process

As the transition from regime I to regime II contains much information, one would suggest that multistep experiments can drastically improve the sensitivity if it is possible to reproduce the switch from regime I to regime II with boundary condition steps. To switch from regime II back to I, either the conductivity in the upper soil or the resistance of the boundary layer must be increased, or the potential drop on the boundary layer decreased. As the resistance of the boundary layer r_b is constant, this cannot be achieved by lowering the boundary potential. Lowering the boundary potential speeds up the drainage of the sample but does not lead to new features.

If the sample is already in regime II and the water vapour pressure at the surface is increased, a second plateau and a second drop of the flux can be seen if the pressure jump and the time between the steps are chosen adequately. This is illustrated with a twostep experiment, figure 9.4 (red and dashed magenta curves). To make the second step more pronounced, a 20 times higher α was used for this simulation. When the vapour pressure at the boundary is increased, the potential drop over the boundary layer and thus the water flux decreases, the boundary layer becomes limiting again. The flow inside the soil is now higher than the evaporation flux. This leads to an increase of the water content and thus the potential at the soil surface, resulting in a larger potential drop on the boundary layer and therefore an again

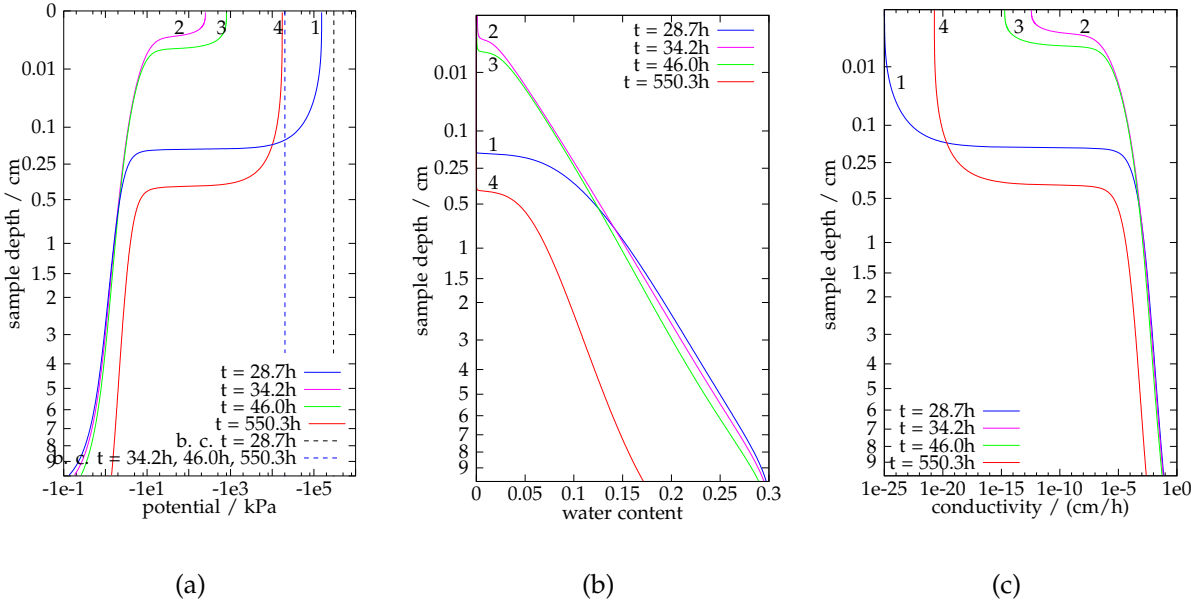


Figure 9.6: Potential $\psi_m(z)$ (a), water content $\theta(z)$ (b), and hydraulic conductivity $K(\theta(z))$ (c) distributions for the twostep experiment with 20 times higher α : (1) directly before the first boundary condition switch, (2) after relaxation, (3) after re-entering regime II and (4) at the end of the experiment. Notice the non-linear scaling of the depth axis.

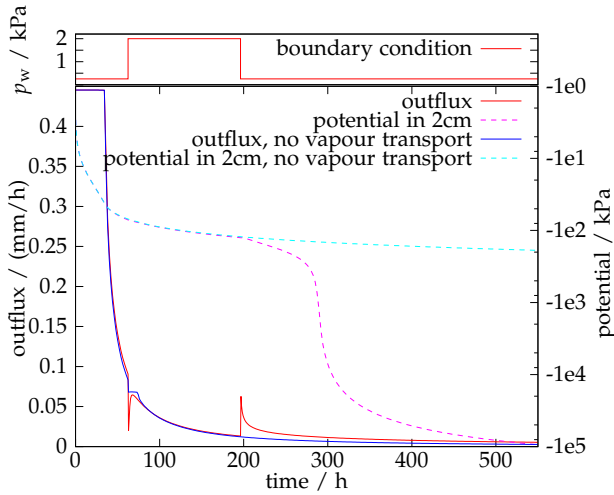


Figure 9.7: Influence of water vapour transport on the results: The threestep experiment was simulated taking vapour transport inside the soil into account (the normal case, red) and without considering it (blue).

higher evaporation flux (see figure 9.6). When this adaptation stage is finished, the increase of water content in the upper soil and therefore the increase of the evaporation rate ends, the delivery from below and the evaporation flux are equal again. If the change in the boundary condition was large enough, the resulting evaporation flux at this point is low enough to be sustained by the soil for a longer time span and regime I is reached again, else the system stays in regime II. This depends on the relation between the new potential drop on the boundary layer and the hydraulic conductivity in the soil (soil water state).

The same effect also occurs with the normal value of α , but it is harder to see (figure 9.7 red line). The higher α results in a less negative potential in the soil before the switch and a relaxation to a higher water content after the switch. Therefore it takes longer until the conductivity drops low enough to reach regime II again.

Acknowledge that any change in the direction of flow leads to hysteresis, which was not

considered in the simulation. Further work is needed to investigate if the influence of hysteresis can be seen in simulations and experimentally and how severe it changes the system.

The twostep experiment has the disadvantage that the potential range covered is too small during a reasonable measuring time. This can be compensated by applying a third step after the experiment has entered the hydraulically dominated regime again to speed up drainage. This results in a threestep experiment (figure 9.7, red and dashed magenta curves) which has identical general features as the twostep experiment but has a much larger potential range in 2 cm depth during the same measuring time.

The same boundary conditions may not produce a second regime switch with any type of soil. If the boundary conditions are changed too early, the plateau just changes its level as the potential difference changes. In this case the threestep experiment does not enhance the estimation of the hydraulic properties. However, it still gives more information about the resistance of the boundary layer. Prior knowledge about the soil hydraulic properties is required to choose the optimal boundary condition steps in a multistep experiment. This knowledge could be obtained by first performing a onestep experiment and then using this information to design a multistep experiment. However, a major disadvantage of this scheme is the long time required to conduct two experiments.

To study the importance of water vapour flow inside the soil compared to the flow of liquid water, also a simulation where the effective conductivity contributed by water vapour flow was disabled was performed (figure 9.7). Water vapour flow inside the soil is especially important at later times, when the soil becomes very dry after the third step of the boundary condition at $t = 196.3$ h. Without the water vapour transport the hydraulic conductivity is already too low to get an increase of the evaporation flux when the vapour pressure at the boundary is reduced. The sample is effectively sealed by a very dry layer at the sample surface with very low conductivity, which prevents the further drying of deeper regions. A second feature not present in the simulation without vapour transport in the soil is the “undershoot” of the evaporation at the transition back to boundary layer dominated regime at $t = 62.5$ h. With vapour transport, the evaporation before the switch is higher and thus the potential at the surface is lower. This results in a more pronounced drop of the evaporation and a longer time till the dynamic equilibrium in the soil is reached again.

Sensitivity analysis

While there is no big change in the relative sensitivities for the potential and the water content 2 cm below the soil surface compared to the onestep experiment, the sensitivities of j_w increase substantially for experiments that re-enter regime I (figure 9.8). Multistep experiments which do not re-enter regime I show no strong effect in the sensitivities (data not shown here). Thus, with the restriction that the effect of hysteresis must be investigated in future studies, multistep experiments are a good tool to increase the sensitivity. Considering the larger potential range which is covered by the threestep experiment, it is also considered superior to the corresponding twostep experiment.

To determine the optimal position of a tensiometer or permittivity probe, profiles of relative sensitivity have been analysed. Figure 9.9 (top) shows profiles of the relative sensitivity of the matric potential to changes in α . Generally, the sensitivity is lower near the sample surface especially at later times. After the onset of regime II, a large sensitivity peak appears, which moves downward with time. At $t = 44$ h the sensitivity drop below the peak even leads to a zero-crossing of the sensitivity. After the increase of water vapour pressure at the upper boundary and the transition back to regime I the sensitivity peak vanishes for a short time and reappears after the fall-back to regime II.

The peaks are located at the drying front (as can be seen for $t = 176.9$ h from a comparison

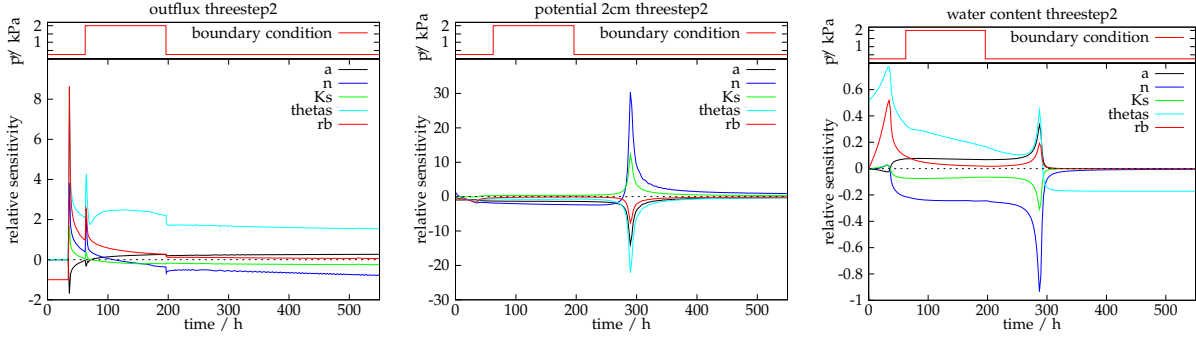


Figure 9.8: Sensitivity coefficients $s(t)$ for the threestep experiment for the outflux (left), the potential in 2 cm (middle) and the water content in 2 cm (right).

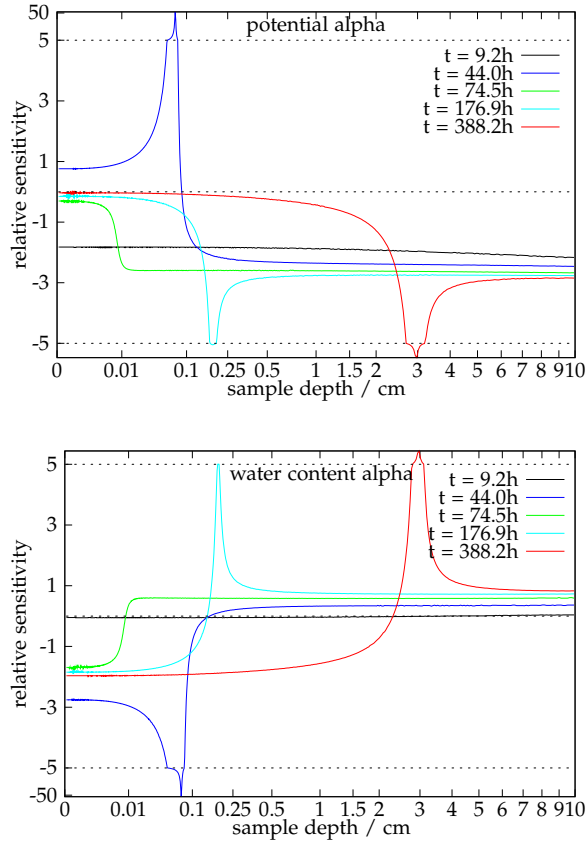


Figure 9.9: Relative sensitivity $s(z, t_i)$ on α of the potential ψ_m (top) and the water content θ (bottom), respectively, versus height z for different times t_i , both for the threestep experiment. Note the switch in the sensitivity scale, marked with a dashed black line, and the non-linear scaling of the depth axis.

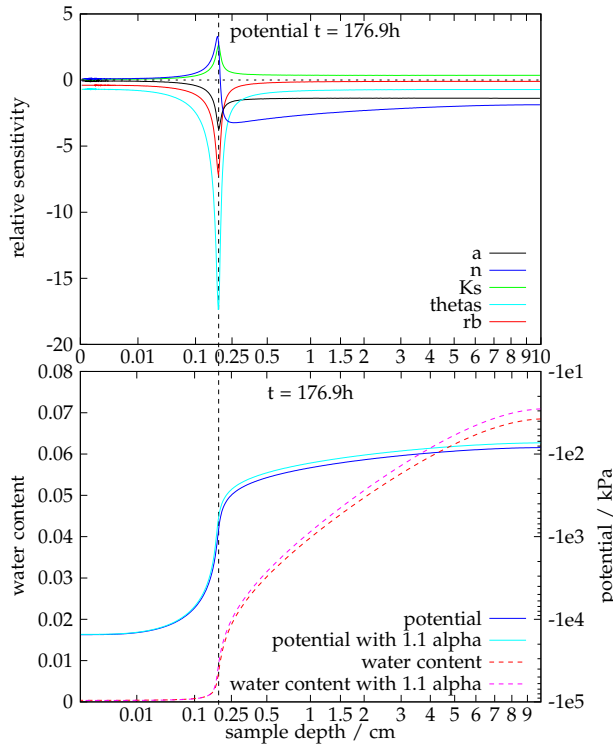


Figure 9.10: Relative sensitivity $s(z, t)$ of the potential ψ_m on all variables at $t = 176.9h$ (top). The potential and water content profiles for the undisturbed parameters as well as with α changed by 10 % (bottom) illustrate the source of the huge sensitivity peaks at $z \approx 0.15$ cm in the plot above, the location is marked with a dotted vertical line. Note the non-linear scaling of the depth axis.

with figure 9.10) where the gradients are particularly high and thus small deviations in the parameters lead to large changes in the solution. As changes of the parameters also affect the position of the drying front, small parameter changes generate huge potential differences in its proximity. Figure 9.10 illustrates the change in the profiles of matric potential and water content at $t = 176.9h$ if α is increased by 10 %. A zero-crossing of the sensitivity occurs if the parameter change leads not only to a different position, but also to a change in the steepness of the drying front.

This is in accordance to Romano and Santini (1999) who reported that sensitivities of the matric potential in traditional evaporation experiments in the uppermost part of the soil show increasing curvatures and a drop to zero. They noticed that this especially happens at larger times when $\Delta\psi_m$ “changes its sign in close proximity to the evaporating surface”. A change in the sign of $\Delta\psi_m$ corresponds to a change of the sign of the corresponding sensitivity coefficient as it was found in our study.

The sensitivity of the water content to changes in α is similar (figure 9.9, bottom). However, there is always a relevant sensitivity at the soil surface and a zero-crossing which is located above the peak at the drying front for later times.

In principle the profiles of the relative sensitivity are similar for all parameters. As shown in figure 9.10 the profiles of the sensitivity of the matric potential at $t = 176.9h$ for different parameters mainly differ in the size and the sign of the peak. The sensitivity to changes of n has a very pronounced zero-crossing as n always influences the steepness of the drying front. The sensitivity peaks are most pronounced for the parameters θ_s and r_b as these parameters have the strongest influence on the propagation speed of the drying front, by determining the speed of drainage in regime I and thus the starting time of the drying front movement.

For the permittivity probe a position nearer to the surface than the 2 cm used here would be advantageous as the sensitivity of the water content is always high there, but this is hard to realise experimentally. For the tensiometer a depth of 2 cm is quite fine, as the sensitivity at this depth is high except for very late times after the passing of the drying front. When the drying front passes, the potential drops so low (-10×10^4 kPa – this corresponds to -1 km

water column – or less, see figure 9.10, bottom) that it is outside the measurement range of traditional tensiometers. Thus, for a real measurement tensiometers have to be removed before the drying front passes to avoid a leakage of water into the soil and therefore the sensitivity peaks of potential cannot be utilised with traditional tensiometers.

Especially the sensitivity peaks of the water content measurements open up the possibility to “scan” different sample layers during one experiment, as the sensitivity is focused on a very small height interval and penetrates with time. If the soil has different layers, the sensitivity would penetrate through these layers with the drying front. However, measurements of the whole water content profile would be necessary to exploit this. If only point measurements are available, the gathered information is not sufficient to distinguish influences at different depths, i. e. if changes in the drying front propagation are caused by another layer above or by different parameters of the same layer. The measurement of water content profiles could be done e.g. by X-ray, neutron or gamma adsorption.

9.1.3 Response surfaces

Figure 9.11 shows the response surfaces of the threestep experiment with j_w and ψ_m as target variables. The tensiometer was removed at -30 kPa. Generally, there is a single global minimum and it is relatively well-defined. Only the combinations (α, K_s) , (n, K_s) , (K_s, θ_s) , and (K_s, r_b) where K_s is involved have small valleys and the slope to the absolute minimum in the direction of the valleys is low. This is a consequence of the low sensitivity of the measurements to changes in K_s . As $n = 2$ and the van Genuchten/Mualem parametrisation does yield non-physical hydraulic conductivity functions for $n < 2$ (Ippisch et al., 2006), the part of the response surface with $n < 2$ must be regarded with care. For the other target variable combinations and boundary conditions the response surfaces look similar and thus are not all shown here.

To identify the essential measurements for a good estimation of the parameters the response surfaces of the threestep experiment are analysed for K_s and n for combinations of the three measurement types (figure 9.12). The parameter combination was chosen because K_s is especially hard to estimate due to its low overall sensitivity. If only the evaporation flux j_w is used the residual does not have a well defined minimum but more a banana like shaped extended region. The minimum in the direction of n is better defined if only the matric potential in 2 cm depth is used but there is still a very long valley in the direction of K_s . This valley becomes much shorter if a combination of j_w and ψ_m is used. The addition of permittivity (i. e. water content) measurements improves the situation only in the combination $\epsilon_c + \psi_m$, in all cases involving the evaporation flux the changes are very small. However, it should be noted that accurate permittivity measurements (e. g. with TDR probes) in the dry range are not feasible. The reasons are: (i) Since the traveltime error is constant, the relative error increases with decreasing ϵ_c . (ii) The compound permittivity ϵ_c is a function of the soil matrix permittivity ϵ_{soil} , of the porosity ϕ , and the actual geometry. For low water contents, the uncertainty of the measurement diverges since the permittivity contribution of the remaining water becomes equal to or even lower than ϵ_{soil} , ϕ is not known accurately, and the geometry is unknown. (iii) For thin films of water as found in the dry region, ϵ_{water} is different from the one of bulk water. Additionally, the measurement volume which was neglected in the simulations will smear out gradients, and results are expected to be worse than in the ideal case of a point measurement.

As j_w is the derivative of the total water content, it is reasonable that adding a water content measurement does only give slightly more information. In contrast, ψ_m is an independent observable and therefore gives more information about the soil water capacity curve. However the fundamental difficulties with tensiometers must be regarded. The information is only provided in a small potential window and great care must be taken that the tensiometer is

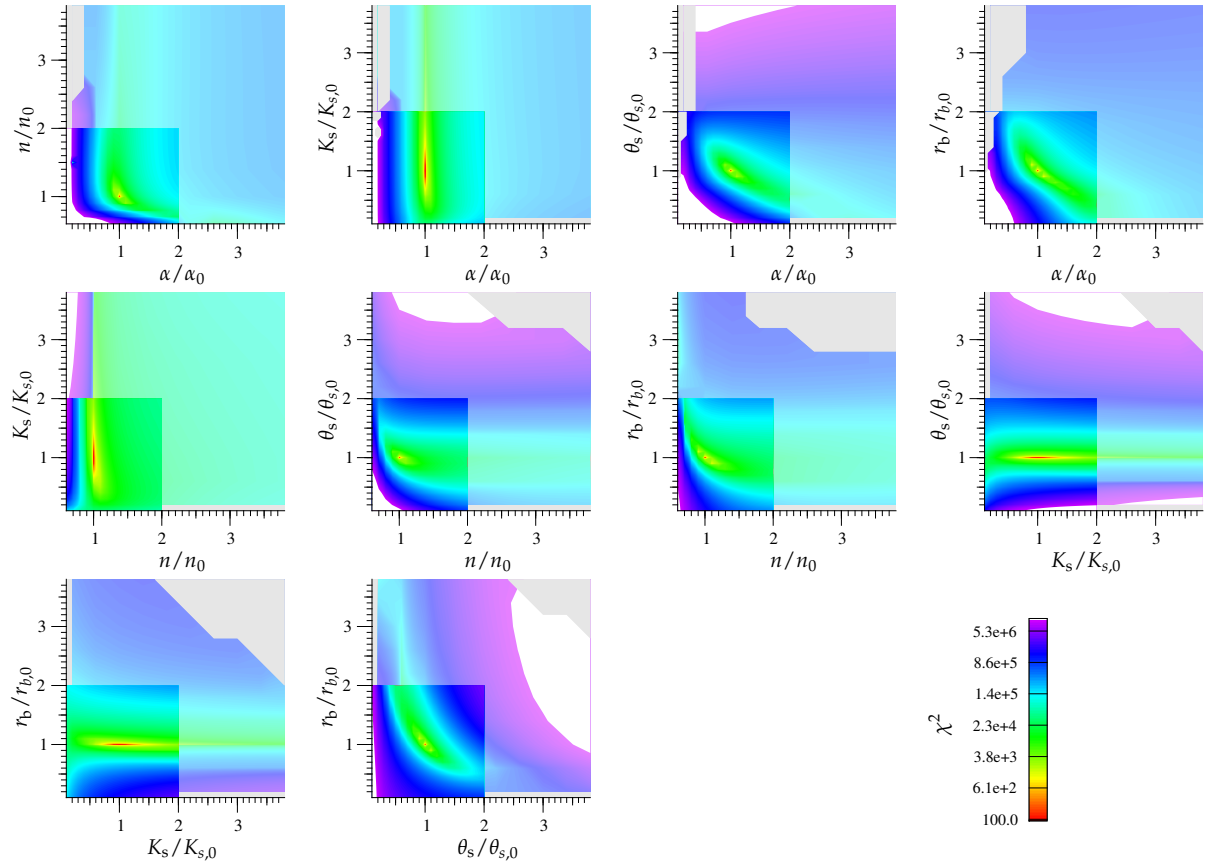


Figure 9.11: Response surfaces for the threestep experiment with j_w and ψ_m as target variables. The light regions with deviation coefficients larger than 2 were calculated with a smaller resolution of 0.4. Non-matching structures between the two regions are caused by the different resolution. Note that a χ^2 above 10^7 is displayed in white, points where the model did not finish after 120000 time steps is marked grey.

removed before the potential in the soil drops below its air entry point, or the water vapour pressure, whatever is higher. Therefore it was also investigated whether a tensiometer which can measure up to $\psi_m = -70$ kPa or even without a limit makes a considerable difference. The results on the response surface are illustrated in figure 9.13. The lower the tearoff of the tensiometer, the better the data. Going from $\psi_m = -30$ kPa to -70 kPa gives a significant enhancement. For the hypothetical case of unlimited potential measurement, the minimum is so localised that χ^2 is above the upper colour scale limit for all values but the minimum itself (data not shown). Apparently, tensiometers with a much wider range of measurements like the ones presented by Bakker et al. (2007) would be most helpful for this type of experiment.

9.1.4 Convergence study

The convergence study revealed the existence of local minima which are not visible in the response surfaces, because they are not located in two-dimensional sub-planes of the five-dimensional parameter space. When initially running the inverse fits with potential and outflux data, the convergence was poor. The reason was that when the outflux peaks did not fit initially and the algorithm slightly modified the solution vector to improve it, the potential in the new solution fitted worse. Because the potential has a relatively low standard deviation, the residuum was not improved and therefore the inverse fit could not improve the solution.

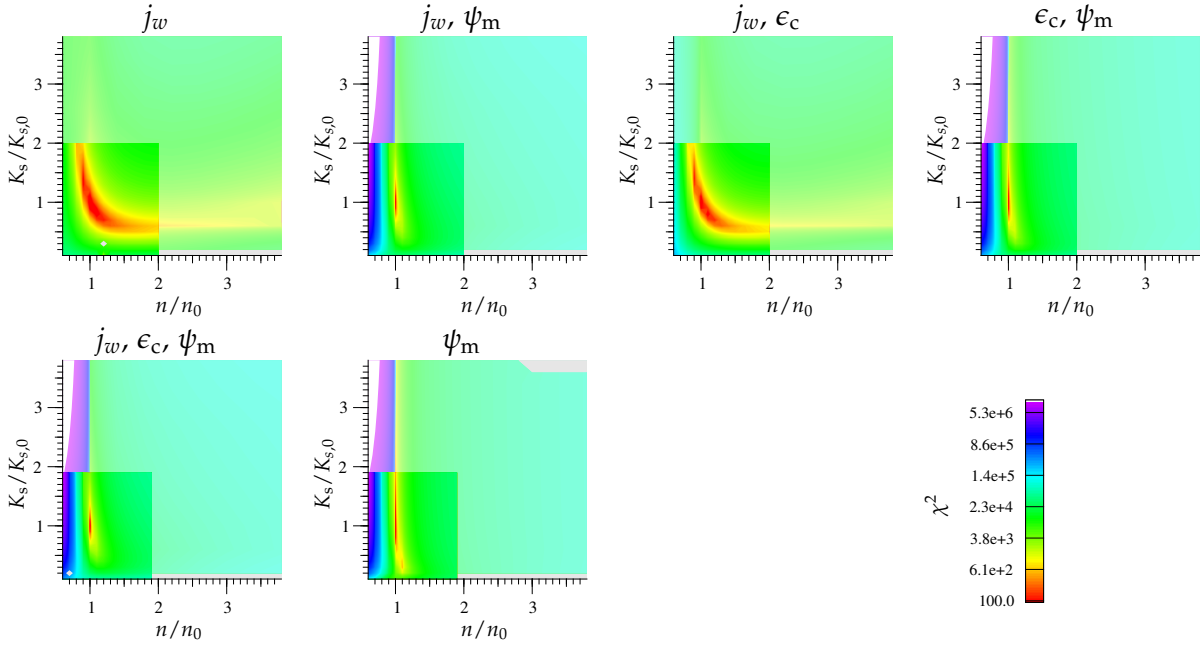


Figure 9.12: Same as figure 9.11 but with different target variable combinations. Note that real permittivity measurements have large errors in the dry range, which would lead to much worse results for the case $\epsilon_c + \psi_m$.

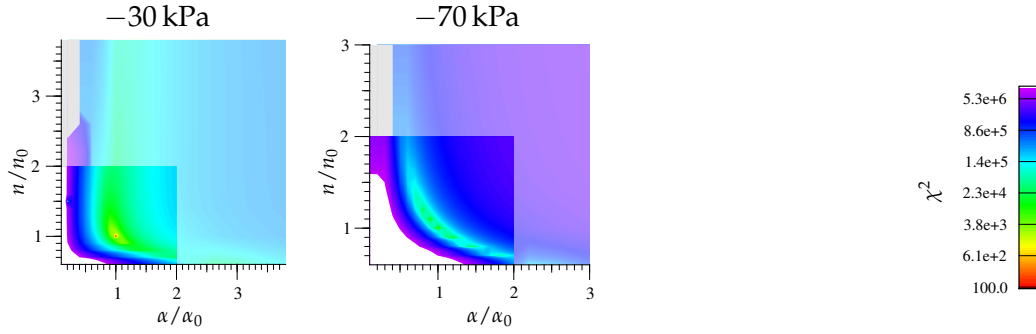


Figure 9.13: Same as figure 9.11 but for different tensiometer tearoff values ψ_{limit} with j_w and ψ as target variables.

To resolve that problem, the inversion was first run with the outflux data only to obtain a reasonable start parameter set. After convergence, the potential data was added and the inversion was restarted.

Using that modified approach there were still a few non-converged fits. However these could be identified clearly because the system response was apparently not fitting the data. These fits were taken out manually. The criteria for sorting out were (1) the outflux peaks were not represented, (2) the potential systematically deviated by more than 10σ (this corresponds to 1 kPa), and (3) the outflux plateau at the beginning of the experiment was not reproduced at all. For criterium (3), if a fit was just at the limit it was accepted to avoid too strict sorting. figure 9.14 shows a comparison between the accepted fit with the highest residuum and the rejected fit with the lowest residuum for all investigated soils.

Using this procedure, the parameters were reasonably reproduced. The mean and standard deviation of the estimated values for each parameter is given in table 9.1, the hydraulic functions for all converged fits together with the ones for the true parameters are shown in fig-

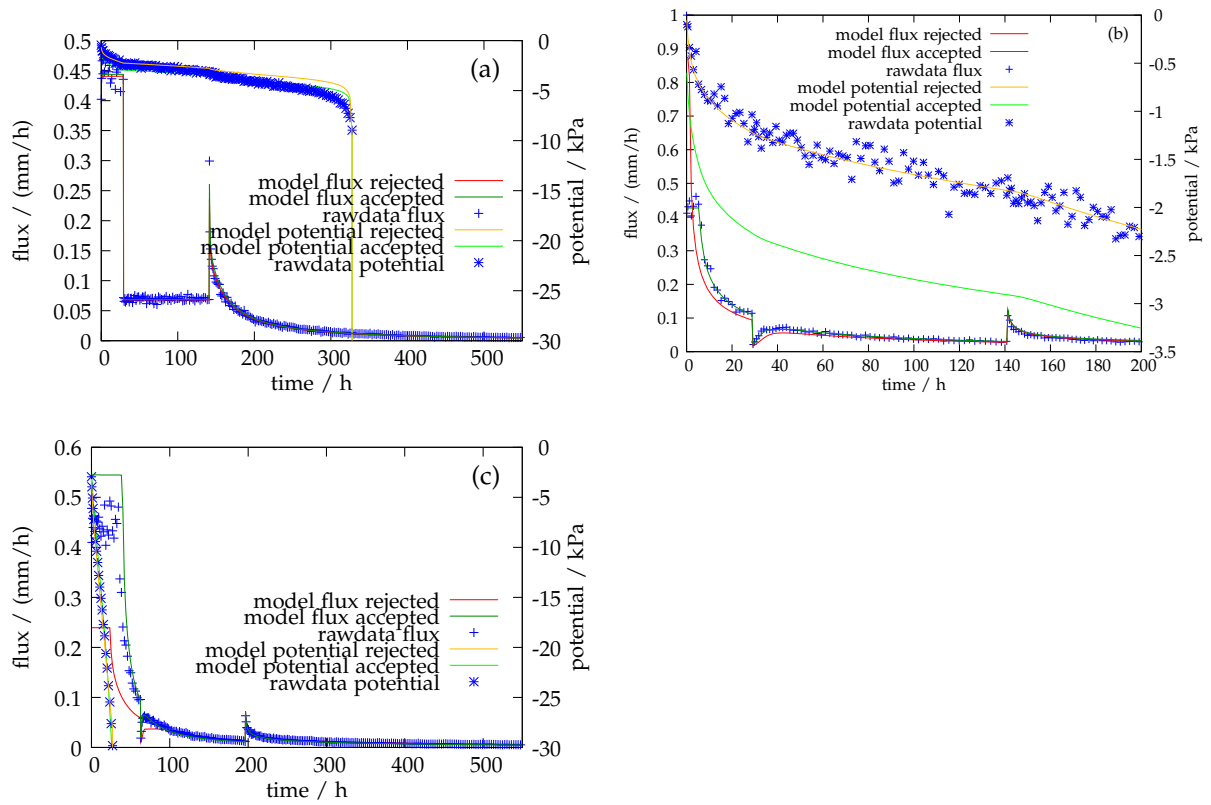


Figure 9.14: Comparison between the accepted fits with the largest residuum and the rejected fits with the lowest residuum, for sand (a) (residuum 4900 (accepted) vs. 9700 (rejected)), loamy sand (b) (residuum 28000 vs. 3300), and silt (c) (residuum 4600 vs. 7300), respectively. The reason for rejecting the curves were for (a) the deviation of the tensiometer, for (b) the first plateau was not represented, and for (c) the first peak was not represented, respectively. Note that for (b), although the potential of the accepted fit matches worse than in the accepted one, its deviation is just at the sort-out limit and limit cases were still accepted, while the plateau in the rejected fit is definitely not represented at all.

ure 9.15. The deviations in each column of table 9.1 were calculated from the ensemble mean of converged results. When comparing the results of the inversion and its standard deviation calculated from the analysis of the sensitivity matrix with the real parameters and the variance encountered by the different fits in the ensemble, one can see that the standard deviations calculated from the analysis of the sensitivity matrix are too small in almost all cases. This is attributed to the fact that the sensitivity matrix does only give a linear approximation of the uncertainties. Deviations of the resulting fit parameters are quantified by the deviation coefficient d . No systematic deviations are found for the silt, however small systematic deviations are present for the sand and the sandy loam. This is expected, since the evaporation method is most sensitive in the dry range, where dynamics of the silt actually take place, but it is relatively insensitive for the wet range, the region of most of the dynamics of the sand and sandy loam soil. The silt has the smallest deviations since its dynamics are reaching most into the dry region of the three soils. This is also confirmed by the plots. For the water characteristic of the sand with 42 fits, 38 practically overlap while 4 deviate quite a bit. The sandy loam has the most significant deviations of all three soils in the water characteristic, only 12 of 27 fits are practically overlapping and 3 have small deviations, while the remaining 12 fits are relatively

Table 9.1: Results of the convergence study with j_w and ψ_m as target variables. For the resulting parameters and standard deviations from the model output, as well as the deviation coefficients d defined by equation (8.4), mean and variance of the inversion of the start parameter sets with each of the data sets are shown.

Sand (42 fits)					
parameter	unit	real value	result	standard deviation	deviation coeff. d
α	m^{-1}	5	4.90 ± 0.04	0.024 ± 0.001	-0.020 ± 0.009
K_s	cm h^{-1}	2	2.11 ± 0.07	0.014 ± 0.002	0.05 ± 0.04
ϕ	$\text{m}^3 \text{m}^{-3}$	0.3	0.2985 ± 0.0003	0.00076 ± 0.00001	-0.0049 ± 0.0009
r_b	mm	3	3.013 ± 0.003	0.0086 ± 0.0002	0.004 ± 0.001
n	–	4	4.2 ± 0.1	0.0218 ± 0.0008	0.05 ± 0.02
Sandy loam (27 fits)					
parameter	unit	real value	result	standard deviation	deviation coeff. d
α	m^{-1}	10	8.6 ± 0.5	0.045 ± 0.005	-0.14 ± 0.05
K_s	cm h^{-1}	0.1	0.08 ± 0.01	0.0010 ± 0.0001	-0.2 ± 0.1
ϕ	$\text{m}^3 \text{m}^{-3}$	0.3	0.309 ± 0.003	0.00090 ± 0.00001	0.03 ± 0.01
r_b	mm	3	2.83 ± 0.08	0.020 ± 0.002	-0.06 ± 0.03
n	–	2	2.18 ± 0.07	0.0062 ± 0.0004	0.09 ± 0.03
Silt (34 fits)					
parameter	unit	real value	result	standard deviation	deviation coeff. d
α	m^{-1}	0.5	0.503 ± 0.003	0.0012 ± 0.0001	0.006 ± 0.006
K_s	cm h^{-1}	0.1	0.12 ± 0.02	0.0007 ± 0.0002	0.2 ± 0.2
ϕ	$\text{m}^3 \text{m}^{-3}$	0.3	0.298 ± 0.003	0.00074 ± 0.00005	-0.01 ± 0.01
r_b	mm	3	3.02 ± 0.05	0.011 ± 0.001	0.01 ± 0.02
n	–	2	2.000 ± 0.009	0.0052 ± 0.0004	0.000 ± 0.004

bad. The silt is nicely represented in all fits, only 2 of 34 fits are deviating slightly. Considering that the estimation of the hydraulic conductivity is difficult, the conductivity function is well represented for all soils. The analysis shows that applying the Monte-Carlo Levenberg-Marquardt approach for real experimental data would lead to a reasonable estimate of the parameters, since most fits converge to the correct parameters and the few deviating fits could be identified in the ensemble.

9.2 Inversion of experimental data

The sample was taken from a field site directly into the PVC sample cylinder. It was then slowly saturated from below with deionised water with $0.352 \text{ mmol/l CaCl}_2$ until the hydraulic potential at the bottom of the sample was equal to the static gravimetric pressure of the sample height η , $\psi_w = \rho_w g \eta$, and no more water was flowing in. Then bottom and top of the sample were closed and the sample was allowed to equilibrate for some days. Finally the evaporation chamber was installed on top and the experiment was started. It was run for 640 h.

Notice that while the boundary condition (figure 9.16 upper frame) suggests that the experiment only covers very low potentials – the maximal value of the potential boundary condition is -39 MPa – in the soil itself the whole potential range from 0 kPa onwards is encountered as the experiment started at saturation.

The measured flux and potential is shown in figure 9.16. Integrating the flux resulted in a cumulative outflow of $(27.1 \pm 1.3) \text{ mm}$. Weighing the sample before and after the experiment yielded $(26.5 \pm 0.5) \text{ mm}$. The aberration in the boundary condition around $t = 300 \text{ h}$ was

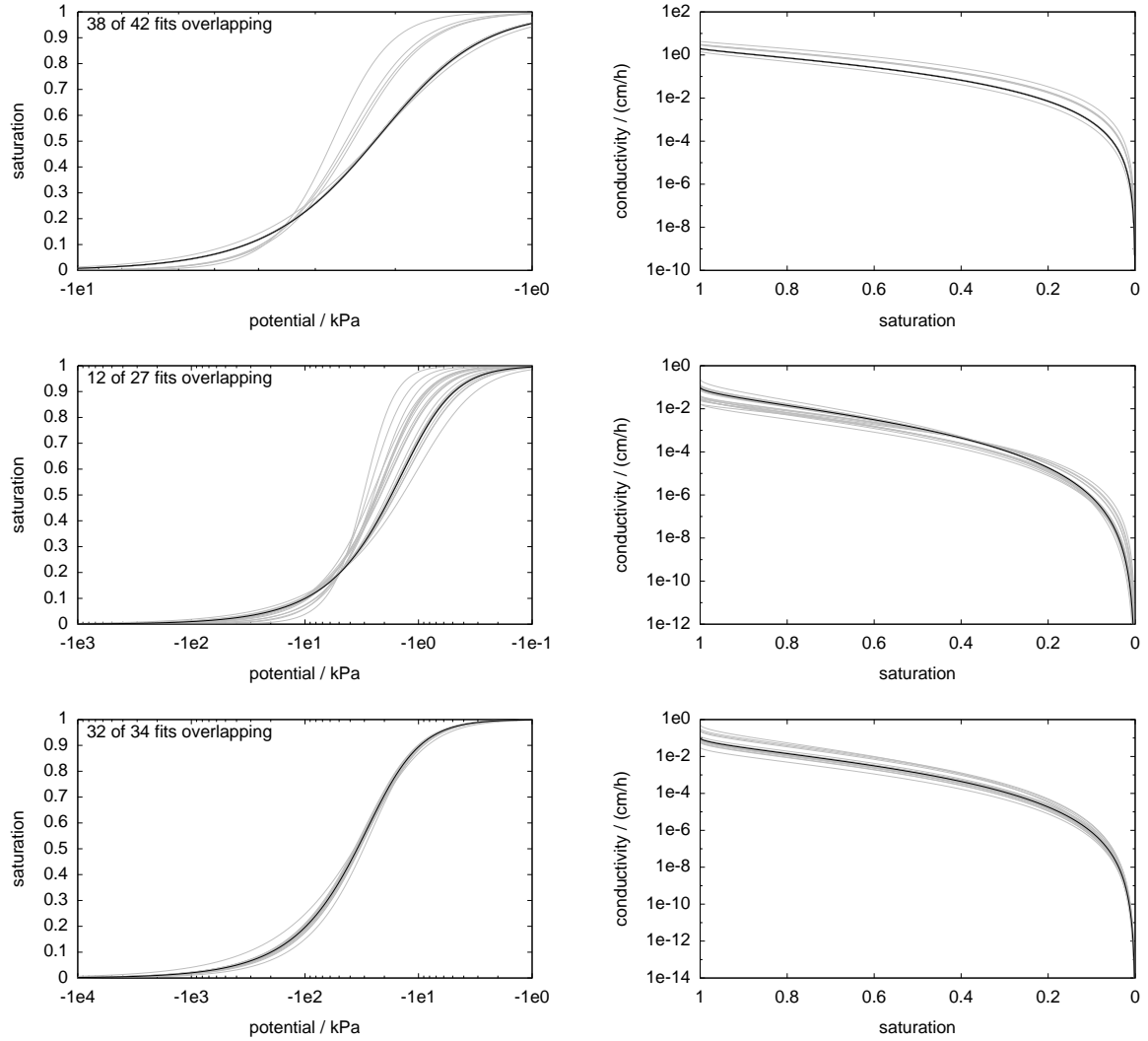


Figure 9.15: Hydraulic functions of the inverse solutions (gray), and the functions for the true parameters (black line), for the sand (top, 42 fits, 38 practically overlapping), sandy loam (middle, 27 fits, 12 practically overlapping), and silt (bottom, 34 fits, 2 deviating slightly), respectively.

caused by heating of the laboratory. Its result on evaporation can be seen in the flux rate and further corroborates the correctness of the data.

Figure 9.18 shows the total water content, measured by TDR. The permittivity of the soil matrix was determined based on volumetric porosity and water content measurements as well as TDR bulk permittivity measurements on the field site when the sample was taken during an excavation. The total change in water content is (25 ± 3) mm. This is consistent with the flux and weight measurements. The figure also shows the mass balance $\eta\theta(t) + \int_0^t j_{w,\text{exp}}(\tau) d\tau$, where η denotes the height of the sample. It is assumed to be constant during the whole measurement. The decrease at the beginning is attributed to larger pores around the TDR rods which were created by the insertion of the probe, analogous to the sand sample. Because these larger pores are drained first, the measured water content was not representative for the whole sample. As the potential became more negative, the smaller pores in the undisturbed part of the sample were drained as well and the real water content again matched the one measured by TDR.

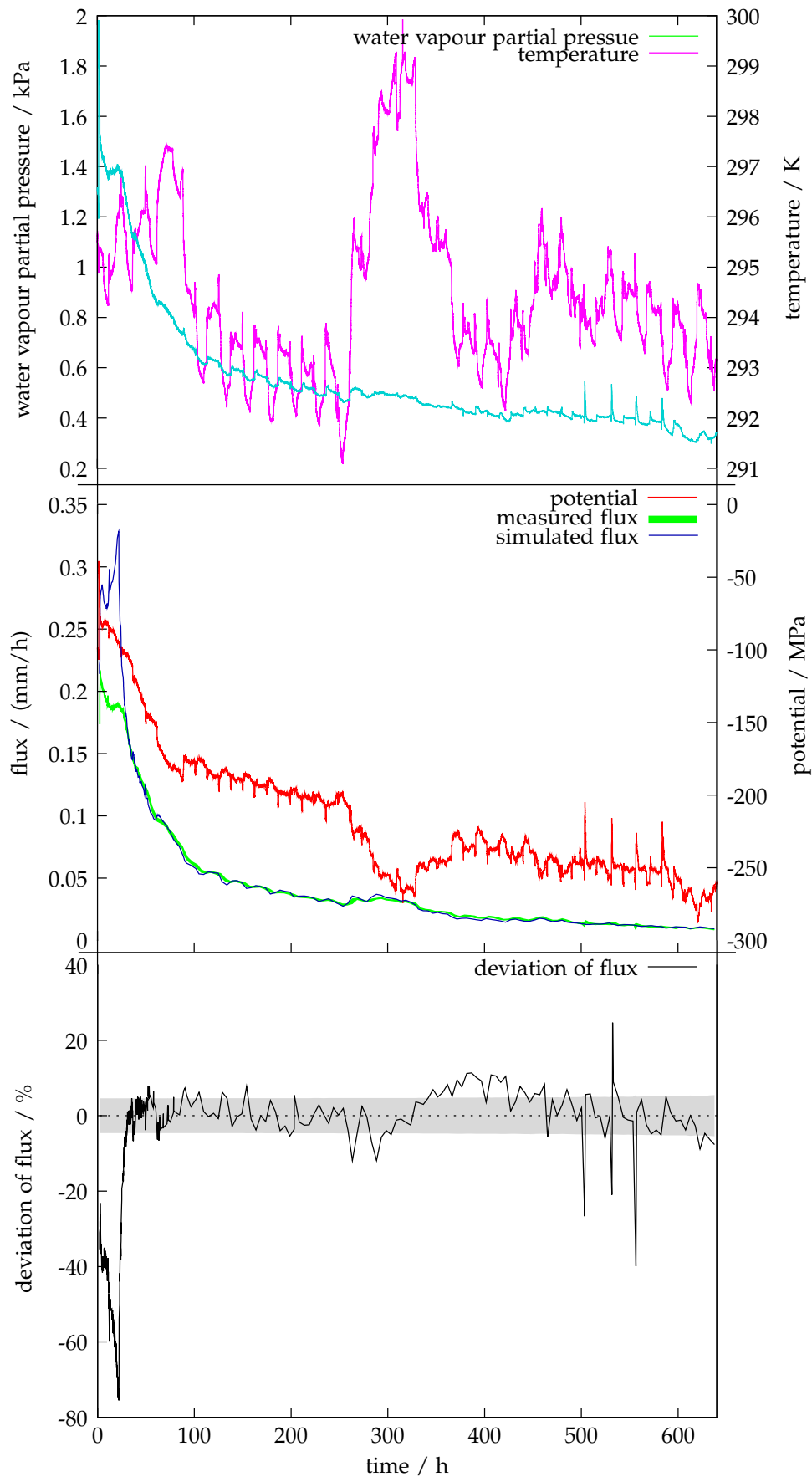


Figure 9.16: Results of the evaporation experiment with the undisturbed soil sample. Matric potential of the air (boundary condition) was calculated from water vapour partial pressure and temperature (upper frame). Notice that measured and simulated flux (middle frame) practically overlap. The relative deviation $(j_{w,\text{exp}} - j_{w,\text{model}})/j_{w,\text{exp}}$ is shown in the bottom frame together with the measuring uncertainty from figure 9.17

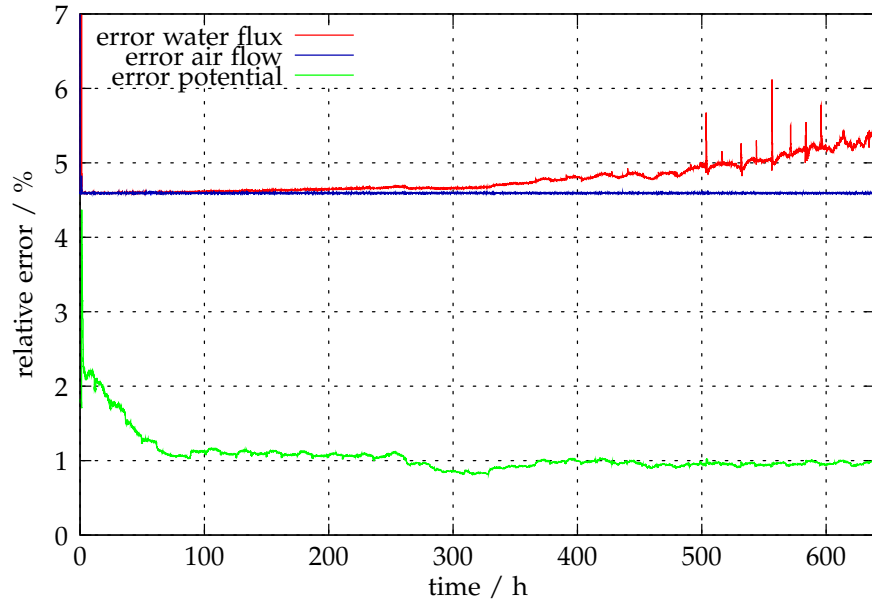


Figure 9.17: Relative error (1σ) of the evaporation measurement.

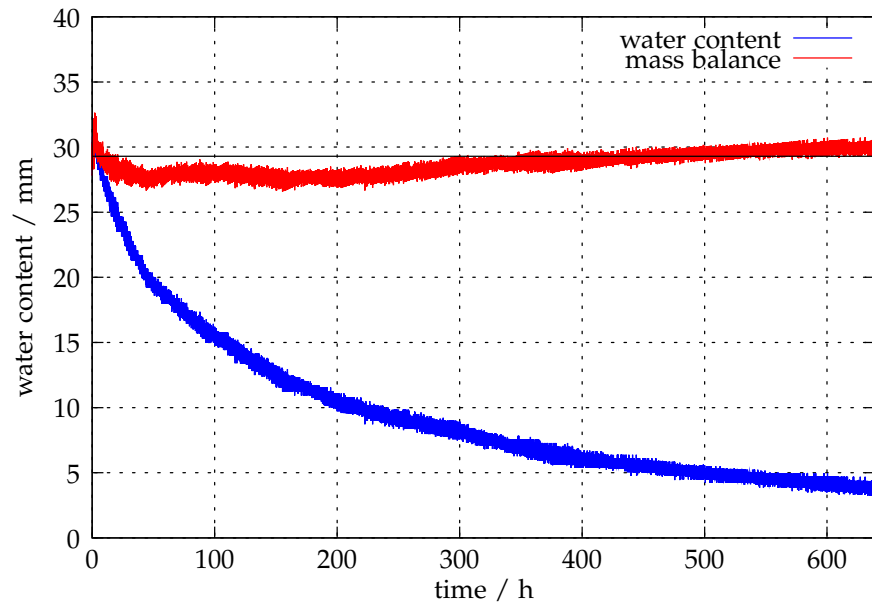


Figure 9.18: Total water content (expressed as equivalent height of water in the soil column) of the undisturbed soil sample, measured by TDR (blue), and the mass balance $\eta\theta(t) + \int_0^t j_{w,\text{exp}}(\tau) d\tau$ (red). The black horizontal line represents the initially measured water content.

Table 9.2: Resulting parameters and their standard deviations resulting from the analysis of the sensitivity matrix for the undisturbed soil sample.

parameter	fitted value	standard deviation
λ (–)	1.03	± 0.1 (10 %)
ψ_e (Pa)	1020	± 40 (4 %)
K_s (cm/h)	0.116	± 0.009 (8 %)
$\theta_s - \theta_r$ (m ³ /m ³)	0.306	± 0.002 (1 %)
r_b (mm)	2.74	± 0.05 (2 %)

The relative error is shown in figure 9.17. In this particular experiment, the absolute pressure at startup had an error of 1 kPa. The error of the potential is 1 % to 2 % and water flux error 4.5 % to 5 %. This high data quality is crucial for the quality of the result of the inversion process.

Hydraulic parameters were estimated from the measured values using the inverse model described in section 8.2. Only evaporation rates were used as target variables. Fitted parameters are the Brooks-Corey parameters λ and ψ_e , the saturated hydraulic conductivity K_s , the available water content $\theta_s - \theta_r$, and the effective thickness r_b of the boundary layer. The value of τ was fixed at 0.5 as suggested by Mualem (1976). The amount of data used for the inversion process was reduced by filtering to keep the model runtime reasonable. Data points were retained if the difference to the last kept point was $\Delta p_w \geq 15$ Pa partial pressure or $j_w \geq 0.005$ mm/h water flux, or if the time step Δt was greater than 5 h. A grid convergence study gave a necessary spatial resolution of 0.125 mm equivalent to 880 grid points.

Figure 9.16 illustrates that for times $t > 30$ h, the optimised model response is in excellent agreement with the data. For shorter times, however, the model is obviously not capable to describe the data. This may be explained by thermal processes that are not represented in the model. At the start of the experiment, the entire column is in thermal equilibrium. With the onset of evaporation, latent heat is consumed right at the saturated soil surface. As a consequence, the temperature drops there and with it the vapour pressure of water. Hence, with equation (8.1), the evaporation flux will be reduced. In its current formulation, the model does not include the effect of latent heat and therefore yields a gross over-prediction of the evaporation flux. An analytical estimation of the order of magnitude of that effect fitted well with the observed one. As a consequence, the data for $t < 30$ h were not used for the inversion.

Once the soil surface starts to dry, the effect of latent heat drops very rapidly because: (i) the evaporation flux becomes smaller, due to the decreasing hydraulic conductivity, and with it the rate of latent heat consumption, (ii) with the rapid widening of the region where evaporation occurs, heat is extracted from a larger volume such that the local temperature depression decreases, and (iii) in the initial phase the strongly temperature-dependent boundary layer limits the evaporation flux whereas later the less temperature-sensitive hydraulic properties become determining. This leads to the eventual excellent agreement between model and data. In particular the aberration caused by the heating of the laboratory is easily reproduced by the model. While such an agreement is no prove that the model is correct, it is a strong hint that it may be used as an effective representation of the real system under similar conditions as those encountered during the experiment. The resulting parameters and their standard deviations resulting from the analysis of the sensitivity matrix are given in table 9.2. While no independent confirmation is available, they appear reasonable for the soil under examination.

10 Conclusions

Data of an undisturbed soil sample measured with the novel evaporation experiment presented by [Schneider \(2005\)](#) were inverted with the numerical model presented in section 8.2 to obtain soil hydraulic properties. The excellent agreement between measured and simulated data for longer times are a strong indication that all the relevant processes, within our window of view, are captured by Richards' equation with an effective hydraulic conductivity function that explicitly incorporates vapour transport and with a constant effective diffusive boundary layer at the soil-atmosphere interface. For shorter times, the discrepancy between data and model requires the inclusion of additional processes. Qualitatively, the deviations can be understood in terms of latent heat consumed by the evaporating water. In the current model formulation, this process and the associated transport of heat is not included.

Virtual experiments for sand, sandy loam and silt parameters were conducted with the numeric model. In accordance with experimental results from the literature, the physical analysis of the evaporation experiment model showed two different evaporation regimes. In regime I the outflux is limited by the diffusive soil-atmosphere boundary layer, while in regime II it is limited by the hydraulic properties. The sensitivity analysis showed that the sensitivity is especially high at the transition between the two regimes. In regime II, a drying front penetrates through the soil. At the location of this front, sensitivity is very high but this sensitivity peak can only partially be exploited as the potential drops below the measurement range of traditional tensiometers. However, if a profile measurement of water content is available the sensitivity peak allows to scan different layers of the sample. Positioning the tensiometer and the permittivity probe 2 cm below the sample surface gives a good sensitivity over the whole experiment.

As the transitions between these two regimes can be induced by boundary condition changes, a three-step experiment, where the water vapour pressure at the boundary is temporarily increased after regime II was reached, yields an increased sensitivity and a high measurement range.

The analysis of the response surfaces exhibited a single minimum for all parameters, which was mostly well localised. The estimation of the saturated hydraulic conductivity was improved significantly by adding a potential measurement to the outflux measurement. In contrast, adding a water content measurement yields no noticeable improvement. A combination of water content and tensiometer measurements gave nearly as good a response surface as a combination of evaporation flux and tensiometer measurements. However TDR probes which are typically used for water content measurements are no point measurements but have a rather high sampling volume, and the uncertainty for very dry conditions diverges. Therefore it is expected that for real measurements the response surface will get much worse. Furthermore uncertainties in porosity and soil matrix permittivity would introduce additional errors which were not considered in the simulations. Hence measuring j_w will probably be much more accurate in real experiments. There is a significant improvement if the measurement range of the tensiometer is extended to -70 kPa. A hypothetical tensiometer with unlimited range shows a huge advantage. This emphasises the importance of extended-range tensiometers.

When excluding obviously diverged fits, the inverse model converged for the silt to correct solutions from all initial values. For the sand and sandy loam, some fits of the ensemble deviated significantly while others converged reasonably well to the correct parameters. The

deviations were attributed to the small sensitivity of the evaporation experiment in the wet range. Despite of these difficulties, usage of the Monte-Carlo Levenberg-Marquardt approach allows a meaningful parameter estimation. Comparing the hydraulic functions estimated using the mean MCLM estimated parameters with the true functions showed that only the water characteristic of the loamy sand deviated slightly and all others overlapped. The distribution of the parameters caused by measurement errors and crosscorrelation of the parameters was acceptable.

Part III

The capillary fringe

11 Introduction

The capillary fringe (CF) is the interface region between the water saturated groundwater and the partially water saturated vadose zone. Here, the transition from primarily vertical flow in the vadose zone to horizontal flow in the groundwater occurs. Due to the irregular forcing by rain and groundwater level changes, the CF is highly dynamic.

It is expected that the CF has a major impact on flow and transport from the vadose zone into the groundwater. Due to the dynamic nature, air is entrapped in the CF region, which probably leads to a relatively high supply of air and thus also oxygen, which is needed by aerobic micro-organisms. On the other hand, water content is high, which is also beneficial for micro-organisms. Therefore it is probable that the CF plays a major role for the biological degradation of infiltrating contaminants. This has a large practical environmental impact. Biological activity is tightly coupled with chemical reactions and hydraulic transport. To quantitatively understand and model these linked processes, it is a prerequisite that the physical processes of flow and transport in the CF under these transient boundary conditions are known. Biologic products like surfactants are known to change the hydraulic properties and therefore back-link the biological activity to the hydraulic processes, which additionally complicates the system.

The CF was for a long time of limited interest for both soil scientists and groundwater hydrologists, as it was just the limit of their respective domain of interest, although sometimes a disquieting one. This is already manifest in the very definition of the CF, or more precisely in the lack of it. In the most strict meaning, the lower end of the CF is the water table and the upper limit is the level where the continuous air phase begins. In soil physics, this is sometimes referred to as the saturated zone (Corey, 1994). In the broadest meaning, the CF is bound from below by the level where all the entrapped air has become dissolved and from above by the maximum height of the capillary bound water, which is used synonymously with “residual water content” (Ronen et al., 2000). The thicknesses given by these two definitions are typically orders of magnitudes apart.

Only very few studies provided limited insight into the operation of the CF and its function for diverse processes. The rapid reaction of the CF to infiltration was demonstrated in several lab- and field-experiments, e.g., in the context of river recharge (Gillham, 1984). Lehmann et al. (1998) studied the impact of a fluctuating water table and demonstrated, experimentally and with numerical simulations, that the response of the CF was highly asymmetric. However, this study was done using tensiometers and TDR-probes and the influence of the sampling volume of the TDR probes was neglected. The disadvantage of using traditional sensors like tensiometers and TDR-probes is that measurements with high spatial resolution are very difficult to achieve. However, these are required for detailed process analyses. Also the simulations did not fit to the experiments. This is because Lehmann et al. (1998) used a conventional Richards’ model. As shown in this study, multi-phase and non-equilibrium processes which are not represented in Richards’ equation play a major role in the CF. Nielsen and Perrochet (2000) investigated the response of water table inside a soil column to harmonic forcing of an external water table in steady-state and reported a harmonic response. Since only flow from the external reservoir into the column is involved here, this is expected.

As an alternative to traditional sensors, dye tracer experiments are employed and combined with a qualitative analysis of optical images (e.g., Berkowitz et al., 2004). With this approach the spatial resolution is excellent and is only limited by the resolution of the image sensor. The

price, however, is a rather low quality of the data which typically only support some qualitative assessment of the observed phenomena.

The studies mentioned above and a few others provide perspectives on the range of processes in the CF and on their phenomenology. However, Berkowitz et al. (2004) in their review conclude that there is a lack of flow and transport process studies, particularly at the local scale, and that “the dynamic interplay of flow and transport mechanisms across and within the CF is not well understood”.

The aim of this work is to get an insight into the fundamental physical processes which govern the response of the CF to transient forcing. The dynamics of the CF is investigated experimentally using the light transmission method. A detailed error analysis which propagates measurement errors shows the significance of the results. Although this should be a standard for all measurements, unfortunately it is not common in the field of soil science. Tidwell and Glass (1994) discussed their statistical and systematic errors but did not conduct a complete error analysis. To the knowledge of the author, this is the first light transmission experiment with a complete error analysis. Measurement uncertainties are propagated and systematic error sources are also discussed.

Since per definition the CF is a highly interdisciplinary topic involving hydraulics, transport, chemistry and biology, an interdisciplinary approach is required to understand the system in its full beauty. Therefore, recently a DFG-funded research group was initiated incorporating physicists, micro-biologists, chemists, and mathematicians with the aim of quantitatively modelling the coupled hydrological-chemical-biological processes within the CF. This thesis presents a first part of important physical results necessary to understand this interesting complex and important system.

12 Materials and methods

12.1 Experimental setup – overview

Dynamics of the capillary fringe were observed in a sand sample in a Hele-Shaw cell (section 12.3.2). The water content distribution was observed using the light transmission method (section 12.3) with high spatial and temporal resolution. Since the aim of the experiments was to investigate the dynamics of the capillary fringe under transient boundary conditions, these had to be computer-controlled. For the lower boundary, the water table was set to a defined height using a water reservoir. The reservoir height could be controlled using a stepper motor and a spindle. The water height in the reservoir was kept constant even if water was extracted from or put into the reservoir by constantly pumping water from a second reservoir into it which then flowed back through an outflow in the height of the defined water table. At the upper boundary, a flux boundary condition could be imposed by pumping water with a computer-controlled peristaltic pump. The water was extracted from a reservoir on a digital balance. Weight measurements were used to automatically correct the pump rate. This procedure allowed to accurately define the flux. Temperature was measured at the glass of the Hele-Shaw cell, in the room and at the light source, to exclude temperature effects. A sketch and a photo of the setup is shown in figure 12.1. Pumps, balances, and temperature measurements are described in section 12.8. In some of the experiments, a NIR camera for microscopic measurements was additionally installed beside the visual camera. This also required the use of another light source, as described in detail in section 12.3.

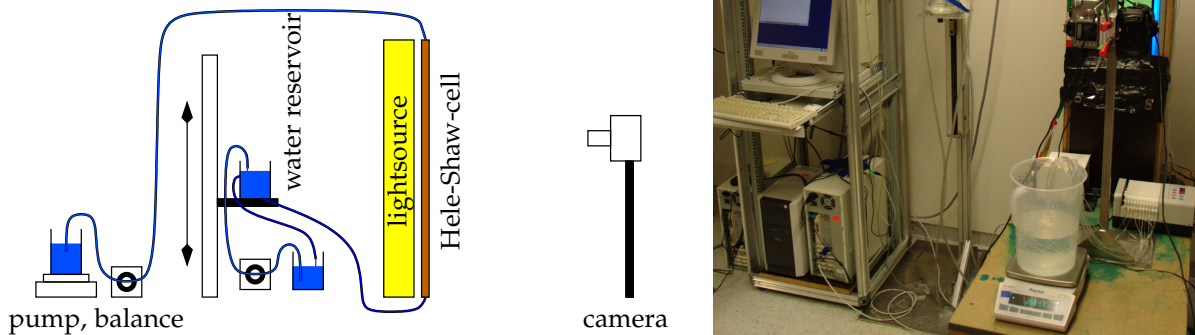


Figure 12.1: Experimental setup used for capillary fringe investigations. The sand soil is contained in a 3 mm thick Hele-Shaw cell. A homogeneous light source illuminates the medium. Transmitted intensity is observed with one or more digital cameras. A stabilised, computer-controlled movable water reservoir defines the lower boundary condition. At the upper boundary, a flux boundary condition can be set with a computer-controlled peristaltic pump coupled with a balance.

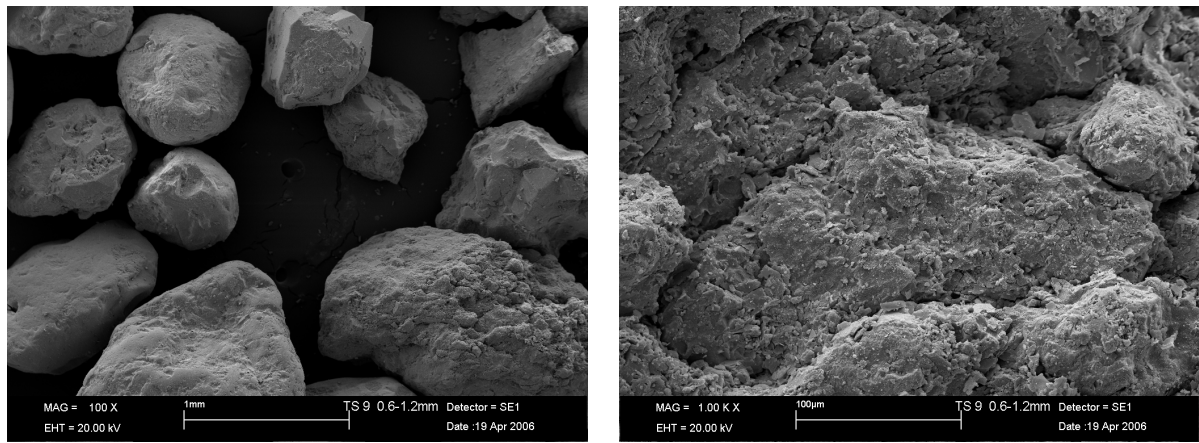


Figure 12.2: REM images of the TS9 sand used in the experiments.

12.2 Porous medium

As porous medium, sieved sand with a grain size distribution between 0.63 mm and 1.25 mm was used. The intention here was that a sand medium is more natural than a glass bead medium, since glass beads have a very smooth surface. Figure 12.2 shows electron-microscopy images of the used sand.

Filling of the medium

Among the available procedures for sample preparation, raining is probably the method which simulates a soil fabric most similar to the one found in natural deposits formed by sedimentation (Rad and Tumay, 1987). When filling the medium into the cell, care must be taken to avoid artificial layering which would affect the results due to structures in the porous medium. Layers can be created when the medium is filled in several steps. Therefore, a filling funnel (figure 12.3) was designed which allowed to fill in the sand in one run. The density of the filled medium can change if the falling height changes due to different specimen velocities. After a certain height, however, velocity will become constant and relative density will not change (Rad and Tumay, 1987). Therefore, below the funnel, an extra falling height of 20 cm was installed. The funnel could be mounted onto the Hele-Shaw cell with a specially designed plastic mount. A toggle allowed opening and closing the funnel. During the filling procedure, the cell was mounted on a shaker which supported the homogeneous grain depletion. After the filling was complete, the cell was again vibrated for settling the grains. This was done to avoid grain settling during the first imbibition/drainage cycle, which would introduce a medium change during the experiments.

REV

To characterise the statistical properties of the medium and obtain the REV size, the semi-variogram was estimated. The visual images from the prewetted fluctuating water table measurement were used for the estimation. Image resolution was (0.346 ± 0.001) mm/pixel. The Bayer downsampled dry and wet images were divided by the lightsource image. This step eliminated possible trends from inhomogeneous illumination. A 100×100 sub-region was cut from the normalised images and the semivariance was estimated. Figure 12.4 shows the result. The correlation length of the medium is about 2.5 mm for the dry and 3 mm for the wet sample (vertical black lines). The data of the dry image had a skewness of 1.6 and the wet image of 0.6. The skewness quantifies the deviation from a Gaussian distribution. 1.6 is at the upper range of acceptable values. Due to that deviation from a Gaussian distribution, the correlation

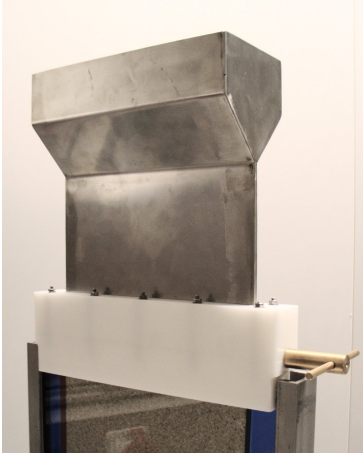


Figure 12.3: Filling funnel for Hele-Shaw cells.

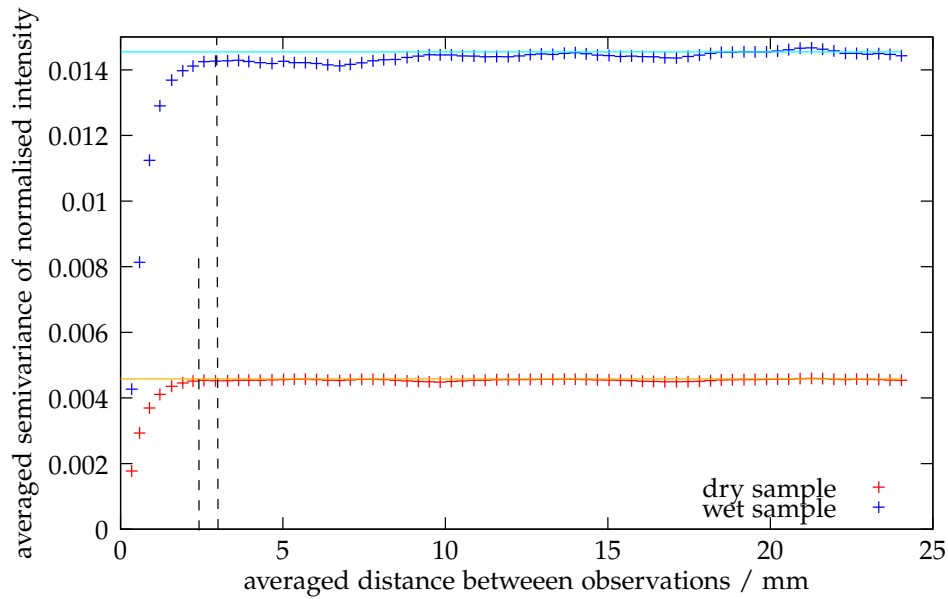


Figure 12.4: Averaged semivariance of the normalised intensities of the dry and the wet sample.

length may be slightly underestimated. This effect is more pronounced for the dry sample, which may be an explanation for the smaller correlation length there. The skewness may result from the used consumer camera which is known to be non-Gaussian. Accordingly, 3 mm is considered to be the more appropriate correlation length. The wet sample has a larger variance due to higher contrast between wet spots and grains than dry spots and grains. The estimated correlation length is reasonable for the grain size distribution in the medium, since the correlation length is expected to be about a few grain diameters. The width of the point spread function (standard deviation of fitted Gaussian) was estimated to $\sigma = (1.6 \pm 0.2)$ mm, which is smaller than the correlation length and indicates that the correlation is not only caused by multiple scattering, although a part of it may be attributed to that process. For macroscopic water content estimation, the radius of the averaging area should be at least a bit larger than the correlation length. The small variations still present at higher distances in the semivariogram are attributed to small heterogeneities in the medium which cannot be avoided and small changes in cell thickness.

12.3 The Light Transmission Method

12.3.1 Introduction

As described in section 5, light transmission can be used to quantify water content in thin cells with porous materials, since the transmission and reflection properties depend on water content. This allows high resolution measurements spatially and temporally, which is a prerequisite for assessing the processes in the capillary fringe. Processes are expected to be highly dynamic and feature spatial fluctuations and large gradients.

In normal measurements where radiative transfer is used as a proxy for water content, the exact matrix-water-air geometry is unknown (if it was known, there would be no need for the measurement). Thus the relation between measured intensity and water content has to be identified. Several approaches can be found in the literature. The first light transmission method (LTM) by Hoa (1981) was only a point measurement. He used a diaphragm to select a small light bundle and measured transmitted light with a photo resistance. For the evaluation, he assumed normal light incidence and modelled the porous medium as parallel boxes with different refractive indices. Pores were assumed to be either full or empty. Transmission factors for the sand/air/water interfaces were calculated using Fresnel's law. Absorption was neglected. The number of water-filled pores was then calculated by $I = I_{\text{dry}}(T_2/T_1)^{2p}$, where p is the number of water-filled pores, $T_1 = 0.946$ the transmission coefficient for sand-air and air-sand, and $T_2 = 0.991$ the transmission coefficient for sand-water and water-sand. For the number of pores to water content relation, a calibration measurement with gravimetric water content measurements was done. The method was extended by Glass et al. (1989) to a two-dimensional field by supplying one side of the soil slab with a diffuse light source. Images were taken with a standard video camera. Only relative water content measurements were done and the evaluation of Hoa (1981) was used without the calibration. Tidwell and Glass (1994) used the same technique with a CCD grey value camera as measurement device and an improved evaluation. The measured intensity was corrected for light source drift and normalised with the standard two-point method, $I' = \frac{I - I_{\text{dry}}}{I_{\text{wet}} - I_{\text{dry}}}$, to correct inhomogeneities of the light source. They abandoned the calibration measurement of Hoa (1981) and used the simple model $p = k\theta$ to calculate the number of water filled pores, where k is the average number of pores in direction of the light beam and was determined empirically. Tidwell and Glass (1994) compared this measurement technique with X-ray measurements and found the data to agree within 5% saturation. Niemet and Selker (2001) developed several intensity-saturation relations, based on different linear box models and one based on a pore model. Niemet et al. (2002) also included liquid films in their consideration.

However, all these authors assumed linear propagation of light and thus box-shaped sand, water, and pore configurations. Diffuse light sources have many incident angles and due to the irregular shaped grains and curved water-air interfaces, even light with a normal incidence angle is refracted and reflected into all directions. The transmitted light has been travelling on an irregular path with many refractions and reflections ("multiple scattering", figure 5.6 on page 46). This also leads to a significant broadening of an initial δ -peak light source, as schematically shown in figure 4.5 on page 34. This effect can be seen in figure 12.28 on page 113 (a) and (b) which shows measured images with a slit positioned between (a) the sample and the camera and (b) the light source and the sample. That broadening effect can be described mathematically by convolution with a point spread function and is explained further in sections 4.5 and 12.6. Rezanezhad et al. (2006) were the first to account for multiple scattering. They used an X-ray measurement to calibrate the intensity-saturation relation and found a non-linear function, thereby showing that the assumption of a linear relationship is invalid. As described in section 5.1, X-rays do not exhibit multiple scattering in the porous medium.

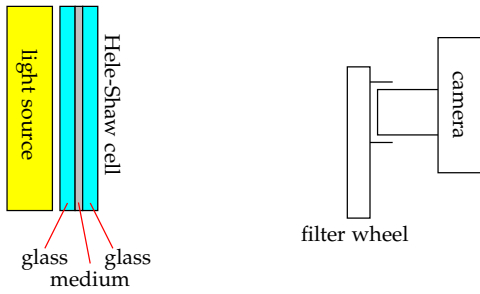


Figure 12.5: Sketch of the setup for light transmission experiments. A homogeneous light source illuminates the sample. Images are taken with a digital camera. Optionally, a filter wheel is used to select appropriate wavelengths.

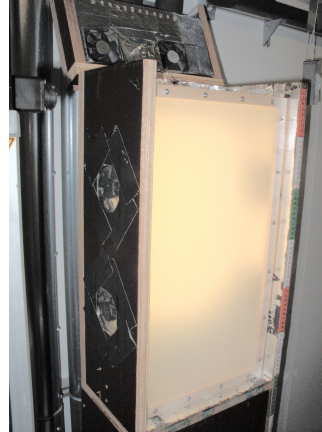
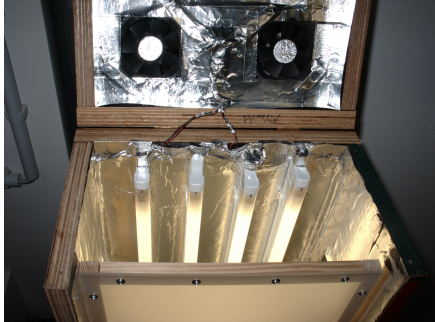


Figure 12.6: Visible light source used for water-content light transmission measurements.

For the investigations presented in this thesis, the calibration idea of [Rezanezhad et al. \(2006\)](#) was used. The measurement procedure was improved and adopted to the light transmission evaluation as described in section 12.7.

For more precise measurements, near infrared (NIR) imaging spectroscopy has been developed. With this method, absorption bands of water are used to quantify the amount of water, more precisely the absorption length, as described in section 5.3.1. In addition to normal light transmission, the wavelength of the transmitted light can be selected. Spectral information is then used to directly extract the absorption length. Even at one single absorbing wavelength, the water can be seen much clearer than with the traditional method. A drawback of the method is that available cameras in the near infrared are expensive and nevertheless have a relatively low resolution (standard is 320×256 pixels).

12.3.2 Experimental setup

The setup used for light transmission experiments is sketched in figure 12.5. The soil is contained in a Hele-Shaw cell as shown in figure 12.10 and illuminated with a homogeneous light source. A digital camera is used to capture images of the transmitted light. Optionally, a filter wheel is used to select appropriate wavelengths with a optical bandpass filters.

VIS light source

The visible light source (figure 12.6) consists of a wooden box with four fluorescent lamps (Osram 75W White) at the rear. Fluorescent lamps have the advantage of limited heat production compared to thermal emitters. The front side consists of a diffusion foil to homogenise the light before it enters the Hele-Shaw cell. The other walls of the box are covered with a reflective aluminium tape to reflect most of the light into the forward direction. For cooling, six fans are installed inside the box which blow fresh air inside the box from the sides and out again on the top. This avoids heating of the medium.

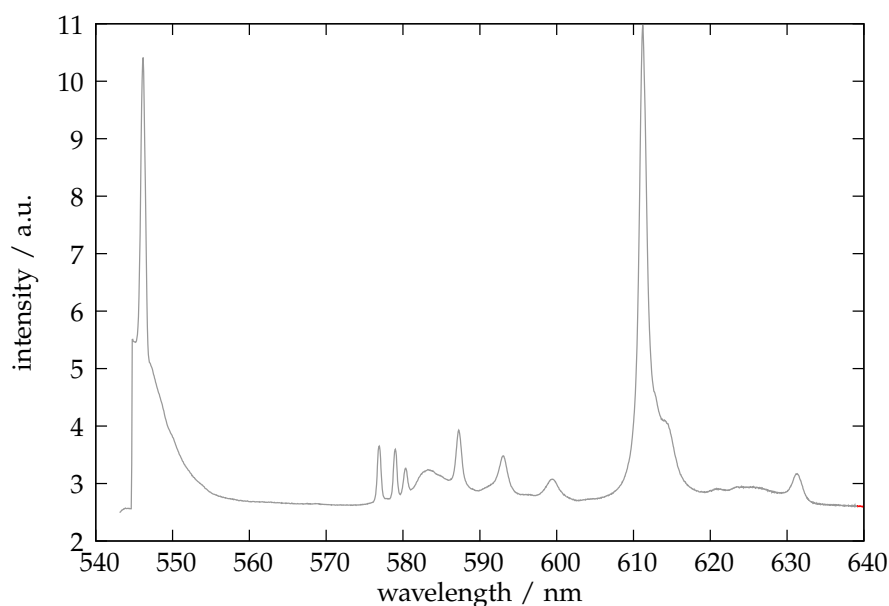


Figure 12.7: Spectral emissivity of the visible light source.

The spectral emissivity of the light source is shown in figure 12.7. The measurement was made with an Ocean Optics USB2000 spectrometer with a wavelength resolution of 0.6 nm. Emission lines of the gases in the lamp can be seen clearly, with main components at 546 nm (green) and 611 nm (red).

Drift test: To estimate the stability of the light source, a test experiment was conducted. After powering on the light source, images were taken every 10 min. Images were dark corrected and the grey value variance was estimated using the calibration measurement. Each channel of the Bayer pattern (see section 12.4) was then multiplied with its corresponding Bayer multiplier. The actual light source was cropped from the image. Images at times t_i were divided by the first image at $t = 0$ to estimate the change of illumination ζ . The mean of all pixels is a measure for the change of the average intensity compared to the first image, while the variance is a measure for the inhomogeneity. Figure 12.8 shows the temporal evolution. After a first intensity increase of about a factor of 2 in the first 20 min after power on the fluctuations remain in a relatively narrow but still significant band. The initial intensity increase is attributed to the equilibration of the lamps. Hence, it was always waited for at least a few hours for the light source to stabilise. Remaining fluctuations are corrected. Spatial intensity changes (inhomogeneity) are relatively small. These cannot be corrected during experiments and introduce a systematical error. The variance seems to be anti-correlated with the mean intensity. Looking at the images of ζ shows that the changes are larger in the central region with high intensity.

NIR/VIS light source

Measuring longer wavelengths where water has significant absorption requires an appropriate light source. Thermal emitters were chosen, since they emit in all required wavelengths. 28 halogen light bulbs with a power of 50 W (Osram HaloStar 64440 ES) were installed in a 4×7 grid. The corresponding spectrum is shown in figure 12.9. As expected it shows the shape of a thermal emitter. Irradiance in the NIR range is reasonably high, even at 1450 nm. A diffusion foil (Plexiglas Satin Ice transparent 0D010 DF) was used to homogenise the irradiance. Since the emitted irradiance of the lamps sensitively depends on the current (the relative uncertainty of the emitted light irradiance is about 6 to 10 times the relative uncertainty of the current regulation), all bulbs were operated in series. This ensured that all lamps exhibit the same

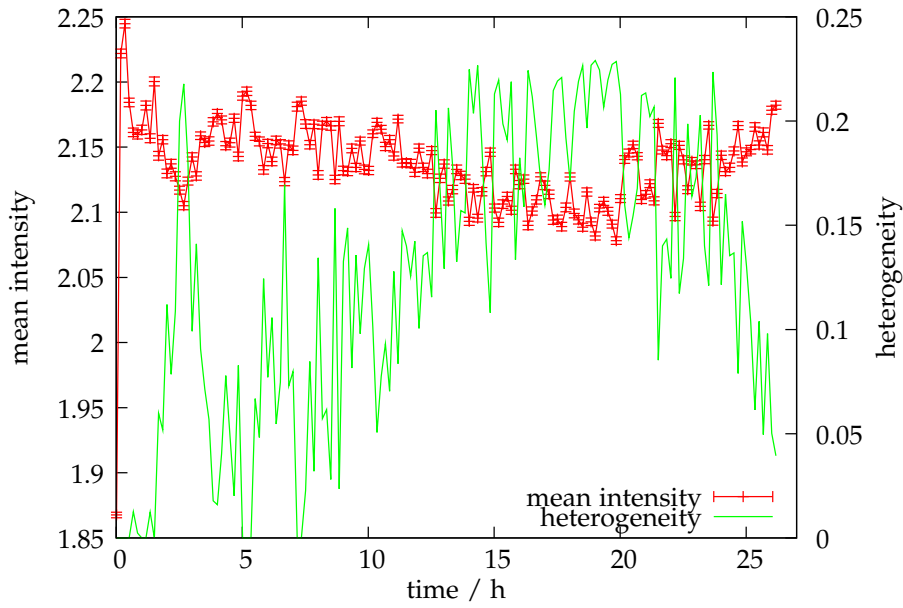


Figure 12.8: Temporal evolution of the mean visual light source intensity and inhomogeneity.

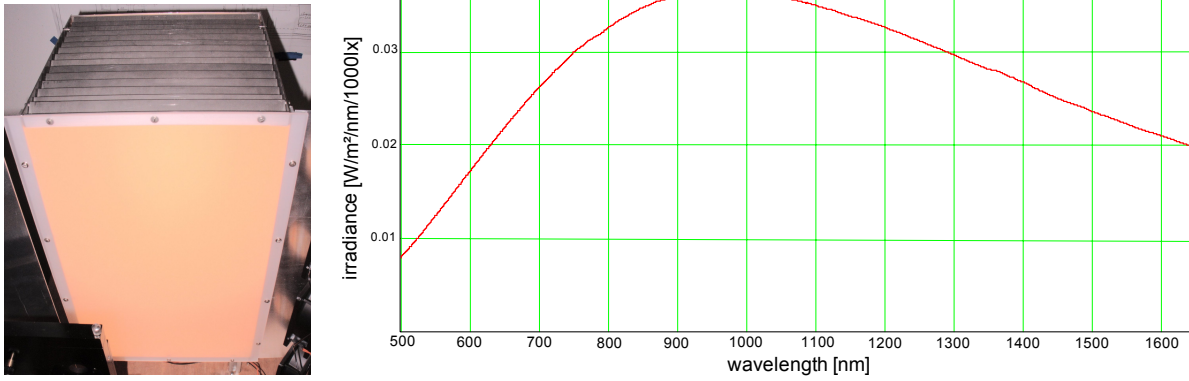


Figure 12.9: Picture of the NIR light source (left) and spectral emissivity of the Osram HaloStar 64440S bulb at a current of 4.35 A and an irradiance of 345 lx (right). *Emissivity spectrogramm was provided by Osram GmbH, München.*

change in intensity. One regulated high-voltage current source (Delta Electronics SM 400-AR-8) was used. The temporal stability was specified as 9×10^{-5} , line regulation as 0.2 mA (0.007 %). The lamps were operated at 3.5 A. Since brand-new bulbs settle in the first 50 h of operation (during that phase, the irradiance at constant current increases), the light source was operated for about 55 h before it was actually used in the experiments.

Hele-Shaw cell

To ensure transparency of the medium, the thickness of the soil to be investigated is limited. Therefore, a semi-two-dimensional sample in a Hele-Shaw cell (figure 12.10) was used. The Hele-Shaw cell consists of two parallel glass plates (500 mm \times 300 mm) positioned at a distance of 3 mm with rubber gaskets. These gaskets also served as sealant. The glass plates were mounted in a steel frame. The medium is then filled into the space between the glass plates. At the bottom, the medium was separated by a porous separator (sintered glass, mean pore size 120 μ m). At the right and left borders, two windows were separated with additional gaskets.

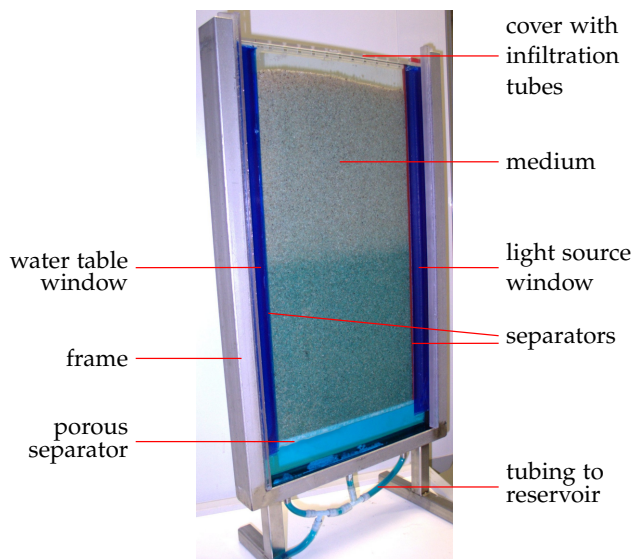


Figure 12.10: Hele-Shaw cell used for light-transmission experiments. In this photo, water with Brilliant Blue has been used for visualisation purposes.

The one on the right was sealed completely and served as a light source reference, while the one on the left was open at the bottom such that free water entered the window to mark the water table height. Blue absorber foils were mounted in front of these two windows to reduce the light intensity. They were necessary to prevent over-exposing the corresponding pixels in the digital image, since the porous medium weakened the light intensity by about a factor of 1000. A lid covered the sample on the top to prevent evaporation and simultaneously served as mount for infiltration tubes.

VIS Camera

A Canon EOS 300D, a consumer digital single-lens reflex camera with a 6.3 MPixel 12 bit CMOS sensor, was used for macroscopic water content measurements. All automatic functions were disabled and images were saved as raw data. Image processing was then used to calculate the corresponding quantities of interest (section 12.3.4). Further information about the consumer camera is reported in section 12.4 and calibration measurements in section 12.5.

NIR camera

A Xenics Xeva, an industrial/scientific NIR camera with a 320×256 pixel 12 bit InGaAs sensor with $30 \mu\text{m}$ pixel pitch and a wavelength range of 900 nm–1700 nm was used for detailed microscopic measurements. The camera allows adjustment of the ADC central and window width voltages, to optimise the quantisation for the actual measurement. Possible frame acquisition frequencies are up to 100 Hz via USB. A thermo-electric cooler allows controlling the sensor chip temperature. The camera has a low gain mode, which uses a large integration capacitor in the trans-impedance amplifier circuit to reduce overall system noise. However longer integration times are needed. Enabling low gain mode enlarges the sensitivity by a factor of 20, but also reduces the output conversion factor by a factor of 20.

During measurements, images were acquired and saved as-is without any preprocessing. Image processing was then used to calculate the corresponding quantities of interest (section 12.3.4). The calibration measurement is reported in section 12.5.2.

Filter wheel and optical bandpass filters

A filter wheel (Optec IFW) was used to select appropriate wavelengths for NIR measurements. The wheel can hold 8 standard 25 mm filters and automatically senses the correct wheel position by using small magnets and hall-effect sensors. The position can be computer-controlled via a serial RS232 connection, which allows automated measurements.

Standard optical bandpass filters (Andover Corporation) were used to select appropriate wavelengths. Wavelengths and bandwidth (full width half maximum) are given in table 12.1. The effective absorption is given by

$$I(\lambda) = \int I_0(\lambda) \exp(-\sigma(\lambda)\rho_w d_w) d\lambda$$

$$\approx I_0 \exp(-\rho_w d_w \underbrace{\ln \int T(\lambda) \exp \sigma(\lambda) d\lambda}_{=:\sigma_{\text{eff}}}) , \quad (12.1)$$

where T is the transmissivity of the optical bandpass filter. Since T is non-zero only in a very small region (typical full width half maximum is 25 nm), the emissivity of the thermal emitter I_0 and the absorption length d_w are approximately constant in that range, thus these can be taken out of the integral. The error introduced by this approximations is considered neglectable. Thereby one obtains an effective cross-section σ_{eff} , which can be used for determining the absorption length.

12.3.3 NIR imaging spectroscopy wavelength selection and test

As explained in section 5.3.1, absorption bands of water can be used to calculate the absorption length (equation (5.21)) if spectral information is exploited. To explore which wavelengths are best suited for water measurements and to test the measurement principle, test measurements were made in a Fourier Transform Infrared Spectrometer (FTIR) (Bruker Vertex80). The Hele-Shaw cell was put into the FTIR at the point of focus (diameter of the ray was about 1 mm) and measurements of dry, totally saturated, and partially saturated sand were made. Aperture was 4 mm and 25 scans were averaged. The transmission coefficient in respect to the dry sample was calculated by dividing the water measurement data by the dry measurement data.

Figure 12.11 shows the wavelength-dependent transmission coefficient. As expected, the transmitted intensity increases at wavelengths with neglectable absorption. At other regions, the intensity increase is masked or even over-compensated by the decrease of light due to absorption. Comparing with the absorption coefficient of water (figure 5.4 on page 45) shows a good qualitative agreement of the shape of the absorption. A wavelength with as little absorption as possible must be chosen as reference to optimise the absorption differences. Using different absorbing wavelengths allows to change the sensitivity, since a higher absorption coefficient difference has a higher sensitivity but may become saturated if the absorption path is too long. For the measurements, the 1075 nm absorption minimum was used as reference, since it is not too far from the absorbing wavelengths (remember that wavelengths should be close enough that the real part of the refraction index does not change significantly). The following absorbing wavelengths were chosen: (i) the 1455 nm maximum, (ii) 1505 nm which has about half the absorption coefficient of 1455 nm, and (iii) 1555 nm with lower absorption for long absorption paths. These wavelengths are available as standard bandpass filter for imaging measurements. Table 12.1 shows the corresponding absorption coefficients as interpolated from the data of Palmer and Williams (1974) and the bandwidth of the optical filters used in the experiments for wavelength selection.

The absorption of pure water was also measured in the test but the measurements of Palmer and Williams (1974) are of much higher quality and therefore were used further on.

To check the wavelength's dependence on the results, the absorption length was calculated according to equation (5.21) on page 48 from the test measurement using 1075 nm as reference wavelength. The result is shown in figure 12.12. Data is only shown for wavelengths $\lambda > 1200$ nm, since below the absorption coefficient difference is too low which leads to large errors. The range of values of the absorption length is reasonable for a medium of 3 mm thickness and

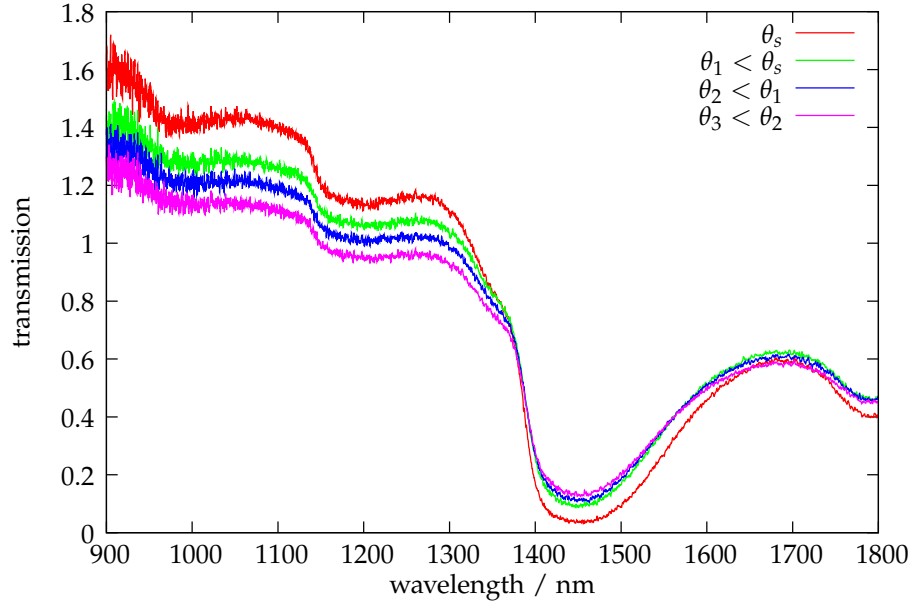


Figure 12.11: Spectroscopic measurement of the light transmission for different water contents. Due to lower light source intensity, noise is higher for lower wavelengths.

Table 12.1: Absorption coefficients at the chosen wavelengths. Values were interpolated using the data of [Palmer and Williams \(1974\)](#). The bandwidth (full width half maximum) of the corresponding optical bandpass filters are also shown.

wavelength / nm /	absorption / cm^{-1}	filter bandwidth / nm
1075	0.15	25
1455	31.4	25
1505	18.1	20
1555	10.4	30

a porosity of about 0.4, at higher wavelengths the values seem a bit low. Ideally without optical dispersion the path length should be constant throughout all wavelengths. The low values at higher wavelengths are attributed to the change of the real part of refractive index (see figure 5.5 on page 46), which leads to a change in multiple scattering, while the “dips” are assumed to be caused by the shift of absorption wavelengths with respect to the reference measurement due to the bound state of the water. However, these effects only affect the absolute value of the absorption length, but not the relative changes due to changing water content at one particular wavelength. Also, since the relative change due to water content changes is identical for different wavelengths, this measurement can be used to correct for these effects.

12.3.4 Image processing

Image processing was done with C++ code, using single precision floating point numbers to eliminate numerical errors caused by integer arithmetic. Single precision was chosen as an optimum between data size and precision (the raw sensor data were 12 bit integers). As first step, the raw measurement data was dark corrected to remove fixed pattern noise, where underflows (when the dark current was higher than the value due to statistical fluctuations) were explicitly avoided by setting these values to zero. Pixel value errors were estimated. Further on, errors were propagated using Gaussian error propagation. The further processing described below

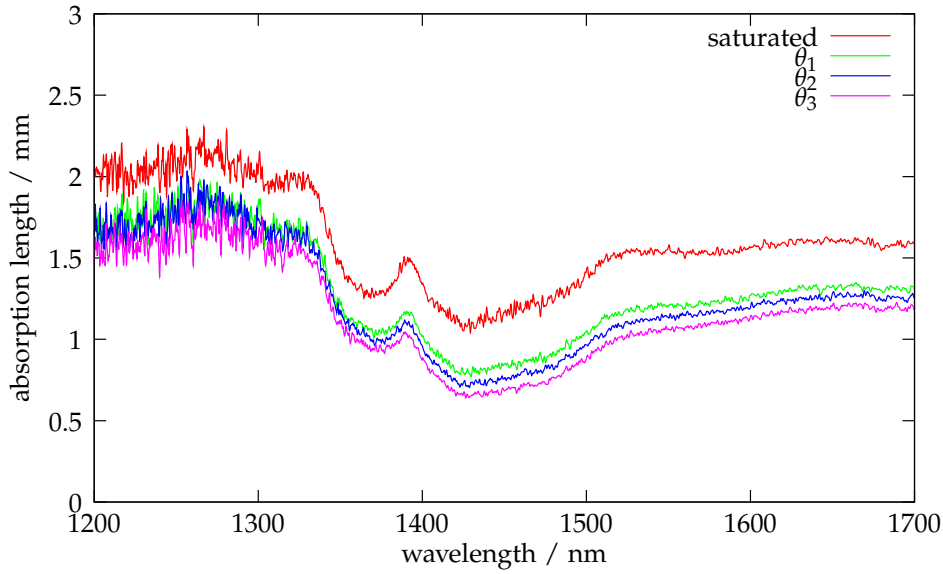


Figure 12.12: Absorption path length as estimated from the test measurement using 1075 nm as reference.

differed for different cameras and different measurements. Covariances between pixels which are introduced for example by filters were neglected.

Water content measurement using a consumer camera

For every measurement i at time t_i , one image without wavelength selection was taken. Pixel errors were estimated according to the camera calibration measurement (section 12.5.1) assuming a linear variance model. Images were white-balanced and Bayer-downsampled as described in section 12.4 and the region with the medium was extracted. Light source fluctuations were corrected using the data from the light source reference window. A scalar correction factor was calculated by $\chi_i = \langle w_{\text{ref}}(x) / w_i(x) \rangle$, where w_{ref} and w_i are the light source window of the dry cell which served as reference, and of measurement i , respectively. $\langle \cdot \rangle$ denotes the average of all pixels. All pixels were multiplied by χ_i . A radius 1 median filter was applied to remove salt and pepper noise. If the measured and the dry and/or wet images did not overlay due to camera and/or cell movements, registration was used to correct it. This was done with Matlab using double precision floating point arithmetic. It was assumed that the movement could be represented by a projective transformation, which describes the changing projection by the camera. Cross-correlation was used to tune the transformation parameters. Water content was then estimated using equation (5.20) on page 48 by applying the inverse of F , $\Theta = F^{-1}(I)$. In the polynomial evaluation for F^{-1} , the error of the normalised intensity as well as the fit coefficient errors including covariances were propagated. If normalised intensities were below the validity range of the calibration, which corresponds to $\Theta = 0.1$, but larger than zero, Θ was estimated by linear interpolation between zero and the lower validity limit. This was done to avoid ambiguities. All these values thereby fall into a known saturation range and can easily be recognised when looking at the results. It is important to notice that values in the saturation range below 0.1 have a very high degree of uncertainty. Normalised intensities outside the range $[0,1]$ are outliers and were kept as-is. Afterwards, a median filter with radius 1 was applied to eliminate outliers. These were caused by imperfectly overlaying images, which is unavoidable since the registration is not perfect, and by changing grain structure. If the grain structure in two images does not match, artefacts in form of large intensity peaks are produced in the output image. Figure 12.13 shows the normalisation procedure graphically. The relation

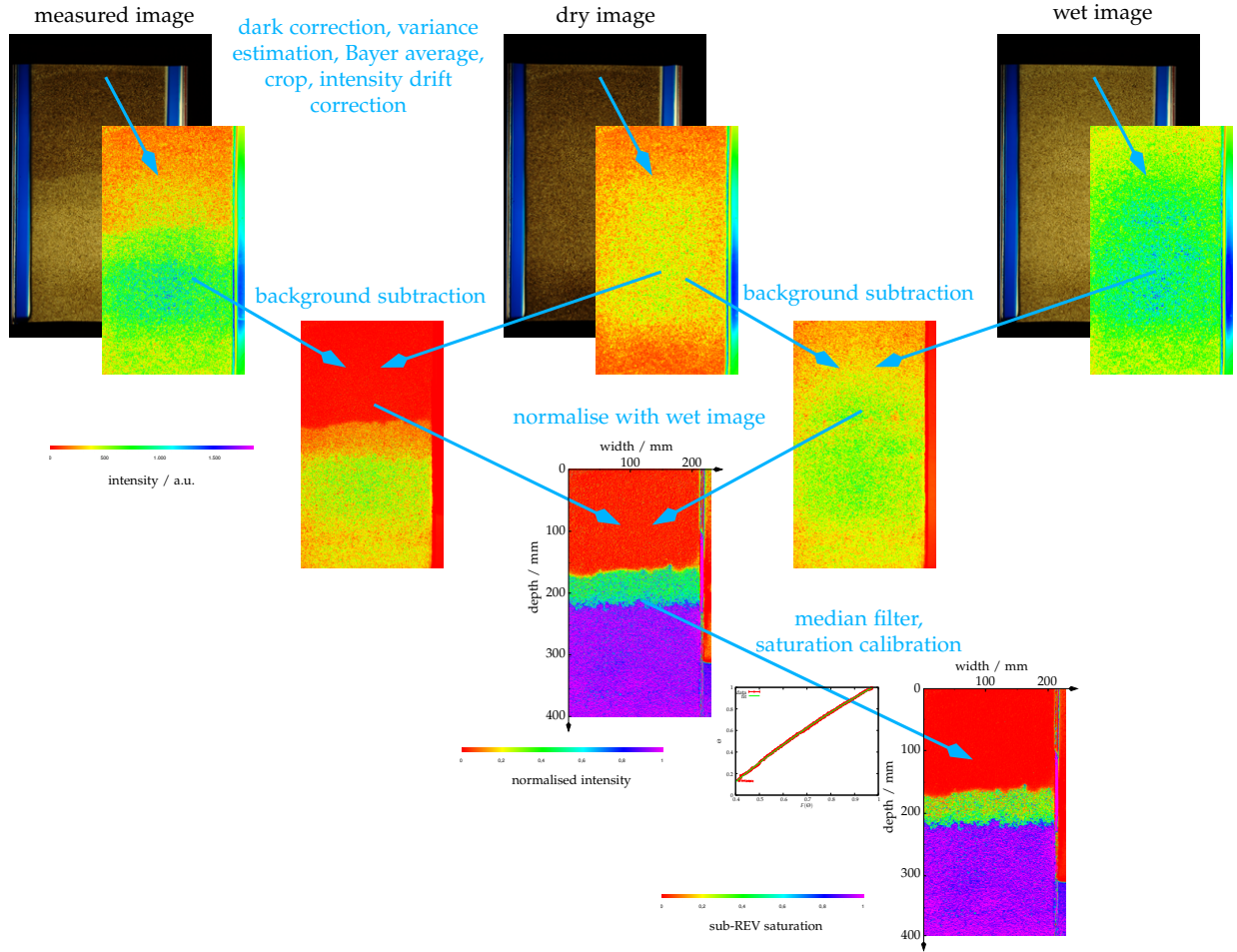


Figure 12.13: Image normalisation for water content estimation.

$F(\Theta)$ had previously been estimated using an X-ray calibration measurement (section 12.7.1).

The remaining non-correctable image dislocation due to camera and medium changes introduces a background noise that is not covered by the error propagation of uncertainties. It is difficult to quantify, since it overlaps with real information about heterogeneous water content changes, therefore it was not incorporated into the uncertainty of the values.

Water measurement using NIR camera (imaging spectroscopy)

The absorption length as a measure for water content was estimated using two different wavelengths. With the spectroscopic measurement, a much higher precision is possible than in normal LTM measurements. The non-linearity correction as determined from the calibration measurement (section 12.5.2) was applied. For every measurement time t_i , the absorption length was calculated by solving equation (5.21) for d_w ,

$$d_w(x) = \frac{1}{\rho_w \Delta \sigma_w} \alpha \log \left(\frac{I(x, \lambda_2) I_0(x, \lambda_1)}{I(x, \lambda_1) I_0(x, \lambda_2)} \right), \quad (12.2)$$

where $\rho_w \Delta \sigma_w$ is the difference in absorption coefficient between λ_1 and λ_2 and was determined from the measurement by [Palmer and Williams \(1974\)](#), and α is a correction factor which accounted for different exposure times for dry (I_0) and normal measurements. It was assumed that the gray value is linearly related to exposure time, which is true after the non-linearity cor-

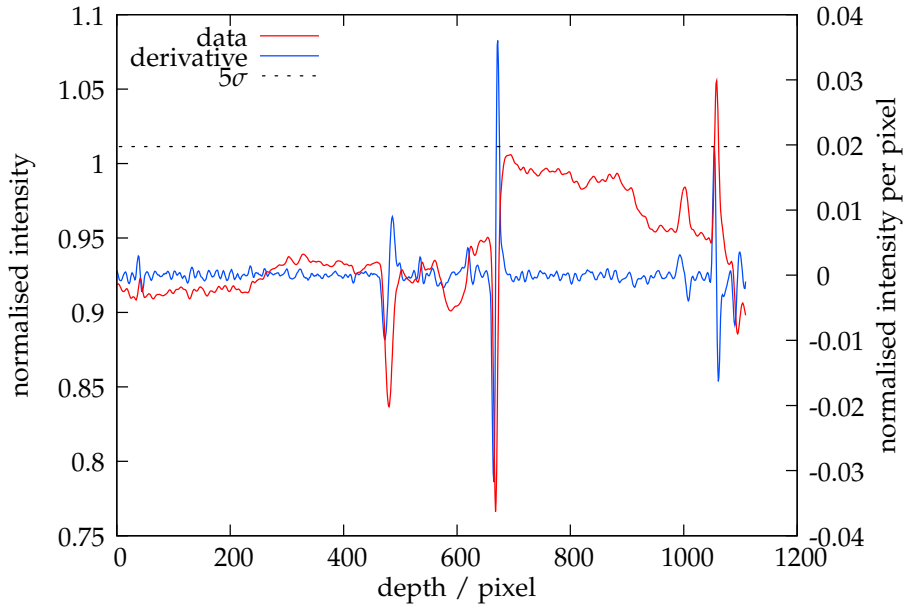


Figure 12.14: Detection of the water table with edge detection.

rection has been applied, thus α linearly scaled the gray value according to the corresponding exposure times as $\alpha = \frac{t_1 t_{0,2}}{t_2 t_{0,1}}$.

Switching between two filters needed about 3 s. If the system changes during this time, a systematical error is introduced because the evaluation assumes that images at both wavelengths show exactly the same state. The extent of that error depends strongly on the processes to be observed and cannot be quantified. This error will only be of considerable importance on very fast transitions, e.g. directly at the edges of rectangular forcing. During the rest of the time, processes are relatively slow and resolution is considered sufficient.

High-speed water measurements at one NIR wavelength

Since the time required for changing the optical filter limits the time resolution (switching between two filters needs about 3 s), measurements at one absorbing wavelength were made to look into fast processes. The camera allows framerates up to 100 Hz, depending on the integration time. Non-linearity correction was applied. Then the images at time steps t_i were divided by an image of the light source at the same wavelength and settings, which served as flat field correction.

Water table estimation

Water table heights were extracted from the water table reference window for every measurement. The meniscus of the free water table in the window can be seen as a dark line and the water-filled region has a higher intensity due to better refractive index matching of water-glass than water-air. To obtain the water table height, the window region was extracted from the image and Bayer-downsampled. For every height the image was averaged horizontally since the data is horizontally homogeneous. A Gaussian filter of radius 6 was applied to reduce random noise. Edge detection was then used to detect the water table. The profile was derivated using the standard symmetric filter. Mean and standard deviation were calculated, and a peak with at least $n\sigma$ was searched. The water table height was defined as the middle between the crossings, the error was estimated from the two crossings of the peak with the threshold. n was initially set to 5. If no peak could be detected, n was reduced gradually until the search was successful. Figure 12.14 illustrates the image processing.

To obtain the amplitude of the water table fluctuation, different strategies were used for sinusoidal and rectangular forcing. For the sinus, a sinus function $z = z_0 + A \sin(2\pi t/T)$ was fitted into the data and the amplitude A and zero position z_0 of the fit were then used. The adjusted function fitted well to the measured data for all experiments. For the rectangle, every half-cycle was identified using the measurement time and the corresponding values were averaged. The first 3 points were not taken into account, since due to finite conductivity, the water table in the cell did not reach the value instantaneously. The amplitude was then estimated as half the maximum-maximum difference and the zero point as the middle between the maxima.

Detection of the capillary fringe boundary

As mentioned in the introduction, a process-based definition of the upper boundary of the capillary fringe is the point where the air phase becomes continuous. However, due to measurement limitations, this criterion cannot be assessed experimentally, since the macroscopic water content measurement does not resolve microscopic features like air phase continuity. Therefore, a more practical approach must be used. Here, the upper boundary was defined by a threshold value of water saturation. The reason was that this is a relatively stable indicator for the edge of the capillary fringe, while edge detection was found to be unstable, even if the data was smoothed before edge detection was applied.

The goal was to estimate the mean position of the edge of the capillary fringe in the cell. Therefore, the data were averaged horizontally. Since the capillary fringe is heterogeneous, the microscopic water content values at different horizontal positions are not the same and the values cannot be treated as different realisations of the same random variable. Normal error propagation was used to estimate the error of the mean since different values are combined to obtain the mean.

Starting from the bottom, the data were searched for contiguous points where the error bars were touching the threshold value. A parabola was fitted into these points to check if it was a true intersection, or if the water content was going back again. If no intersection was found, the search was continued. Else, a line was fitted using weighted linear regression. It was tested if the slope was larger than $0.0005 \text{ pixel}^{-1}$, to reject situations where the parabola intersected but the points were very close to the apex and the water content was going down again nevertheless. If the test succeeded, the intersection point was calculated. Errors were estimated using data errors of the pixel closest to the intersection point and the fit errors of the line including covariances. If no intersection is found, the search is repeated and the rejection criteria are relaxed by also allowing points where the parabola does not intersect. This is required for a few special cases with large groups of intersecting points. The procedure is shown graphically in figure 12.15. The threshold was optimised to 0.8.

Ambiguities occur if several intersections with the threshold are found. One of them must be chosen arbitrarily. Here, the lowest was used. This leads to jumps in the solution if the data move into the threshold or out of it vertically. Figure 12.16 depicts such a situation. At the first timestep, one real intersection is present and it is correctly identified by the algorithm. In the next timestep, however, the water content decreases due to drying and a second intersection arises. Since it is at a lower position than the first one, it is chosen by the algorithm and the solution jumps. This effect cannot be fully eliminated and is a drawback of the particular definition of the capillary fringe boundary.

CF boundary half minimum-maximum distance

To estimate the half minimum-maximum distance η of the CF boundary for rectangular forcing, a line was fitted into the last points for each half-cycle and then extrapolated to the end of the half cycle. Even if the CF boundary was still raising, this strategy obtained the correct end

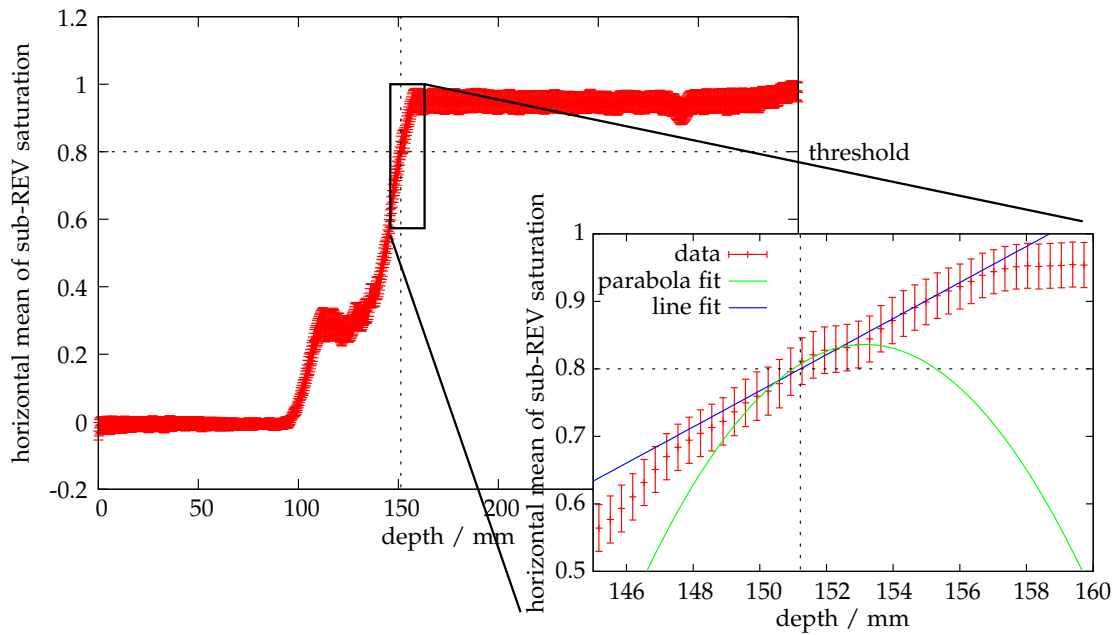


Figure 12.15: Determination of the upper boundary of the capillary fringe. A threshold value (here 0.8) was used to define the boundary. The data was searched for the points where the error bars touch the threshold. A parabola was fitted to check if it was a true intersection. If that test passed, a line was fitted and the intersection with the threshold was estimated.

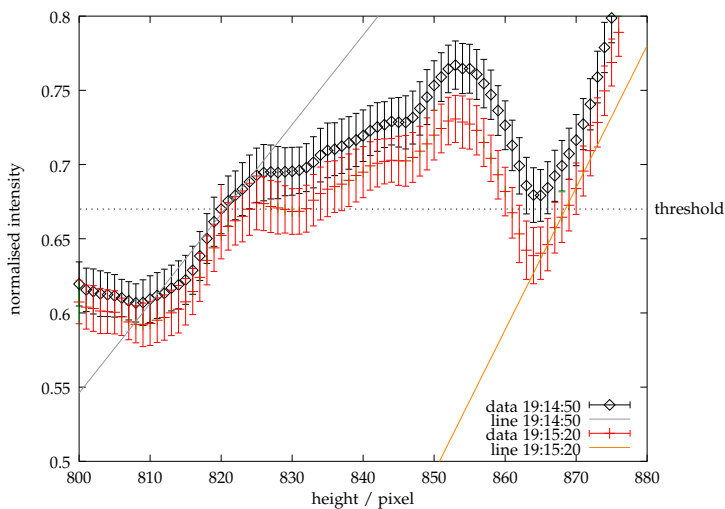


Figure 12.16: Ambiguity in the capillary fringe boundary definition. The data moves into the threshold from above and introduces two intersections. This introduces a jump in the edge estimation, since the lowest intersection jumps from the left point at 19:14:50 to the right point at 19:15:20.

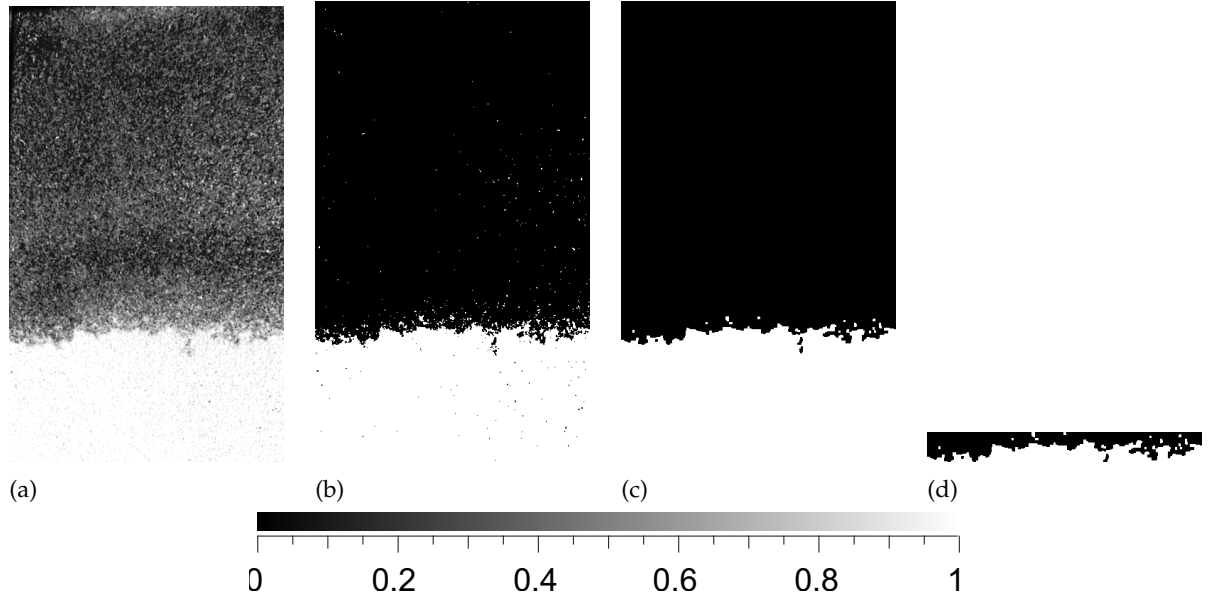


Figure 12.17: CF heterogeneity measure estimation: The saturation image (a) is binarised (b). Opening and closing is used to eliminate outliers (c). The CF border region is determined (d) and the absolute value of the area from the middle to the actual border is summed up.

point before the CF boundary was going down again. Half the distance between these end points was then defined as η . It was obtained separately for each cycle.

Heterogeneity measure of the capillary fringe boundary

To estimate the heterogeneity of the CF and the change of heterogeneity due to transient forcing, a heterogeneity measure H is introduced: the more fragmented the CF, the higher H . A prerequisite for the measure is that it should not depend on the actual discretisation of the data, i. e. if for example the discretisation is doubled but the structure of the border is not altered, the measure should be the same.

Saturation images were binarised with the same threshold as in the CF boundary estimation. Opening and closing (Jähne, 2005, sect. 18.4.1) with a square 3×3 mask are applied to the binarised image to eliminate single misclassified pixels. To find the edges of the CF border, starting from the bottom the first totally black line is searched, and from there the first totally white line below. The area in-between those two lines is defined as the continuous CF border region and it was cropped from the image. This definition for the border region was chosen to avoid black or white spots to spuriously enlarge the border region. For each horizontal pixel, starting vertically from the middle of the CF border region, the distance to the black/white edge is determined and the absolute value of the area defined by that distance and the pixel width as determined by the distance-pixel calibration is then added up and normalised by the image width w :

$$H = \frac{1}{w} \sum_{x=0}^{w-1} |y_{\text{middle}}(x) - y_{\text{edge}}(x)|. \quad (12.3)$$

Figure 12.17 illustrates the image processing.

If the CF border is homogeneous, i. e. it is a horizontal line, the thickness of the CF border region and therefore the measure is zero. The more fractured the border, the larger the area from the middle line to the actual border line and therefore the larger the heterogeneity measure.

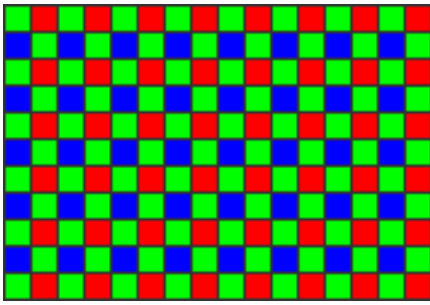


Figure 12.18: Bayer pattern used in consumer cameras to obtain colour information. 50 % of all pixels are green, 25 % red and 25 % blue.

The measure H weakly depends on the size of the opening/closing mask. The larger the mask size, the larger details are lost, but also larger outliers are eliminated and the border becomes smoother. An optimum mask size was found to be 3×3 .

Distance-pixel calibration

To obtain the pixel-distance relation, images with a measuring tape were made. The number of pixels for a certain distance was obtained in several image regions and the millimetre-pixel relation was calculated. To estimate the distance change over the image due to lens distortions, an image with a checkerboard marker put onto the cell was made.

12.4 Consumer DSLR cameras

Consumer digital single-lens reflex (DSLR) cameras with large resolution are commonly available for low prices (less than 1/10th of a professional scientific camera). DSLR cameras use a mechanical mirror system and a pentaprism to allow the photographer to directly view the picture through the optical viewfinder on the back of the camera. Most modern cameras have a 12bit CCD or CMOS sensor with several megapixels resolution and feature remote control via a PC which permits automated measurements. They offer an alternative for measurements on large timescales which do not require a high acquisition rate. Ideally, the raw format would allow to save the sensor-data and camera adjustments as-is with no preprocessing, which would permit the exact knowledge of all image processing steps. This is important for scientific measurements, because the dependence of the results on the measured irradiance must be known. In practise, internal preprocessing is done by most consumer cameras and no information about it is given away from the manufacturers. Here, calibration and test measurements are necessary to ensure quantitatively correct results.

Colour by Bayer pattern

Almost all consumer colour cameras have just one image sensor. To obtain colour information, the chip has colour-filter micro lenses in front of the sensor which form a colour matrix. The colour pattern used by nearly all cameras is called Bayer pattern¹ and contains 50 % green, 25 % red and 25 % blue pixels (figure 12.18), since the sensitivity of the eye is maximal for green light. Of course, as a result not all colour information is present in every pixel. Thus, if all colours are required at all pixels, the different colours must be *interpolated* to obtain a full resolution colour image, assuming that there are only small intensity differences between adjacent pixels of the same colour. The most simple interpolation just averages the missing values between adjacent pixels. However, this method blurs edges. Thus, more advanced methods interpolate alongside edges. Many camera manufacturers have their own proprietary

¹ Bryce E. Bayer submitted a patent in march 1975 on behalf of Eastman Kodak Company ("Color imaging array", United States Patent 3,971,065)

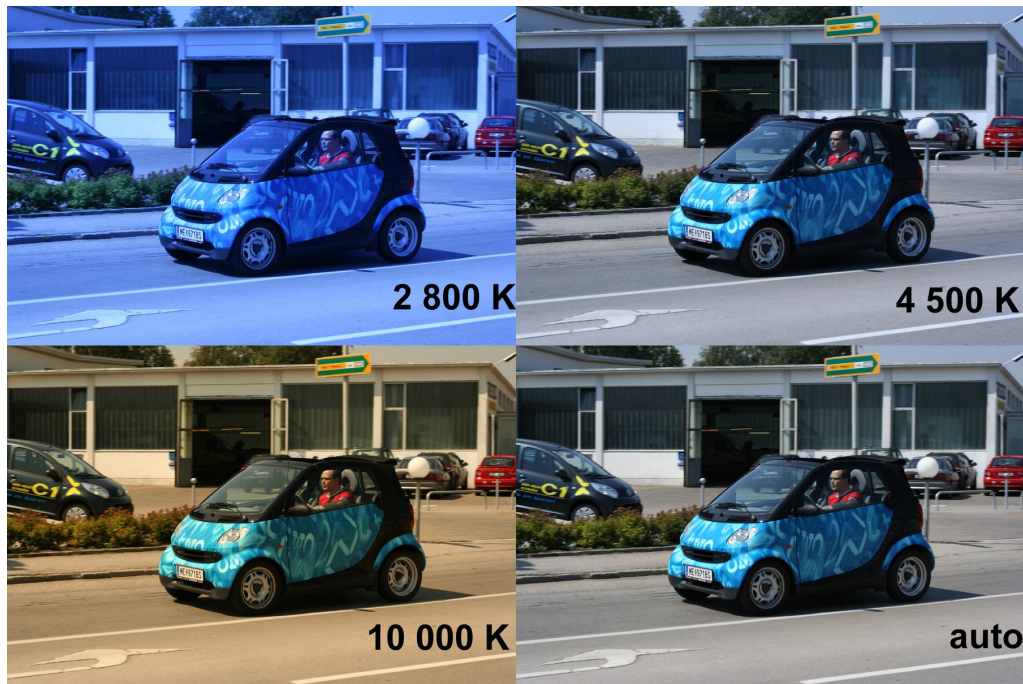


Figure 12.19: Effect of different white balance adjustments using different colour temperatures. *Image source: wikipedia.org*

algorithms. There are also quite a few open algorithms. One of the most popular ones is the adaptive homogeneity-directed (AHD) algorithm ([Hirakawa and Parks, 2005](#)).

White balance

Different light sources have a different spectrum, thus depending on the illumination, also coloured objects have a different spectrum. When looking with the human eye, the visual impression automatically adapts to changing illumination, such that objects keep their colour. This ability is called “chromatic adaption”. In contrast, for digital images this process must be done explicitly. Moreover the colour sensitivity of the image sensor is different from that of the human eye, and again different from the monitor, or printer, which is used to view the image, so additional corrections are needed.

The process of adjusting the intensity of the different colours to fit the “natural” appearance as perceived with the eye is called white balance, because it makes white objects really look white. Mathematically, constants m_i are multiplied to the intensity of each colour channel I_i (i denotes the colour, e. g. red, green, or blue). In normal automatic operation, the camera guesses the m_i according to sensor information by looking for the brightest area, which is assumed to be white, and forces it to be white by balancing the corresponding colour intensities. The m_i can also be determined manually by taking the image of a white object.

Assuming black body radiation, a temperature can be assigned to the illumination, and thereby to a specific colour balance of the captured objects. That temperature is called “colour temperature”. A low colour temperature has more red, a higher one more blue. Normal daylight has a colour temperature of about 5500 K. The effect of different white balance adjustments is shown in figure [12.19](#).

12.4.1 White balance and Bayer pattern averaging in LTM

With LTM, the attenuation of light in the medium is to be quantified. It is assumed that the light intensity change in respect to water content changes does not depend on the wavelength, because in the spectral range covered by a colour DSLR camera, there are no absorption bands of water and the real part of the refractive index of water is approximately constant in respect to wavelength. If every pixel was processed individually to obtain the attenuation, no white balance would be needed. However, this would forbid the usage of image processing filters for image correction prior to the attenuation calculation, because adjacent pixels are illuminated differently due to the colour filters (see figure 12.20 which shows a histogram of an image of the LTM light source) and image processing filters combine different pixels. In addition, adjacent pixels would have different sensitivities, because the quantisation intervals are constant and the number of photons is different. Thus adjacent pixels could not be compared directly. When applying the white balance, poorly illuminated pixels are multiplied by high factors and thus decrease the resolution. The relative variance stays constant since the variance linearly depends on the grey value. However the relative error (which is proportional to the square root of the variance) is lower for higher illuminated pixels. Equal illumination of all colours is not possible because there is no appropriate light source.

To eliminate the need for interpolation between different colour channels and to reduce noise, the four Bayer pixels were averaged. It can be thought of as putting all detected photons from the four different Bayer pixels, weighted by the spectral sensitivity, into one bin, $\mu = \sum_{i=1}^4 m_i g_i$, where g_i is the grey value of pixel i , and $\sigma_\mu^2 = \sum_{i=1}^4 m_i^2 \sigma_i^2$ since the m_i are constants. The result is then divided by 4 to keep the (anyway arbitrarily scaled) grey value number in the same range as the original values. Remember that for dynamic processes, temporal averages are not possible due to the low acquisition rate of consumer cameras.

The m_i are determined using the light source image. To have equal intensities with equal attenuation, the light source must be forced to be white. Therefore, a Gaussian is fitted into the histogram of each colour channel (figure 12.20). The m_i are calculated such that the centres of the Gaussians μ_i of all colours are shifted to the one with the highest grey value, $m_i = \mu_{\max} / \mu_i$. Thus, $m_i \geq 1$. This is a usual convention, because for integer arithmetic as used for consumer images, a multiplier $m_i < 1$ would decrease the number of used quantisation elements and thus the resolution. Although this is not the case here since floating point numbers are used for the LTM evaluation, the convention was used nevertheless.

12.5 Camera calibration

To estimate the error of the measured pixel values according to the sensor noise model (see section 4.4 on page 32) and to check the linearity and reproducibility of the camera, a calibration is needed. The camera is illuminated with a homogeneous light source with different known irradiances, and the measured pixel values are analysed.

12.5.1 Calibration of the consumer camera

The Heidelberg Collaboratory for Image Processing (HCI) has a calibration facility for digital cameras. Here, the camera without lens is directly mounted on a homogeneous LED light source with Ulbricht sphere. An Ulbricht sphere is a sphere with a small hole which is coated with a highly reflective material in the inside. It is illuminated from the inside. The light which comes out of the aperture has been scattered many times in the sphere, thus the source has a very flat irradiance field. The lens is removed to prevent distortions. The light intensity can be controlled by adjusting the LED current. A total illumination measurement can be done with

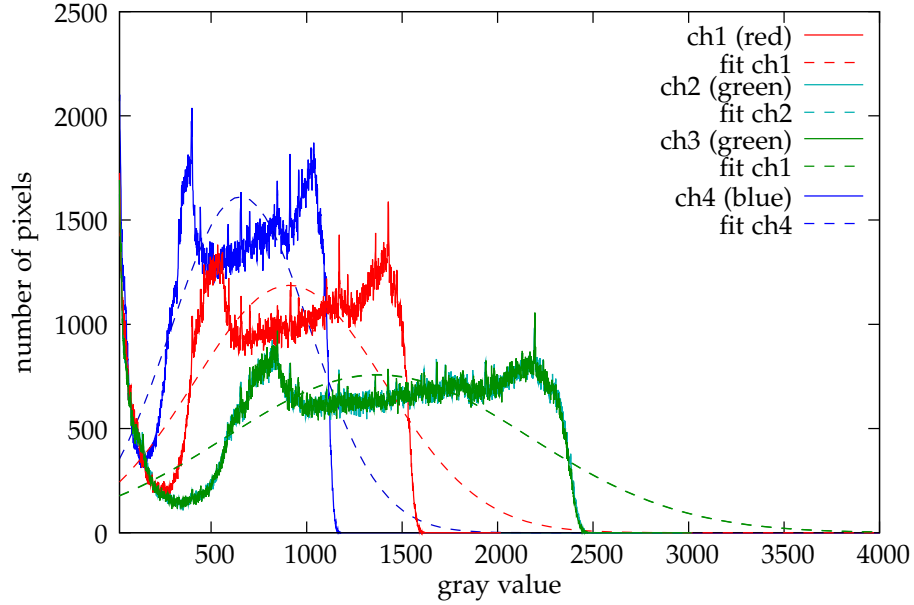


Figure 12.20: Histogram of raw pixel values of the VIS lightsource taken with the Canon EOS 300D, separately plotted for the four Bayer entries. For white balancing, the light source is made white by shifting the histograms of all colours such that they overlay each other. A Gaussian was fitted to each colour channel, and the centre values were used to calculate the multipliers m_i to 1.59 ± 0.02 , 1.00 ± 0.01 , 1.00 ± 0.01 , and 2.18 ± 0.03 . Note that green has two overlaying entries.

a calibrated photo diode. Statistics were collected spatially, assuming that all pixels are equal and the ensemble reflects also the statistics of a single pixel (ergodicity hypothesis).

The light intensity was increased in small steps until sensor saturation and then decreased again to preclude drift. For each measured light intensity, two images were made. Images with overexposed pixels (saturated grey value) were discarded by visually inspecting the histograms. Dark current was subtracted and the images were then separated into four sub-images, one for every Bayer channel. For every light intensity i and every Bayer channel j , the mean μ was calculated using the combined data of the two available images I_{ijk} , $k \in \{1, 2\}$, $\mu = \langle (I_{ij1} + I_{ij2})/2 \rangle$. For the variance, both images were subtracted and the variance was calculated, which results in twice the variance of a single image, thus

$$\sigma^2 = 1/2 \text{var}(I_{ij1} - I_{ij2}) = 1/2 \langle (I_{ij1} - I_{ij2})^2 \rangle, \quad (12.4)$$

since $\langle I_{ij1} - I_{ij2} \rangle = 0$. This variance calculation has the advantage that inherent structure of the images, e. g. due to different pixel sensitivities, is neglected.

An example of a dark-corrected calibration image is shown in figure 12.21. The remaining structure is attributed to different sensitivities of the different pixels. Assuming that the sensor is linear, i. e. the grey value is $g_i = a_i N + b_i$, where N is the number of photons in each pixel, different sensitivities mean that every pixel has slightly different a_i and b_i . The b_i are eliminated with the dark current subtraction. The linearity check is not disturbed by different a_i , since mean is a linear operation. The systematic error in the variance is kept minimal by determining the variance of a difference image as given in equation (12.4). The remaining systematic error is due to the fact that the variance is grey value dependent and different sensitivities lead to different grey values and thus pixel variances in one image. This effect is small (a grey value difference of 200 leads to a maximal variance variation of 18) and can be neglected.

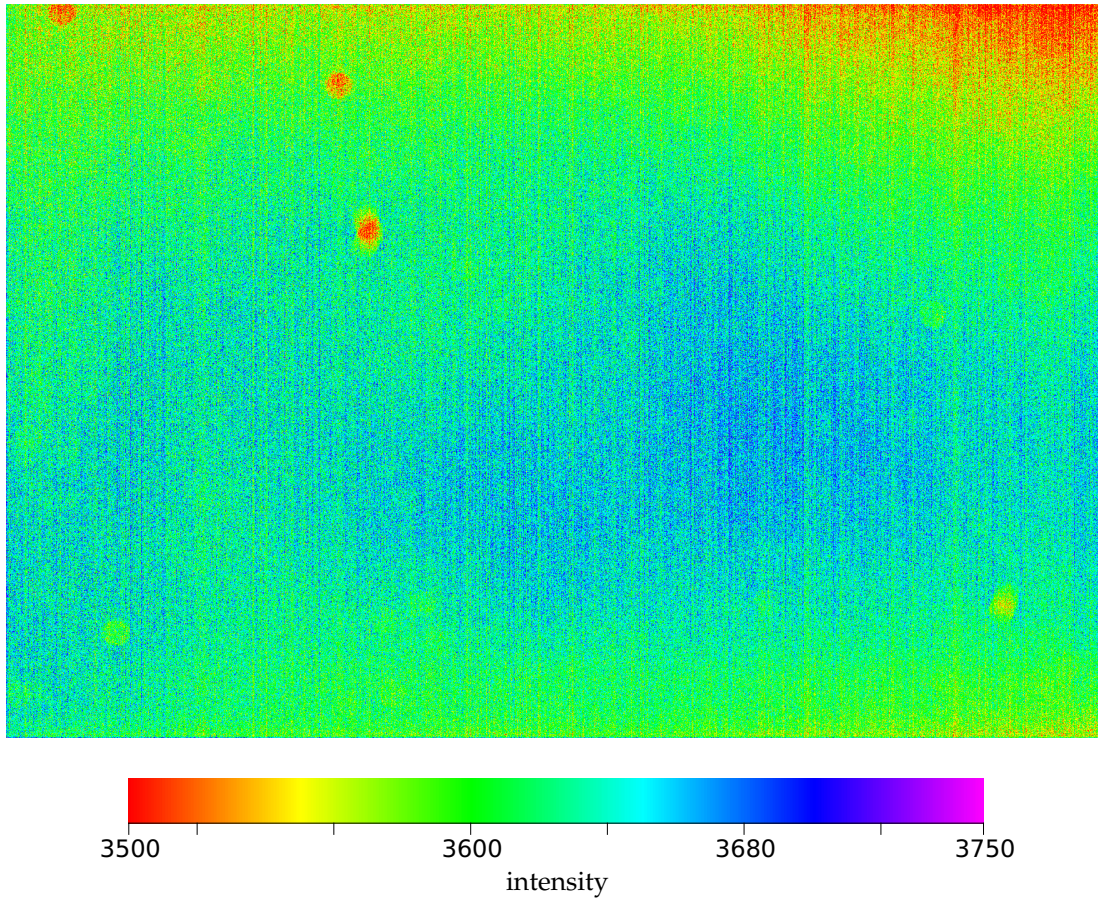


Figure 12.21: Dark-corrected Canon EOS 300D image of the homogeneous LED illumination of the HCI calibration facility (blue channel, blue LEDs, $N = 1.6061 \times 10^2$ photons/ $(\mu\text{m}^2\text{ms})$). The remaining structure is caused by the different sensitivities of the CMOS sensor.

The measurement was done with the smallest available ISO value, ISO 100, and an exposure time of $t_e = 1/80\text{s}$. The linearity test for all channels is shown in figure 12.22. Linearity is reasonable for all channels. A small curvature can still be seen, especially for larger grey values, which is attributed to camera-internal processing before saving as raw.

The fits for the noise model are shown in figure 12.23. The variance is reasonably linear for grey values below about 2700, but substantially deviates for larger grey values. The variance decrease for large grey values is attributed to camera-internal noise reduction. This is also suggested by the fact that saturated images never have the full 12 bit value, but saturation is already achieved at about 3990, and by the systematic deviation from sensor linearity for large grey values mentioned above. As there is no information available about internal preprocessing, this assumption cannot be verified, but it seems reasonable. For the linear fit, larger grey values were neglected. Using the linear fit for error estimation leads to a slight error overestimation of at worst a factor $\sqrt{2}$ for large grey values. Since very large grey values are seldom and the variance values are only used for error estimation, i. e. to see if changes are significant, this seems acceptable.

The results of the fits are given in table 12.2. Given errors are fit errors. Comparing the fit coefficients for the different channels shows that they are, within the uncertainty, nearly equal. Therefore one regression line can be used for all variances. Since the sensor elements are assumed to be equal and the difference between channels is only the microfilter in front,

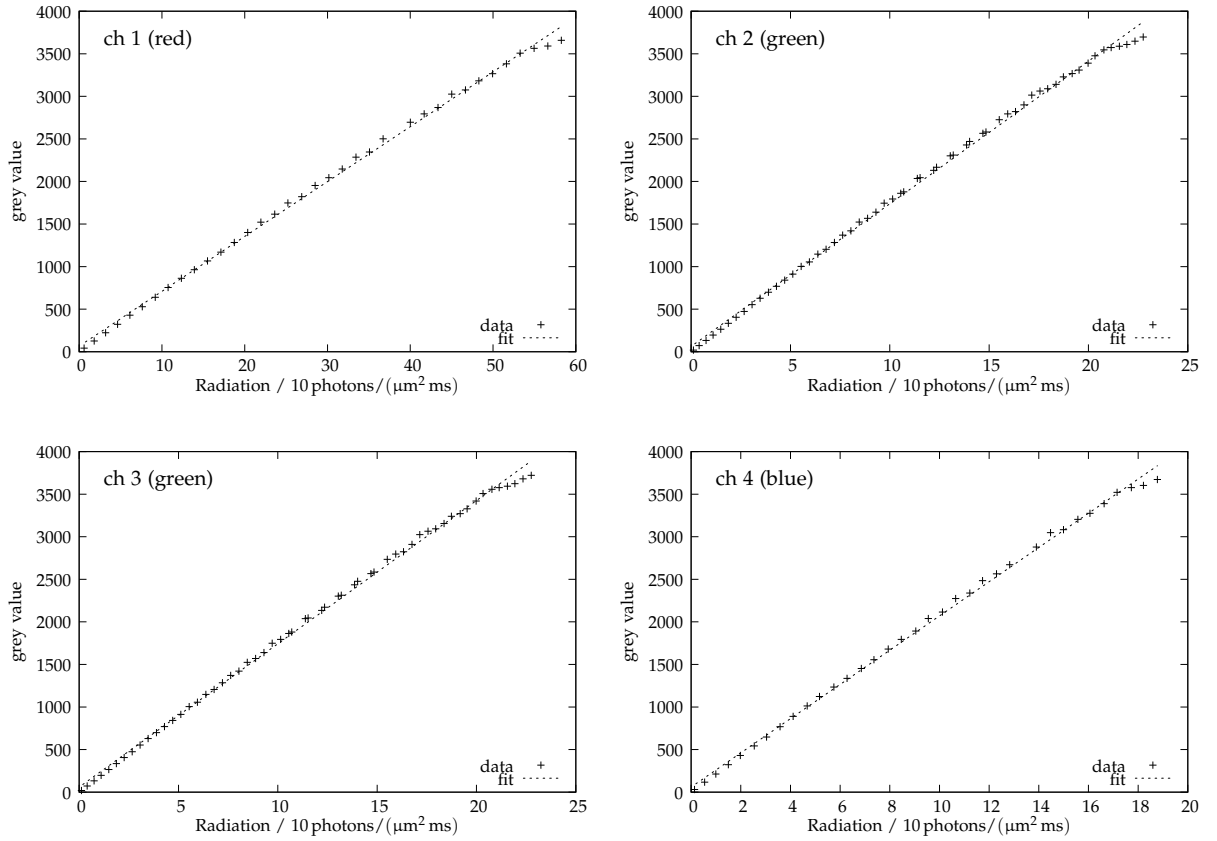


Figure 12.22: Linearity test for Canon EOS 300D with $t_e = 1/80$ s and ISO 100.

Table 12.2: Calibration results for Canon EOS 300D with $t_e = 1/80$ s and ISO 100. Co-variances between slope and offset are -0.85 to -0.86 for all fits. Given errors are fit errors.

channel	linearity fit		variance fit	
	slope	offset	slope	offset
1 (red)	0.168 ± 0.001	59 ± 13	0.089 ± 0.001	14 ± 2
2 (green)	0.168 ± 0.001	63 ± 14	0.092 ± 0.001	10 ± 2
3 (green)	0.168 ± 0.001	59 ± 13	0.089 ± 0.001	14 ± 2
4 (blue)	0.201 ± 0.001	59 ± 16	0.090 ± 0.002	12 ± 3
mean	–		0.090 ± 0.001	13 ± 2

this is expected. Once the number of photons necessary for a given grey value passed the colour filters, the colour channel makes no difference for sensor noise. Thus, the mean variance fit is used for error estimation in measurements.

Temporal pixel statistics

To assess the temporal fluctuations of the pixels, the camera was positioned in front of an Ulbricht sphere illuminated with with four halogen lamps. The lamps were operated with a stabilised power supply to prevent power fluctuations. Therefore, the camera can be assumed to perceive a locally and temporally constant homogeneous irradiance. N images with exactly the same settings were taken and statistics were analysed for every pixel. As the shutter is electronically controlled, the shutter timer error is expected to be constant for all exposure

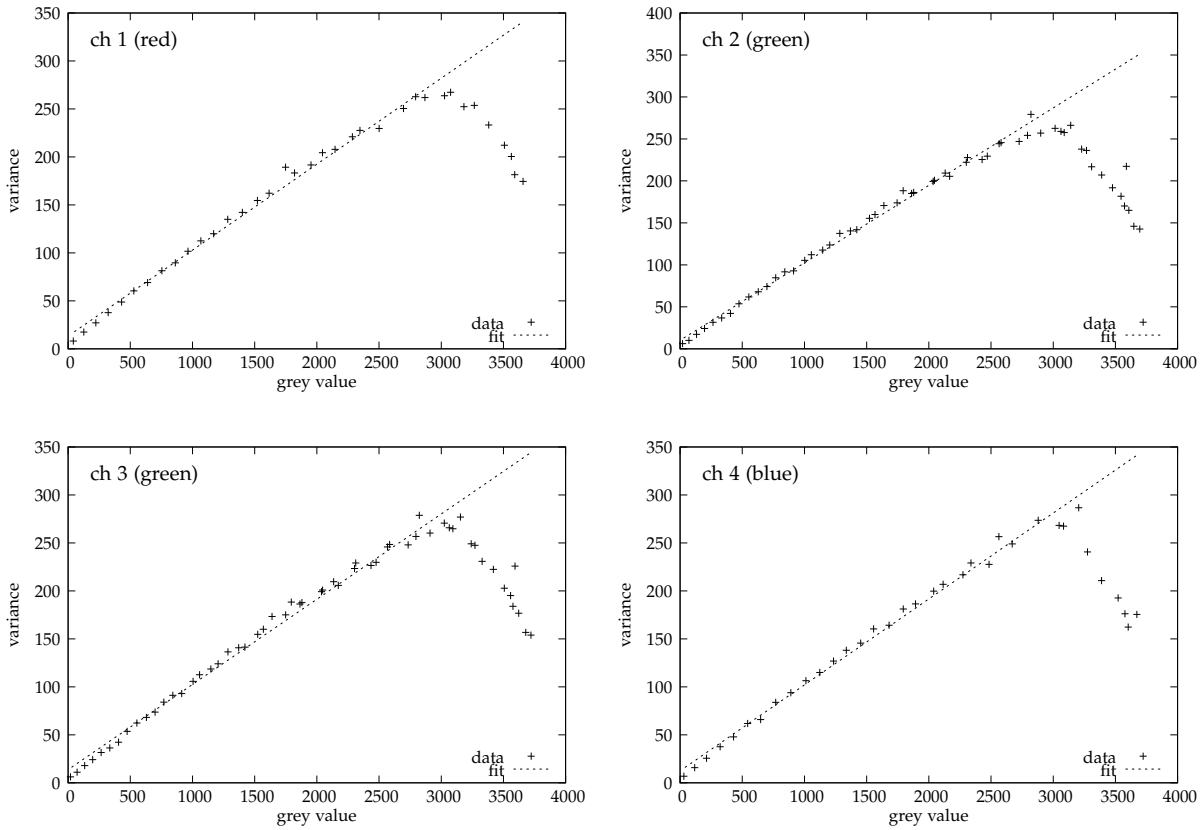


Figure 12.23: Variance model for the Canon EOS 300D with $t_e = 1/80$ s and ISO 100. The deviation at large grey values is attributed to the sensitivity structure.

times.

Since it is not possible to visually inspect all pixels for all exposure times, a few pixels were randomly chosen. Figure 12.24 shows the histogram of one pixel. The shape of the PDF is not Gaussian as would be expected. A reason for the deviation from the expected shape may be internal preprocessing in the camera before it saves the raw images. It is known that some preprocessing is done, but Canon did not give away any information about it. Also a lognormal distribution does not match the observed PDF. A moved lognormal PDF, i. e. the addition of a constant (additional fit parameter) describes the data better, but it is not suited for automatic fitting because the fit required manually tuned start parameters to converge. It is also not clear what this distribution means physically. The slightly distorted distribution has an influence on the error propagation since a Gaussian distribution is assumed here.

12.5.2 Calibration of the NIR camera

The calibration is sketched in figure 12.25. The light of a Xenon lamp was collimated with a lens and a slit. A concave grid separated the wavelengths. Then a second slit let only a small range of wavelengths through. The beam was collimated again with a lens and homogenised with a diffuser. The camera was put into the calibration unit without lens. The grid could be rotated computer-controlled for wavelength selection. The wavelength was calibrated with a spectrometer (Ocean Optics). A calibrated photo diode allowed absolute intensity measurements. For the calibration, different intensities had to be realised using different exposure times.

Three different wavelengths up to 1000 nm were used for calibration. Wavelengths above

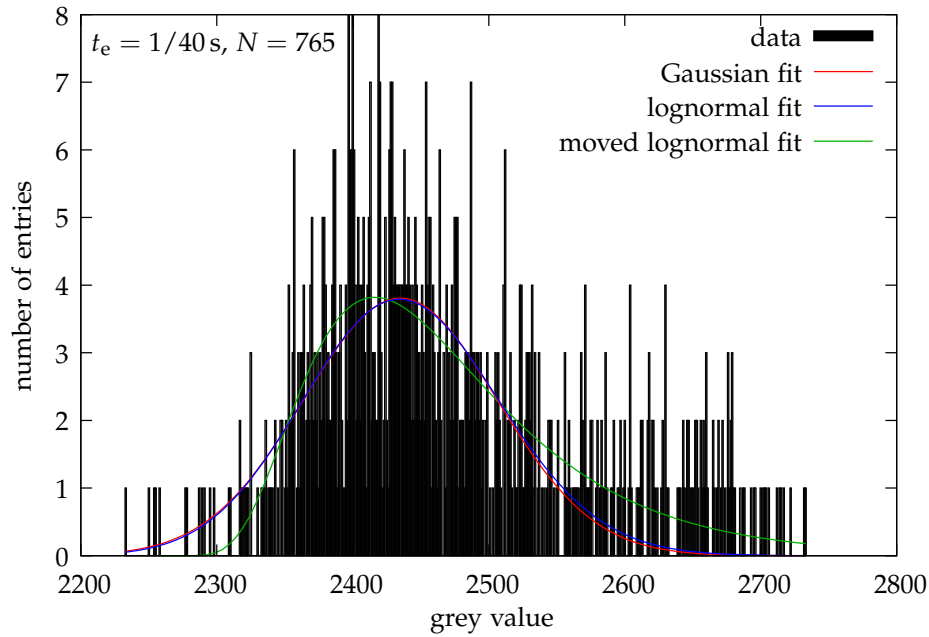


Figure 12.24: Histogram of the temporal fluctuations for pixel (881, 1413) at constant homogeneous illumination. Fits of a Gaussian, a lognormal and a moved lognormal distribution are shown.

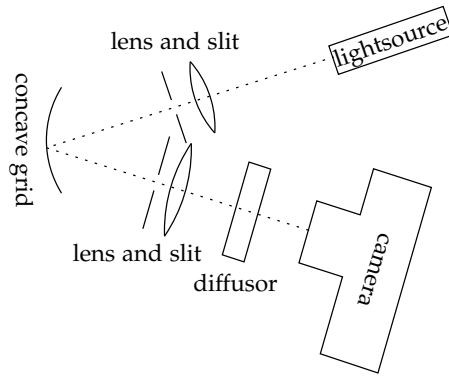


Figure 12.25: Setup used for the calibration of the NIR camera. The concave grid could be rotated to adjust the wavelength.

1000 nm were not possible to select with the setup.

The used InGaAs sensor and readout electronics used in the camera lead to different sensitivities for different pixels. This makes a spatial calibration impossible and every pixel is to be calibrated separately. 500 images were acquired for about 20 different exposure times, each for the low gain and high gain mode. During acquisition, the camera was cooled to 270 K. The acquired images were dark-corrected and mean and standard deviation of the 500 measurements was then calculated for every pixel separately. The manual inspection of the histograms of a few randomly selected pixels showed that the distribution matched closely the expected Gaussian distribution.

Figure 12.26 shows the sensor response and variance for a randomly selected pixel in low gain and high gain mode. All examined pixels looked similar. The linearity of the high gain mode is good, but not as good for the low gain mode, probably due to the additional capacitor. A parabola fits all data well. An F test was used to test if the data were also compatible with a linear fit. The linear fit was rejected at the 1 % level for nearly all pixels of the low gain and a few pixel for the high gain mode. For every pixel, the fit coefficients for the grey value – exposure time relation were saved and used for linearity correction during image processing.

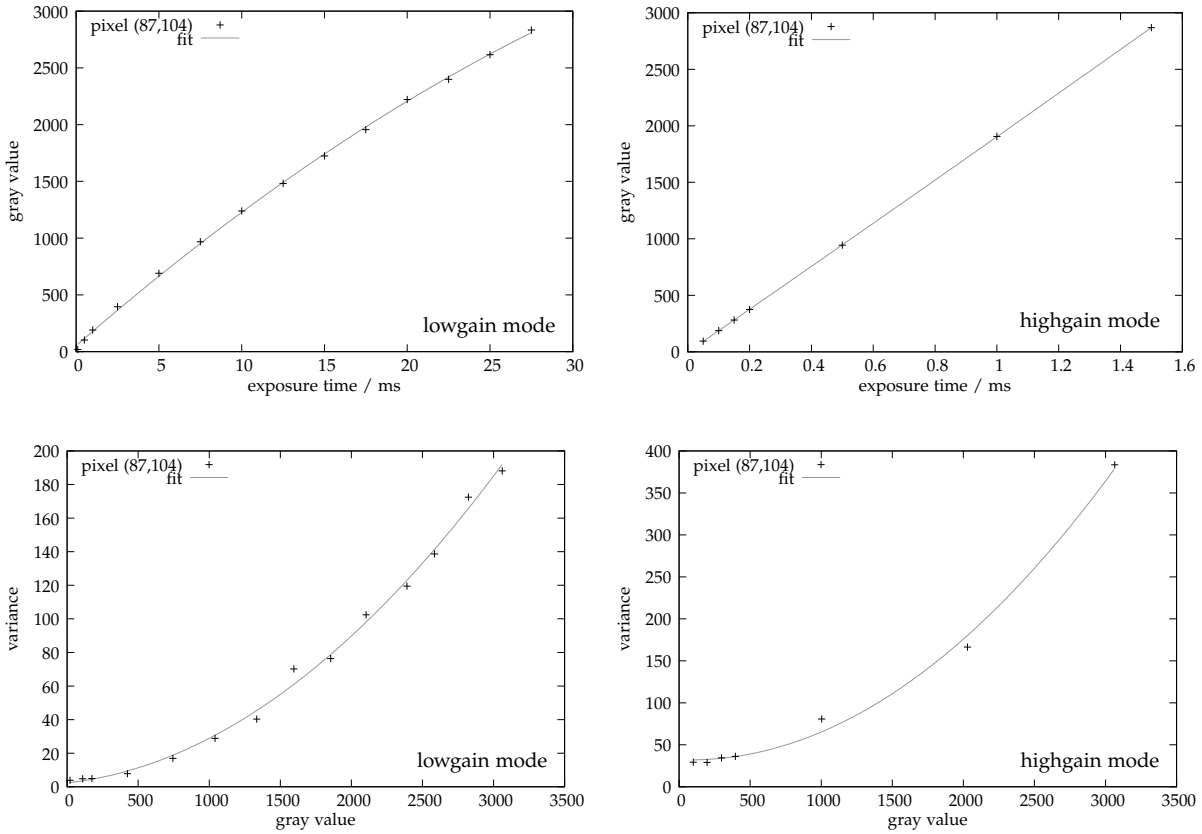


Figure 12.26: Sensor response and variance for a randomly selected pixel in low gain (left) and high gain (right) mode.

For pixels where the linear fit was accepted, the quadratic coefficient was set to zero. The fit errors of the fit coefficients were not used in the error propagation, since the correction of systematic errors is no statistical process and false errors would be introduced.

The low gain mode has a lower noise as expected. The variance is not linearly related to the grey value for both modes. A parabola describes the data well. For every pixel, the fit coefficients were saved for error estimation during image processing.

12.6 Deconvolution

Deconvolution is a method to correct the blurring caused by multiple scattering in the porous medium. Section 4.5 describes the theoretical basis of deconvolution. A deconvolution strategy which optimises sharpening versus noise amplification has to be found.

Due to the isotropy of the medium, the spreading is assumed to be isotropic at one particular water content. Different water content slightly differs the PSF, however the change was found to be in about the range of the uncertainty of the estimated PSF parameters. Since a non-uniform PSF poses a rather difficult problem, a mean PSF was determined and the non-uniformity was neglected. The PSF was assumed to be Gaussian, which closely matched the experimentally observed shape. For an isotropic medium, a Gaussian shape is expected for the limit of many scattering incidents (central limit theorem). Since the Gaussian is separable, $G(x) = G_1(x_1)G_2(x_2)$, the two components of the 2D Gaussian may be determined independently. The PSF was determined for every experiment with slits as explained in section 4.5.1 on

page 35. Vertical and horizontal slits were positioned between light source and measurement cell. The acquired image was dark-corrected and the Bayer pattern was averaged as described in section 12.4. The regions with the slits were cut from the resulting image, and the vertical and horizontal components were determined independently by averaging along the slit, derivating and fitting a Gaussian,

$$G_i(x_i; \mu_i, \sigma_i, a_i) = a_i e^{-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}}, \quad (12.5)$$

to both edges (figure 4.6 on page 36). The two σ_j obtained for the two edges $j \in \{1, 2\}$ of each slit were averaged. The resulting x and y components of σ of the 2D Gaussian, obtained from the horizontal and vertical slits, respectively, were always equal within their standard deviation. Therefore a homogeneous Gaussian with the average σ was used. The 2D Gaussian was normalised to one, i. e.

$$\int_{\mathbb{R}^2} G(\mathbf{x}) \, d\mathbf{x} = 1, \quad (12.6)$$

to avoid changing the average image value by the filter. Further steps depended on the deconvolution strategy and are described below.

All images were dark-corrected, white-balanced and Bayer-averaged before deconvolution was applied.

Band-limiting deconvolution

A Gaussian can be transformed analytically to Fourier space (Bronstein et al., 2001):

$$G(\mathbf{x}; \sigma) = a e^{-\frac{\mathbf{x}^2}{2\sigma^2}} \longleftrightarrow 2\pi\sigma^2 a e^{-2\pi^2\sigma^2 \mathbf{k}^2} = \hat{G}(\mathbf{k}; \sigma), \quad (12.7)$$

where a is defined by the normalisation constraint (12.6). The inverse optical transfer function (IOTF) is thus $\hat{q}(\mathbf{k}) = 1/\hat{G}(\mathbf{k}; \sigma)$. It was then band-limited as described in section 4.5.2 with the free parameters α and β . The special case $\beta = 0$ specifies an instantaneous cut-off with a discontinuity at $\mathbf{k} = \mathbf{k}_b$. The resulting band-limited IOTF was discretised according to the discretisation of the image to be deconvolved. After the discrete Fourier transformation (DFT), the discretisation in \mathbf{k} is $\Delta k_i = \frac{1}{N_i \Delta x_i}$ (Appendix B), where N_i is the number of points in i -direction, i. e. $N_1 = N$ and $N_2 = M$. Since $\Delta x_i = 1$ for the original image (pixel coordinates), the discretisation of \hat{q} must be $\Delta k_i = 1/N_i$. According to the definition of the DFT used by the FFTW library which was used for the calculations (see equation (B.6)), the constant frequency element $\hat{g}'_{00} = \sum_{(i,j)=(0,0)}^{(N-1,M-1)} g_{ij}$ is the total sum of all pixels, thus the normalisation in Fourier space is done by setting $\hat{g}'_{00} = 1$. \mathbf{k}_b is determined from the condition $\hat{q}(\mathbf{k}_b) = \alpha$ as $k_b^2 = \ln \alpha / (2\pi^2\sigma^2)$. With $\mathbf{k}_{ij} = (i/N, j/M)$ the resulting discrete IOTF is therefore

$$\hat{q}_{ij} = \begin{cases} e^{2\pi^2\sigma^2(i^2/N^2 + j^2/M^2)} & , (i^2 + j^2) < \frac{\ln \alpha}{2\pi^2\sigma^2} \\ r(i/N, j/M) & , \frac{\ln \alpha}{2\pi^2\sigma^2} \leq (i^2 + j^2) < (1 + \beta) \frac{\ln \alpha}{2\pi^2\sigma^2} \\ 1 & , (1 + \beta) \frac{\ln \alpha}{2\pi^2\sigma^2} \leq (i^2 + j^2) \end{cases}, \quad (12.8)$$

where r is the polynomial as defined by the requirements (4.27)–(4.30).

The edges of the source images were gradually blurred to zero with the deconvolution PSF to avoid numeric artefacts from artificial jumps at the edges of the source image when the discrete Fourier transform is applied. This was done using the Matlab `edgetaper()` function. Then the image was Fourier transformed, multiplied by the IOTF specified in equation (12.8), and inverse Fourier transformed. The constants α and β were tuned to obtain an optimal deconvolved slit image.

Lucy-Richardson deconvolution

The PSF was discretised in a regular rectangular grid with an edge size of $6\sigma + 1$ and the PSF centre placed in the middle of the region. The edges of the source images were gradually blurred to zero with the deconvolution PSF to avoid numeric artefacts from artificial jumps at the edges of the source image when the discrete Fourier transform is applied. This was done using the Matlab `edgetaper()` function. Then the image was deconvolved using Matlab's `deconvlucy()` function. Different damping factors were tested to achieve an optimum between deconvolution and artefacts. However, the results were not found to be satisfactory, since when the damping was high enough to avoid artefacts, the image was practically not deconvolved anymore.

Wiener filter

The PSF was discretised the same way as described above for the Lucy-Richardson deconvolution. The source image was transformed to have a grey-value independent noise using equation (4.20). The transformation parameters were extracted from the camera calibration (section 12.5.1). Of course the PSF is not transformed since it is normalised. The image was then edge taped as described above, and the Matlab `deconvwnr()` function was applied. Several approaches were used to estimate the noise term $\nu := S_{nn}/S_{ff}$: (i) Since the image was transformed to obtain a constant noise of 1, $\nu = 1/\sigma_h^2$ was used, where h is the transformed source image. This is the straight-forward approach which uses the known properties of the image. Note that the properties of the original image are estimated using the blurred image, since the information about the original image is not known. (ii) $\nu(\mathbf{k}) = NM/|\hat{h}(\mathbf{k})|^2$, which uses the actual power spectrum of the convolved image as an estimator for the power spectrum of the original image. (iii) The image F was estimated by applying a 3x3 median filter to the blurred image, the noise n by the difference between filtered and original image. Then ν was calculated by $\nu(\mathbf{k}) = |\hat{n}(\mathbf{k})|^2/|\hat{f}(\mathbf{k})|^2$.

Influence of noise

Noise has a major impact on the results of the deconvolution process. To investigate this, a test image was convolved with a Gaussian PSF and then deconvolved with the cut-off band-limit method (section 4.5.2). Negative intensities, which are obviously a deconvolution artefact, were clipped to zero. The result is shown in figure 12.27. Without noise, the image can be reconstructed reasonably (d). However, if only a small amount of Gaussian noise ($\sigma = 1$, $\mu = 0$) is added after the convolution, the deconvolution results are relatively poor (e, f).

If the zero border around the original image is not large enough, the deconvolved image is dominated by artefacts (figure 12.27 (c)). This is caused by the smear-out of information beyond the image boundaries when the image border is smaller than the influence of the PSF. Due to the periodicity assumption, that information reappears on the other image boundary, where it interferes with the information at that location.

Comparison of different deconvolution methods

The performance of different deconvolution methods on the acquired images has been analysed to choose an optimal deconvolution strategy.

To visually inspect the blurring, and the performance of the deconvolution, a direct image of the slit was taken by placing the slit between the sample and the camera. With this setup, only the blurring in the optics enters the image. This is only a minor component compared to the effect of the porous medium. The slit image which was also used to determine the PSF was deconvolved and compared with the direct slit image. The results for the optimal choice of parameters are shown in figure 12.28, a profile in figure 12.29.

The Lucy-Richardson deconvolution has the best result. However, it also introduces speckle-type artefacts (section 4.5.3). This behaviour is not observed when deblurring the slit, but

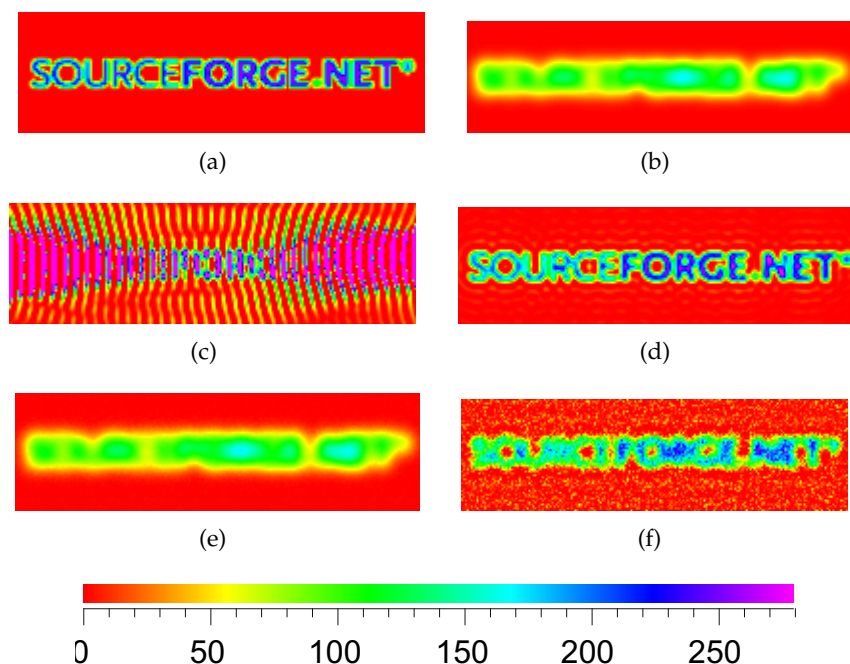


Figure 12.27: Test of the influence of noise on the deconvolution process. A test image (a) was convolved with a symmetric Gaussian PSF of $\sigma = 4$ (b) and then deconvolved with the cut-off band-limit method. With too small a border around the original test image, the deconvolved image contains artefacts (c). If a larger zero border is added to (a), the deconvolution (optimal threshold $\alpha = 10^{11}$) works fairly well (d). If a small amount of Gaussian noise with $\sigma = 1$ and $\mu = 0$ is added to the blurred image (e), the results (optimal threshold $\alpha = 2 \cdot 10^3$) are much worse (f).

enters significantly when deconvolving real images with extended regions, as can be seen in figure 4.7 on page 39. Reducing the number of iterations or increasing the damping parameter decreases the artefacts but also decreases the sharpening. When the artefacts are reduced to an acceptable level, sharpening is so low that the effect of the deconvolution is hardly visible (figure 12.31).

The cut-off method also significantly reduces the blurring, but introduces artificial oscillations. Again, reducing the artefacts reduces the sharpening. The influence of α on the result is shown in figure 12.30. The higher α , the better the sharpening, but the worse the oscillations. When the oscillations have ceased the sharpening is very small.

The Wiener filter blurs the image far too much. According to section 4.5.4 this indicates overestimated noise. However, the filter was very insensitive to the noise-to-signal ratio. Even with the noise-to-signal-ratio set to zero the deconvolution algorithm did the disproportional blurring.

After investigating the different deconvolution strategies, it was concluded that the deconvolution procedure introduces more artefacts than it helps. That impression was augmented by the fact that for example the image analysis groups at the HCI and the MPI for astrophysics in Heidelberg also do not deconvolve, since they think that deconvolution harms more than it helps. Thus, deconvolution was abandoned.

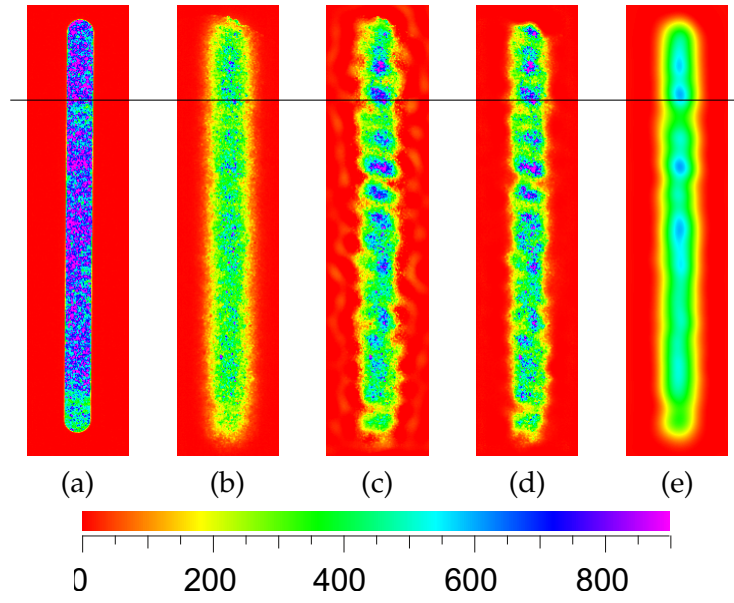


Figure 12.28: Comparison of the different deconvolution methods using the slit image: (a) direct slit image, (b) Original image, (c) Cut-off ($\alpha = 8$, $\beta = 0.2$), (d) Lucy-Richardson (15 iterations, no damping), (e) Wiener filter (NSR 10^{-9}). Negative intensities were set to zero. The black horizontal line marks the location of the profiles shown in figure 12.29.

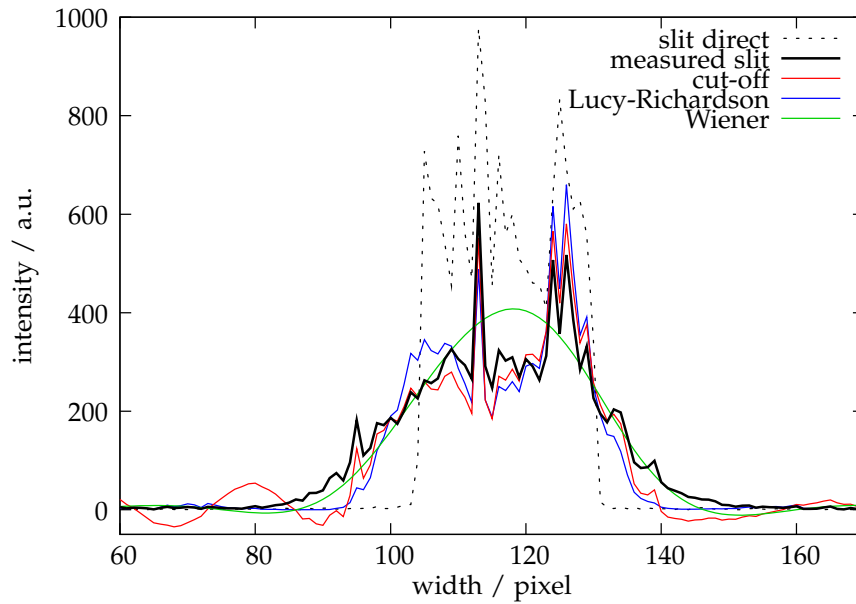


Figure 12.29: Horizontal profiles of the deconvolved slit images shown in figure 12.28.

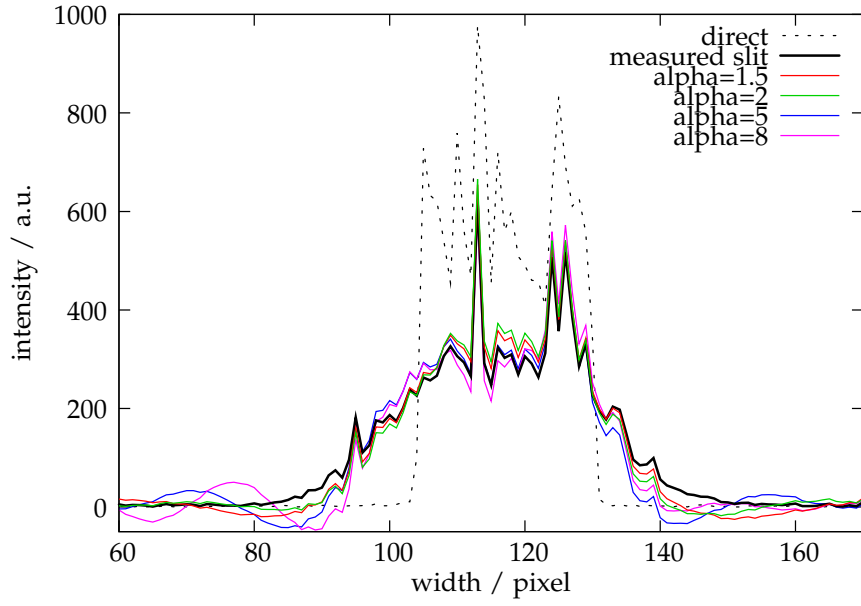


Figure 12.30: Comparison of the influence of different cut-off values α on deconvolution.

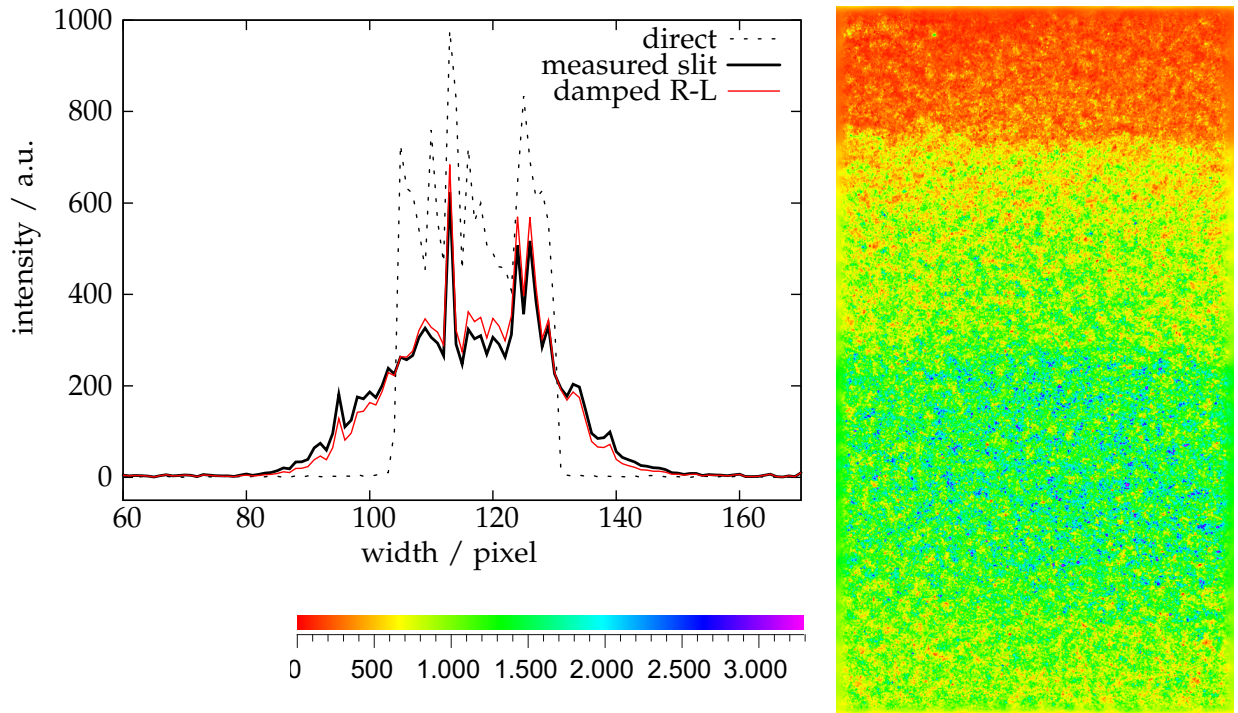


Figure 12.31: Damped Lucy-Richardson deconvolution (3 iterations, damping parameter 4) on the slit (profile, left) and a measured image (right). Although the sharpening effect has become very small, the artefacts are still quite distinct.

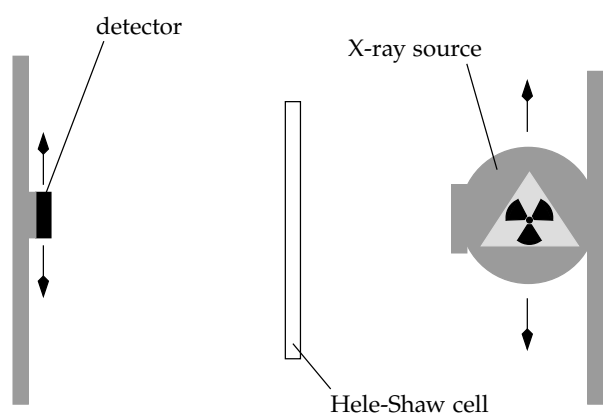


Figure 12.32: Setup of the X-ray measurement device

12.7 X-ray measurements

For calibration of LTM, an X-ray facility was used (Bayer, 2005). It consisted of a medical polychromatic X-ray tube (Philips Optimus Rad) operated at 141 keV, a horizontal 12 bit CCD line detector with 1280 pixels of 0.4 mm side length (Hamamatsu C7390) and a computer-controlled positioning device for tube and detector which allows vertical movement. The Hele-Shaw cell was positioned in-between, in parallel to the detector. To obtain a two-dimensional image, tube and detector were moved synchronously and horizontal lines were acquired sequentially. A sketch of the setup is given in figure 12.32.

To readjust the tube vacuum, a certain procedure has to be executed regularly. To ensure optimal tube operation, this was done before every measurement. An adjustable copper slit directly at the tube is used for beam collimation by absorbing radiation with large incident angles. Yet the resulting radiation is still a fan beam, which limits resolution. For quantification, a copper marker of 3 mm thickness was installed horizontally on the empty Hele-Shaw cell and a measurement with a horizontal resolution of 0.26 mm was made. The edges of the copper block were smeared out over 4 pixel, which indicates a horizontal resolution of about 1 mm for a medium with 3 mm thickness like the medium in the Hele-Shaw cell used here.

Due to the small thickness of the cell, the change in attenuation between the dry and the wet cell is relatively small. The relative change of attenuation between water and air is about $\Delta\mu = 0.03 \text{ mm}^{-1}$. With a porosity of 0.4, the absorption length of pore volume is only 1.2 mm, which leads to a relative change of intensity between fully saturated and dry conditions of $I_{\text{wet}}/I_{\text{dry}} \approx 0.96$. This is not sufficient for significant results. Thus, the contrast enhancing agent potassium iodide (KI) was used. Iodide has a high mass and therefore exhibits high X-ray attenuation. The attenuation enhancement was estimated in a test experiment (table 12.3) as $\Delta\mu = (4.7 \pm 0.6) \times 10^{-4} \text{ l/(g mm)}$. A nominal concentration of 400 g/l was used, which leads to a relative intensity change of $I_{\text{wet}}/I_{\text{dry}} \approx 0.76$ and is still much lower than the solubility (1430 g/l at 20 °C). However iodide may precipitate from the KI solution due to light illumination and air contact. This leads to a yellow colour which changes the visible light transmission, and may also cause a change in the X-ray absorption coefficient. Accordingly, it falsifies the results. Therefore, precipitation must be avoided by minimising measurement and light illumination time.

Due to Compton scattering, photons do not always travel linearly through the medium, which causes a deviation from Lambert-Beer's law. The small cross section of Compton scattering for 141 keV makes this effect small. The polychromatic nature of the X-ray source (beam hardening effect, section 5.1.2) also leads to deviations from Lambert-Beer's law. A measurement of the photon energy distribution of the used tube is given by Bayer (2005, sect. 2.1.3). Taking this non-linear behaviour into account would require a large amount of information

Table 12.3: Change of the attenuation due to the contrast enhancing agent KI. Relative intensities were measured for an attenuation length of $d = (13.5 \pm 1.0)$ mm for several concentrations C . The attenuation due to KI depending on the concentration was fitted as $\mu = (0.035 \pm 0.002) \text{ mm}^{-1} + (4.7 \pm 0.6) \times 10^{-4} \text{ l g}^{-1} \text{ mm}^{-1})C$.

$C / \text{g/l}$	$I_{\text{wet}}/I_{\text{dry}}$	$\mu / \text{l/mm}$	$I_{\text{wet}}/I_{\text{dry}}$ for $d = 1.2 \text{ mm}$
0	0.6154 ± 0.0006	0.031 ± 0.003	0.964 ± 0.003
10	0.5798 ± 0.0006	0.040 ± 0.003	0.953 ± 0.003
50	0.4802 ± 0.0005	0.054 ± 0.004	0.937 ± 0.004
100	0.3106 ± 0.0004	0.087 ± 0.006	0.901 ± 0.007

which is practically not available. As [Tidwell and Glass \(1994\)](#) pointed out, the chamber glass acts as filter, absorbing a significant amount of the low-energy photons and leaving a less polychromatic beam. Therefore, non-linear effects were neglected and equation (5.23) was assumed to be valid.

A calibration measurement of the detector revealed a pixel error of about $145 + 0.5I$, where I is measured intensity. Of course the actual numbers depended on the pixel, for example the offset ranged from -16 to 266 . Anyway the error is large. Remembering that due to the small cell thickness, small attenuation changes must be detectable, this error is clearly not acceptable. Therefore, temporal averaging was done to minimise the error. With $N = 500$ measurements, the error is reduced already by a factor of about 22. This yields a reasonable error. Averaging has the additional advantage that the measurement error can be estimated by the ensemble standard deviation.

Drift

To check the long-term stability of the X-ray measurements, a test measurement with no sample was done every 5 min for 24 h with always the same settings. At every time step $N = 500$ measurements were made and averaged for noise reduction and standard deviation estimation. The temporal evolution of the acquired pixel values was then examined. Ideally, the value should stay constant during the whole measurement. Figure 12.33 (left) shows the value of two randomly selected pixels during the 24 h period. The standard deviation of the measurement as calculated from the ensemble mean is about 1.9 and indicated in the figure by the line thickness. Apart from the high-frequency noise there is a significant drift with more than 10σ . It seems that the different pixels are drifting similarly. However at some times the drift evolves somewhat different for the different pixels, which can be seen at about $t = 12 \text{ h}$ in the left plot. For quantification the correlation of one of the two shown pixels with all other pixels as estimated from the whole measurement time is shown in figure 12.33 left. Indeed the correlation is very high, but some pixels break the correlation. It is not clear what causes the drift. Sources might be changes in detector sensitivity, the detector readout, or of the X-ray source intensity due to temperature.

To correct the systematic error introduced by the drift, the drift is measured at a reference pixel which has a constant attenuation. The other pixels are then corrected weighted with the correlation with the reference pixel, such that less correlated pixels will be corrected less. Otherwise artefacts would be introduced if one pixel is corrected although it changed in another way than the reference pixel. Averaging multiple reference pixels is not possible since this would change the correlation. The correction factor itself is calculated the same way as in the drift correction of the visual images. Thus, pixel values are updated according to

$$I'(x, y) = I(x, y) \left\{ \left[\frac{I_{\text{ref}}}{I(x_{\text{ref}}, y)} - 1 \right] \text{cor}(x, x_{\text{ref}}) + 1 \right\}, \quad (12.9)$$

where I_{ref} is the value of the reference pixel at the reference line. This linearly maps the

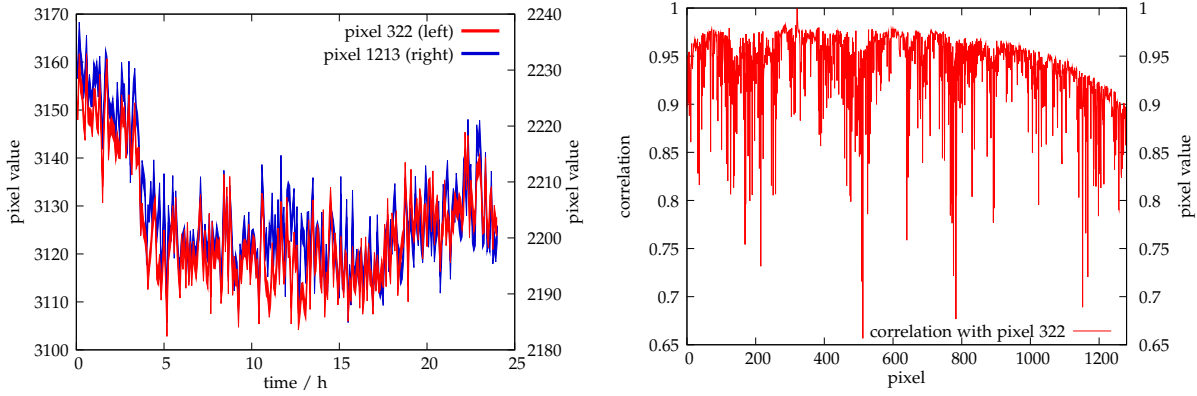


Figure 12.33: Drift test measurement of the X-ray facility: Temporal evolution of the pixel value on two randomly selected pixels for 24 h (left, the thickness of the lines show the 1σ error), and the correlation of the first displayed pixel with all other pixels (right).

correction term such that it fully corrects with $\frac{I_{\text{ref}}}{I(x_{\text{ref}}, y)}$ if the correlation is 1 while the correction factor is 1 (no correction) if the correlation is 0. To avoid line-type artefacts caused from random deviations of the reference pixel, the $I(x_{\text{ref}}, y)$ profile is smoothed using a Gaussian filter with $\sigma = 2$ before applying the correction. A remaining problem is that the high-frequency changes of the covariance may introduce vertical line-type artefacts. However, no smoothing is possible here since it would disturb the statistical properties. These remaining artefacts are small and considered superior to the artefacts introduced by the drift if no correction was made.

12.7.1 Water content calibration

As explained in section 5.3.2, X-rays do not exhibit multiple scattering due to macroscopic material property changes, which allows to use equation (5.23) for directly calculating the water saturation from X-ray data. A calibration experiment was done to relate the LTM normalised intensity data to water content. Since the X-ray measurement takes about 5 h, the sample must be in equilibrium. Otherwise the sample would change during the measurement and different parts of the image would describe different water content distributions. Images were taken with both X-ray and visual light as described in sections 12.7 and 12.3, respectively. Water with (401 ± 6) g/l KI was used for contrast enhancing of the X-ray measurement as described above. To prevent precipitation of the KI, light illumination and total experiment time were minimised.

After filling the sample and installing copper reference markers, measurements of the dry cell were made. The sample was saturated by a slowly increasing water table of 30 mm/h and equilibrated for 1 h. Then the saturated images were acquired. The watertable was lowered again with 30 mm/h and then equilibrated for 7 h. Since the water table had fallen too low, it was again risen by 20 mm with 30 mm/h and again equilibrated for 3 h. Then the water content profile images were acquired. After finishing the last X-ray image, negligible change of colour of the KI solution was detected. X-ray measurements were done with 141 kV tube voltage, 4 mA tube current, and an exposure time of 10 ms. For each line, 250 scans were averaged. I_0 was measured with a tube current of 1 mA and 500 scans were averaged. Visual images were acquired with an aperture of 5.6 and an exposure time of 1/13 s. 50 images were averaged. Resolution was (0.340 ± 0.001) mm/pixel.

If the partly saturated image had been taken before the wet image, the whole water content

range would have been available, whereas for the measurement procedure used here, saturation between residual and dry was not part of the calibration. The reason was that if the sample is equilibrated much longer at one particular water table position, which would be necessary for the measurement, and afterwards wetting is resumed, due to the different water configuration at that point a water content minimum is found just above the position of the capillary fringe during the partly saturated configuration. This would introduce artefacts in the wet image. Since in this work the focus is on the capillary fringe and low watercontents are not of primary interest, the limited calibration range has been considered superior to artefacts in the calibration measurement.

Visual images were evaluated as described in section 12.3.4, however in contrast to the normal evaluation, of course F^{-1} was not invoked. X-ray images were drift-corrected as described above, and coordinates were shifted to overlay the images. Shifting was necessary because the cell was moved from one facility to the other during the experiment, which causes slight changes in position. Registration was performed with Matlab using the edges of the copper markers as reference. To obtain the offset, the images were binarised. The threshold was chosen to be the middle value between the the copper marker and mean surrounding intensities. As a first start, points were set manually and then cross correlation was used for coordinate system mapping. Only shift and no scaling was used for correction. Saturation was then calculated according to equation (5.23) on page 49. Pixel 1268 in the light source window was chosen as reference pixel, and the uppermost line of the dry image as reference line.

Since X-ray and visual images were acquired at different resolutions, registration was needed to overlay. A projective transformation was used for mapping both images. Again, registration was performed with Matlab using the edges of the copper markers as references. After estimating the coordinate system mapping as described above, projection parameters were estimated in a least square sense from the mapped coordinates. The X-ray saturation image was then transformed using that transformation to finally overlay the visual normalised intensity image. This involved interpolation of pixel values. A Gaussian filter with radius 12 pixel (4 mm) was applied to both X-ray and visual data sets to obtain macroscopic REV-scale water content values. This is somewhat larger than the correlation length of the medium of 3 mm (section 12.2). Another reason why averaging is required is that the measurements take place at different scales. While X-rays propagate more or less linearly through the medium, due to multiple scattering, visual light “explores” a broader area of the medium, which is quantified by the point spread function. Measurements can only be compared if a scale larger than the PSF extend is averaged, since only then the same quantity is considered in both measurements. Figure 12.34 shows the two data sets. The lower two images show the propagated measurement uncertainty. It depends on the value since the uncertainty of the measured pixel intensity depends on the grey value. The error is relatively low and the quality of the data is good. The typical relative error σ_{Θ}/Θ is 0.7 % for X-ray and 0.6 % for visual images.

For each pixel, Θ as calculated from the X-ray measurement and $F(\Theta)$ as calculated from the visual light transmission data were assigned to each other. No horizontal averages were made, since a horizontal average will be disrupted by the horizontal inhomogeneity of the water content distribution caused by medium heterogeneity. The result of the matching is shown in figure 12.35 (a). Values are in a plausible range. However, the broad range of $F(\Theta)$ values assigned to each Θ makes a meaningful relation impossible. The range of spreading is much larger than the errorbars and therefore must stem from another source. Looking at figure 12.34 shows large-scale structures in the visual image which are not present in the X-ray image. These are attributed to peaks caused by imperfect registration and changes of the medium during the experiment. If such peaks are present, they bias the Gaussian averaging and cause artifacts on the REV scale which are not present in the original medium. They also cannot completely removed by median filtering, since too large a filter size disturbs the correct

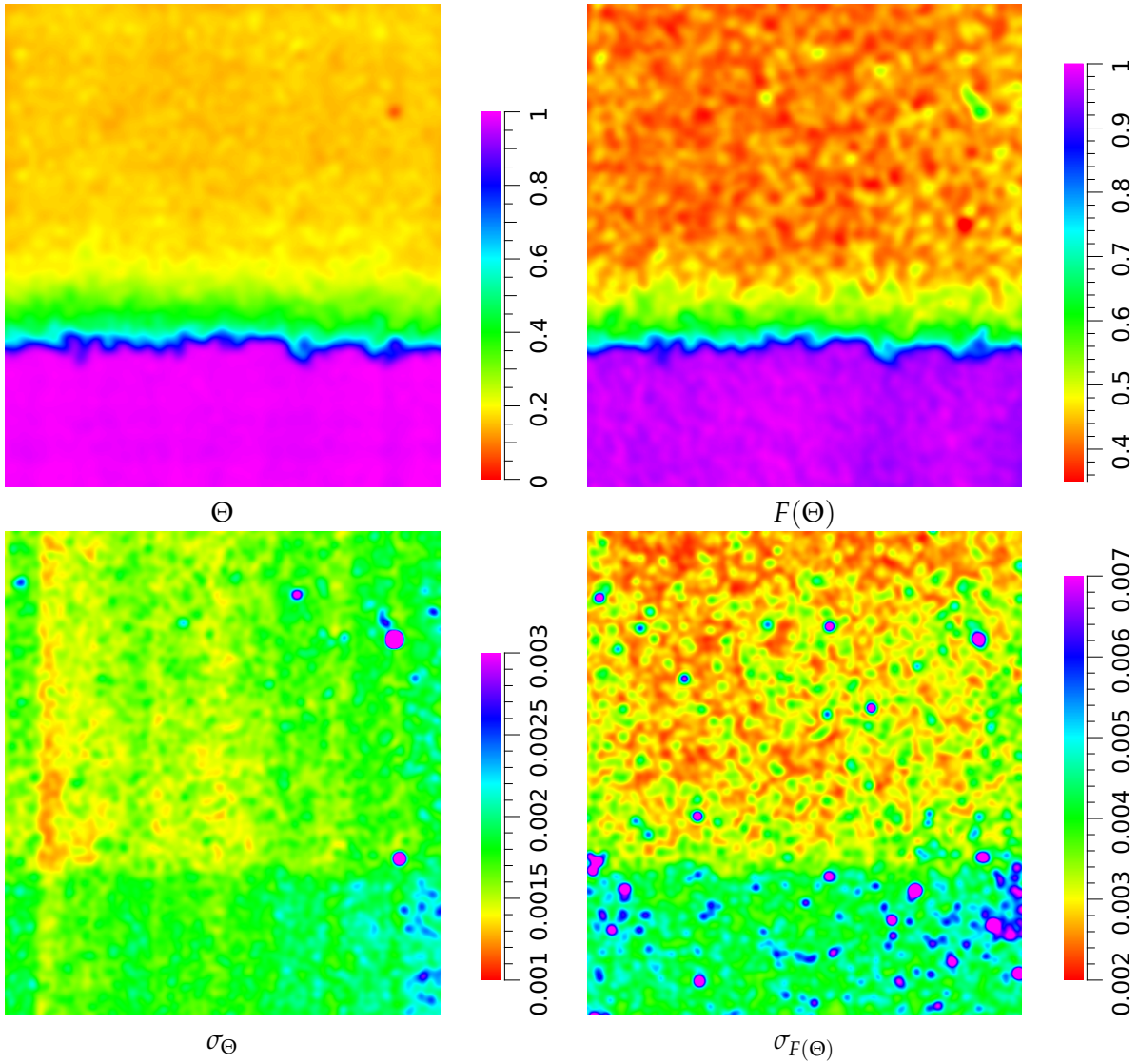


Figure 12.34: Saturation Θ measured by X-ray attenuation (left) and corresponding visual light transmission data $F(\Theta)$ (right). Image dimensions are 192 mm \times 214 mm.

features of the image. Gaussian averaging also introduces large correlations between pixels which may lead to circular structures not present in the original image, as observed in a few locations. This is also seen in the direct matching as round structures. The deviation in the upper right is caused by a “hole” in the medium. Grains were laying such that a large pore was created where direct visual light came through. This resulted in a high intensity for visual light. In the static equilibrium image, this larger pore was drained and had less water than in the surrounding which was correctly identified by the X-rays due to lower absorption. General features, for example the capillary fringe boundary, are well represented in both images and also the structure of the CF is consistent. This shows that, apart from small deviations, both images are consistent and can actually be matched. Therefore, a statistical treatment of the data similar to Buchner (2009) was applied.

If the remaining structure can be described statistically, for one Θ interval a range of corresponding $F(\Theta)$ values is expected. Hence, it is plausible to obtain the most likely $F(\Theta)$ value and then use it for the calibration function. This was done by parting Θ into intervals of the width of the mean error of Θ , which was 0.002. The mean of all $F(\Theta)$ values within this in-

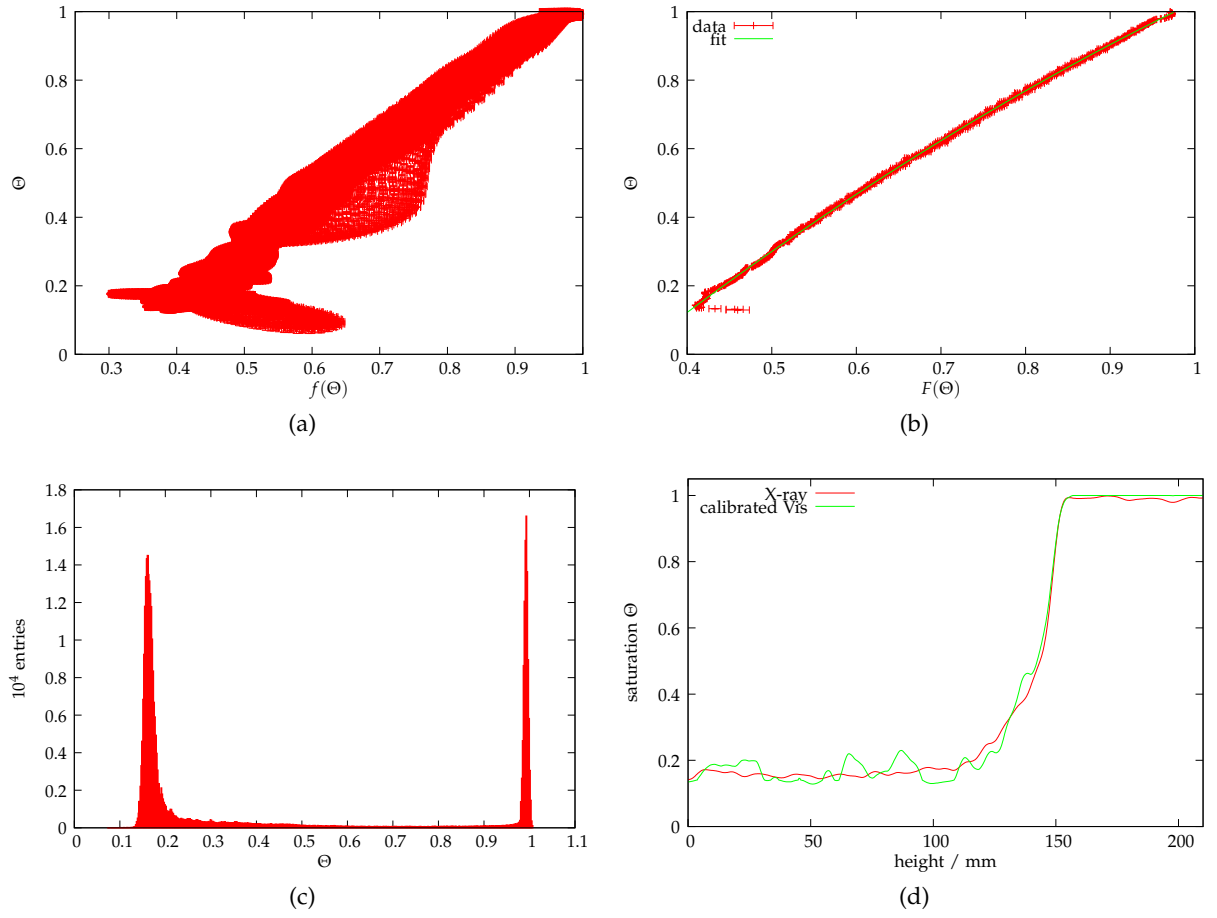


Figure 12.35: (a) Pixelwise assignment of visible light transmission data $F(\Theta)$ to X-ray measured saturation Θ , errors were calculated by error propagation from the measurement errors. (b) Statistical evaluation of the values in (a). Θ was divided into partitions of width 0.002 and the mean was estimated. (c) Number of entries in each interval. (d) Test of the calibration. The visual light transmission image was evaluated as in the normal experiments, a vertical profile with 2 pixel horizontal average was made and the result was compared with the matching vertical profile of the X-ray measured saturation.

Table 12.4: Fit parameters and covariance matrix of the calibration relation F .

i	0	1	2
c_i	-0.53572	2.25961	-0.69589

$$\text{cov} = \begin{pmatrix} 1.8857 & -2.5748 & 0.82793 \\ -2.5748 & 3.5564 & -1.1578 \\ 0.82793 & -1.1578 & 0.38292 \end{pmatrix} \cdot 10^{-4}$$

terval was estimated and used as corresponding value. This resulted in figure 12.35 (b). The error bars were estimated by error propagation. Figure 12.35 (c) shows the number of entries in each interval. Since no measurement with saturation below residual ($\Theta < 0.15$) is available, it is clear that no calibration is possible in that range.

Although figure 12.35 (a) looks rather bad at first sight, the statistical evaluation used here which leads to figure 12.35 (b) is still considered superior to horizontal averages, since the latter introduces systematical errors due to the neglect of heterogeneity, while the statistical approach only averages matching saturations and does not bias the results.

A polynomial $\Theta(I) = \sum_{i=0}^N c_i I^i$ of grade $N = 2$ was fitted into the data. It was then used for saturation calibration of measured visual light transmission data during experiments. Table 12.4 shows the fit parameters and the covariance matrix.

To test the saturation calibration, the visual light transmission data was re-evaluated according to the normal evaluation procedure as described in section 12.3.4. A 12 pixel Gaussian filter was applied to come to the REV scale and a vertical profile at a randomly chosen horizontal location was made. It was then compared with the corresponding region in the X-ray measured saturation. The result is shown in figure 12.35 (d). The data fits well to the actual saturation, which indicates that the calibration is correct. For lower saturation values, the visual light data has much more noise than the X-ray measurement. This is attributed to the artefacts described above. However, since the focus of this work is at higher water contents, this is not considered to be problematic.

An inherent assumption of the calibration is that $F(\Theta)$ is spacially isotropic, i.e. the same throughout the cell, since saturation values at different locations were used as calibration and not different saturation values at each point. This means that the light propagation in the porous medium is assumed to be independent of the position.

12.8 Additional instruments

12.8.1 Pumps

For infiltration experiments, a computer-controllable multi-channel peristaltic pump (Ismatec IPC-12) was used. The pump allowed setting the pump speed in 1 % steps. 11 infiltration tubes regularly distributed along the top of the Hele-Haw cell were used for infiltration. The homogeneity of the infiltration was maximised by adjusting the tube pressure on the different pump channels. Homogeneity was measured by running the pump for a defined time period and collecting the amount of pumped water separately for every channel. Remaining non-fixable flow deviations were typically 4 %. The heterogeneity of the flow field is shown in figure 12.36.

Total flow was calibrated by pumping water from one reservoir to another. Both reservoirs were placed on digital balances which were automatically read out. Pump speed was changed in 10 % steps. Flow rates were estimated by fitting lines on the balance readings for every pump speed. The error was estimated by the fit error. The total flow per percent pump speed was estimated as $(0.00294 \pm 0.00002) \text{ ml s}^{-1} \text{ \%}^{-1}$. This was used as initial guess for the pump speed. During measurements, the actual flow was measured with a balance by weight change of the reservoir and pump speed was corrected according to these measurements. For data evaluation, the actually measured flow was used.

12.8.2 Balances

For flux estimation, digital balances (Precisa XB 10200D) with 0.1 g accuracy were used. The balances could be automatically read out by the control computer via a serial RS232 connection.

12.8.3 Temperature measurements

Temperature was measured using Pt1000 probes. This is a temperature-dependent resistance with a nominal value of $1000\ \Omega$ at $0\ ^\circ\text{C}$ and a temperature dependence of $3.9083\ \Omega\ \text{K}^{-1}$. The tolerance of the sensor was $\pm 0.52\ \Omega$ (1/3 DIN EN 60751 class B) which corresponds to $\pm 0.13\ \text{K}$ at $20\ ^\circ\text{C}$.

The resistance of the sensor is measured with a triplex lead: a current is sent through the sensor through wire A and ground wire, while the voltage is measured between ground and wire B. A commercial measurement amplifier conditions the signal before it is converted in a 16 bit ADC in the control computer.

Calibration

The deviation of the Pt1000 elements from their nominal value is small compared to the deviation of the electronics. Therefore, only the electronics were calibrated by using a fine-tunable known resistance instead of the Pt1000 element. A resistance corresponding to a known temperature was set and the corresponding digit value was read from the ADC. This was done for several values and a line was fitted through the data.

As a test measurement, the probe was put into ice water with a temperature of $0\ ^\circ\text{C}$. The deviation of the sensor was less than $0.2\ ^\circ\text{C}$. Further tests with the tunable resistance showed that the error introduced by the electronics and the ADC was less than $\pm 0.1\ ^\circ\text{C}$. Thus, the accuracy of the temperature measurement was considered $\pm 0.2\ ^\circ\text{C}$.

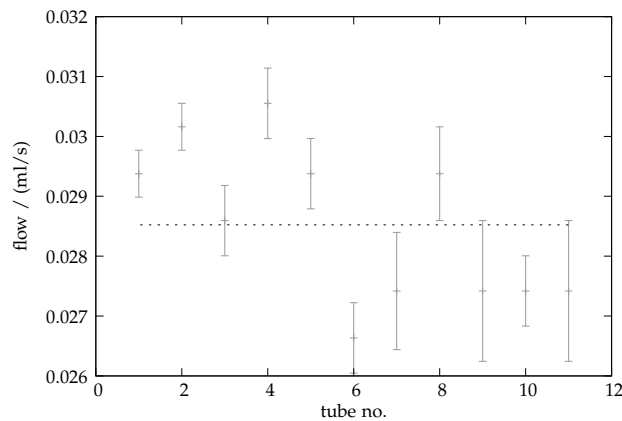


Figure 12.36: Flow inhomogeneity of the different infiltration tubes. The black horizontal line indicates the mean flow.

13 Results and discussion

Scales

Being aware of the difficulties described in section 2.3, an optimal strategy for data evaluation is to be chosen. Figure 13.1 shows the measured water distribution at original resolution and at an REV scale. As expected, large-scale features are still present in the REV images but sub-REV scale fluctuations caused by the grain structure are gone. The blurring of front width in the REV images is also obvious. The resolution to be used for the images depends on what features are to be evaluated. Of course, to look at water content θ , REV scale images must be used since θ is undefined below the REV scale. However, if the focus is on the front movements and sub-REV processes like non-equilibrium effects are to be resolved, the sub-REV scale is more suited than the REV scale, with the drawback mentioned before. As shown in the evaluation, non-equilibrium processes play a crucial role in the dynamics of the capillary fringe.

In the following, processes are viewed at three different scales as appropriate: (i) at the REV scale, where macroscopic water content is investigated, (ii) at a sub-REV scale, where water distribution is used for front movement and dynamics evaluation, but no single grains are resolved, and (iii) a microscopic scale where single grains and pores are resolved.

13.1 Dynamics of a fluctuating water table

13.1.1 Measurement protocol

The sample was prepared as described in section 12.2. Porosity was estimated gravimetrically to $\phi = 0.41 \pm 0.05$. Initially, after taking images of the dry sample, the sample was slowly saturated from below with the water table (WT) raising at 30 mm/h. Images of the wet sample were taken and the water table was slowly decreased at 30 mm/h to a middle position. Then the sample was equilibrated for a few days. Subsequently the water table height z_{wt} was shifted periodically, either rectangular, $z_{wt} = z_0 + R(2\pi t/T)$, or sinusoidal, $z_{wt} = z_0 + A_{wt} \sin(2\pi t/T)$. $R(t) = \sum_{j < t/\pi} (-1)^j h(j\pi)$ with h being Heaviside's function (4.23) is the unit rectangle function. Boundary conditions are shown graphically in figure 13.2. Different amplitudes A_{wt} and periods T were applied. During runtime, water content distribution was observed with a temporal resolution of 2 min. Three initial experiments with a rectangle and $A = 15$ cm amplitude were conducted to bring the experiment in a more natural state. In typical field conditions, the water table is found in a state where many changes already occurred and the region above the water table is already wet. Before every experiment, the sample was equilibrated for 12 h. The amplitudes and periods used in the different runs are shown in table 13.1. The different experiments were run automatically in sequence, the different amplitudes were run in a random order to preclude artefacts which might be produced by increasing or decreasing amplitudes.

To also investigate the dynamics on rapid groundwater changes after a dry period, for example due to heavy rainfall, a few additional experiments with an initially dry medium were conducted. Here, after sample preparation and acquisition of the dry images, the water table was slowly raised from below at 20 mm/h to a middle position in the initially dry medium and equilibrated for about 12 h. Then the experiment was conducted. Afterwards the medium was slowly saturated with the water table raising at 20 mm/h and the wet images were acquired.

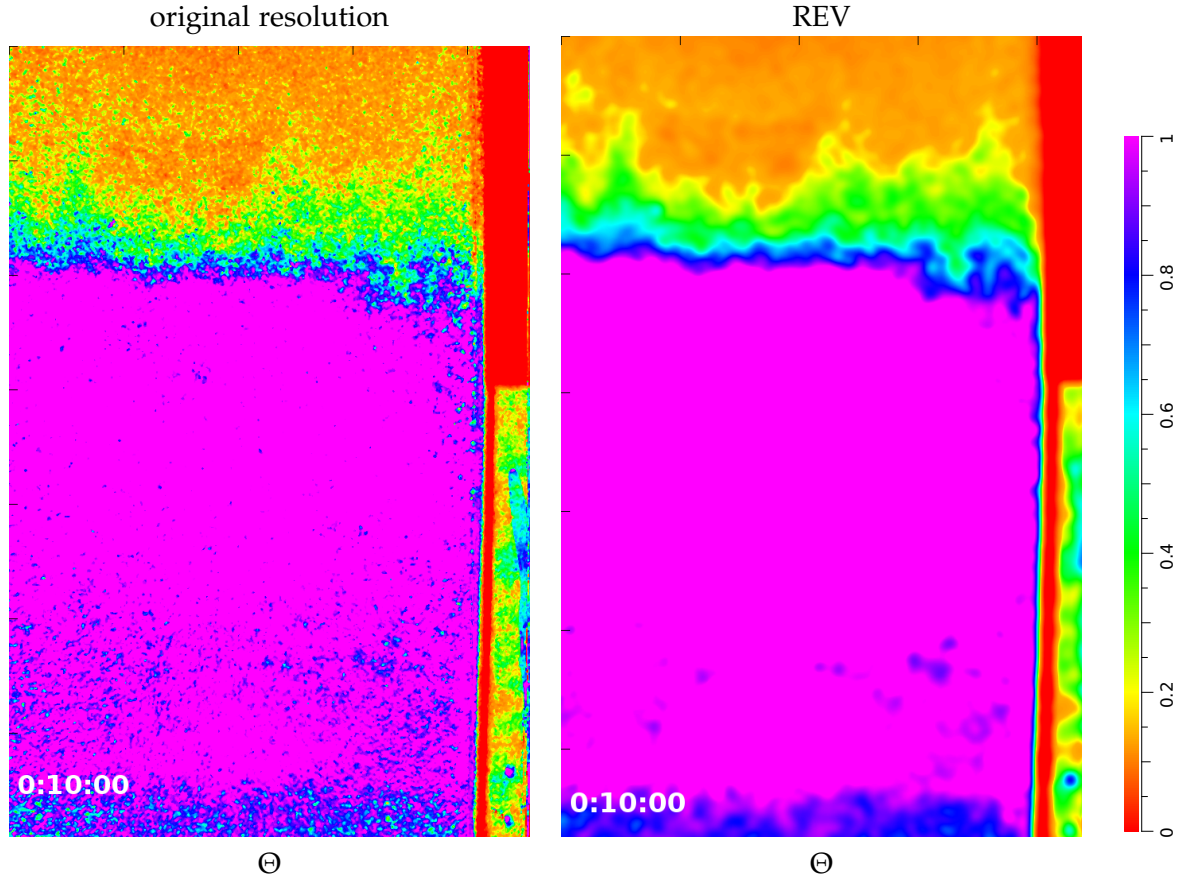


Figure 13.1: Sub-REV original (resolution (0.346 ± 0.001) mm/pixel) and REV scale images of water distribution measured with LTM. REV data was estimated by Gaussian averaging. Typical propagated relative measurement error σ_{Θ}/Θ is 2 % for original and 0.1 % for REV resolution. Black scale ticks correspond to 50 mm.

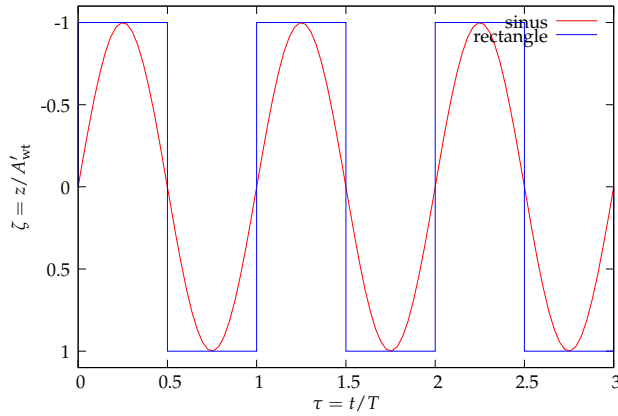


Figure 13.2: Boundary conditions for fluctuating water table experiments.

Table 13.1: Amplitudes A_{wt} and periods T used in the different experiments with a fluctuating water table and initially pre-wetted condition. All experiments were conducted with both rectangular and sinusoidal forcing. Each amplitude was run with both $T = T_1$ and $T = T_2$.

A_{wt} / mm	10	15	20	30	50	100	150
T_1 / h	1	1	1	1	1	1	1
T_2 / h	6	6	6	6	6	6	6

Table 13.2: Nominal amplitudes A_{wt} and periods T used in the different experiments with a fluctuating water table and initially dry condition.

type	sinus			rectangle		
A_{wt} / mm	50	50	100	50	50	100
T / h	6	24	6	1	24	6

Parameters of the conducted experiments are given in table 13.2. Experiments with initially dry medium required new sample preparation for every experiment.

Images were acquired with ISO 100, an exposure time of 1/10 s and an aperture of 5.6.

To obtain a better understanding of microscopic processes, one experiment was conducted where the NIR camera was measuring a small cut-out of the Hele-Shaw cell of $20 \text{ mm} \times 16 \text{ mm}$ with microscopic resolution, such that single pores could be resolved. The fluctuating water table experiment conducted before the infiltration experiments was used for this purpose. The NIR camera was installed as described in section 13.2 and settings are given in table 13.4. Boundary conditions were $A_{wt} = 50 \text{ mm}$ and $T = 1 \text{ h}$.

The real water table amplitude as measured in the water table window deviated from the nominal amplitude by typically about 2 mm to 4 mm and was always smaller than the nominal value. These deviations are attributed to the motor accuracy. Since the water table was measured directly in the cell, the measured value is considered more accurate than nominal values. The general shape of the temporal evolution was always correct. Because the measured water table height was used for evaluation, this did not affect the results. In the following, A_{wt} is used to specify the nominal and A'_{wt} to denote real measured amplitudes. It should be kept in mind that real amplitudes were slightly smaller. With the rectangular forcing, due to finite conductivity, the water table did not always instantaneously reach the final value but changed for about 1 mm during the first few measurements.

13.1.2 Evaluation

Data were processed according to section 12.3.4. Sub-REV and microscopic data were used for evaluation, since the focus was on front dynamics and governing processes (see also section 2.3). To compare experiments with different periods and different amplitudes, dimensionless time $\tau := t/T$ and dimensionless length $\zeta := z/A'_{wt}$ are introduced.

In accordance to the definition in section 2.2.1, in all experiments, the z (depth) axis and therefore also the ζ axis point downwards into the ground.

13.1.3 System response

Films of the system response of all experiments can be found in the accompanying DVD. Figure 13.3 shows the sub-REV water content distribution of the experiment with $A = 50 \text{ mm}$ and $T = 1 \text{ h}$ at different times. Two important features are visible: (i) on the downwards half-period, a saturation minimum at the former upper CF edge arises which can be seen as a red region in figure 13.3 (rightmost image), and (ii) after again raising the water table, a small vertical structure is found in the water distribution at the former lower CF edge.

The first effect is thought to stem from the rapid drainage. When first moving the WT upwards, the interface is pushed upwards and then allowed to equilibrate during the time where the boundary condition stays constant. Subsequently the WT is rapidly lowered, which causes quick outflow. Due to the relatively low conductivity further upwards, the water there cannot follow that quick movement. It stays where it is and the connection to the receding water front further below is lost.

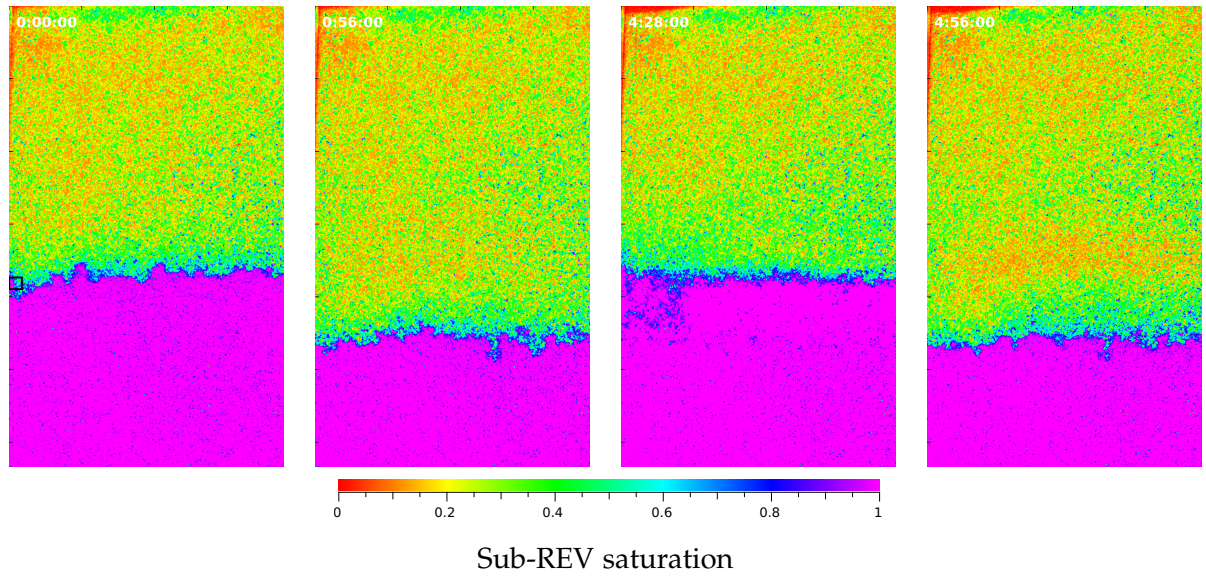


Figure 13.3: Sub-REV saturation $\tilde{\Theta}$ distribution of the moving water table experiment with rectangular forcing of $A_{wt} = 50$ mm and $T = 1$ h at different times. Typical propagated relative measurement error $\sigma_{\tilde{\Theta}}/\tilde{\Theta}$ is 1 %. The black rectangle at the left of the $\tau = 0$ image shows the region covered by the microscopic NIR camera.

Air is entrapped in the following imbibition cycle, when the water front is moving rapidly, which leads to lower water content minima at later periods. This can be seen by the microscopic NIR measurement. Figure 13.4 shows the difference of absorption length between several system states. Negative absorption lengths indicate less water. Since the camera resolved single pores, the drainage of individual pores and the entrapment of air in them can be seen. A relatively large amount of air is already entrapped in the first period and as expected the amount of water does not increase in relation to the saturated sample. In the subsequent periods, air is entrapped in additional pores. A few pores which were previously empty are also filled again as seen by positive absorption length differences. Here, the air could escape because during the drainage a continuous air path was again opened.

Note that these features cannot be described by Richards' equation since, first of all, it assumes local equilibrium. Hence, the continuity of water cannot be lost. Second, the water minimum cannot be represented since the degenerate multiphase regime does not allow entrapped air bubbles but assumes the air phase to be continuous. The minimum would not persist since gradients of the matric potential would lead to flow ceasing the minimum. Persistent air bubbles could be described by a lower effective amount of fluid, but the air bubbles could not change or interact with the moving fluid.

The second effect stems from the different equilibration times. While the boundary condition is constant, the front has time to equilibrate and the menisci can adjust. In contrast, during the rapid infiltration and drainage, the interface is moved so fast that it has no time to adjust. Once it has passed, water-filled flow channels are wide enough to sustain much higher velocities (see also Roth (2005, sect. 5.5.2)). Therefore, at the turning points of the CF movements the hydraulic state is different to that in the inner part. The importance of non-equilibrium effects is further corroborated below.

Figure 13.5 shows examples of the evolution of the microscopic water distribution which was measured with a temporal resolution of 1 s. In accordance with the observations of Heow (2008), large portions of pore space are drained so fast that the drainage of individual pores cannot be tracked at that time scale. Although a vertical front-like movement can still be seen,

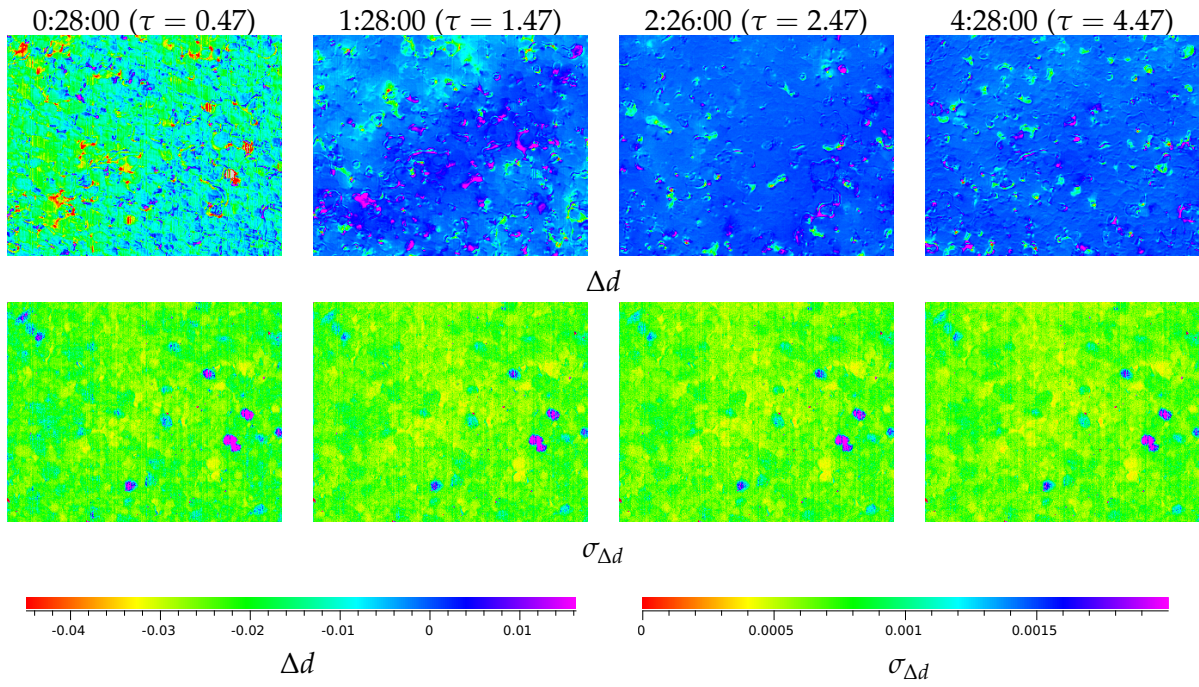


Figure 13.4: Microscopic absorption length difference of the moving water table experiment with rectangular forcing of $A_{wt} = 50$ mm and $T = 1$ h at different times measured with NIRIS. The first image at $\tau = 0.47$ shows the difference to the saturated image, the other images differences to the previous displayed image. A negative d denotes less water (absorption length has decreased). Image dimensions are $20 \text{ mm} \times 16 \text{ mm}$.

no sharp front line is observed, but a heterogeneous boundary area forms. This is an effect of the sample's thickness. The interface movement is very convoluted in 3D and only the average amount of water through the cell's thickness is observed. During the drainage process, liquid connections between grains seem to be preserved. Also during infiltration, a microscopic front is seen which moves upwards. However, even after the front has passed dry spots remain which are filled later in the imbibition process. This is again caused by the 3D effects.

These results show that the front definition issue is much more complicated than the simple reflections in section 2.3 may indicate. Further pore-scale investigations are needed to improve the understanding of these processes.

13.1.4 Capillary fringe edge response

The CF edge was estimated as described in section 12.3.4. The result of a rectangular and a sinusoidal experiment with $A_{wt} = 50$ mm is shown in figure 13.6 for both initially pre-wetted and dry conditions.

The overall shape of the CF response of initially dry and pre-wetted experiments does not differ greatly. The CF edge response is distorted due to the non-linear water characteristic. At the upper reversal point it is flattened due to pinning (see below). Pre-wetted experiments have a higher CF edge. At the border to the dry medium, a large potential gradient is found but the conductivity is very low. Therefore the further upwards movement of the interface into the dry medium is overwhelmingly slow. The water distribution is “cut-off” compared to the water characteristic (figure 13.8 (a)). Conductivity in the pre-wetted experiment is higher due to the residual water, and with the same equilibration time it is much nearer to static equilibrium. When moving the interface upwards, a positive pressure offset is needed to overcome the

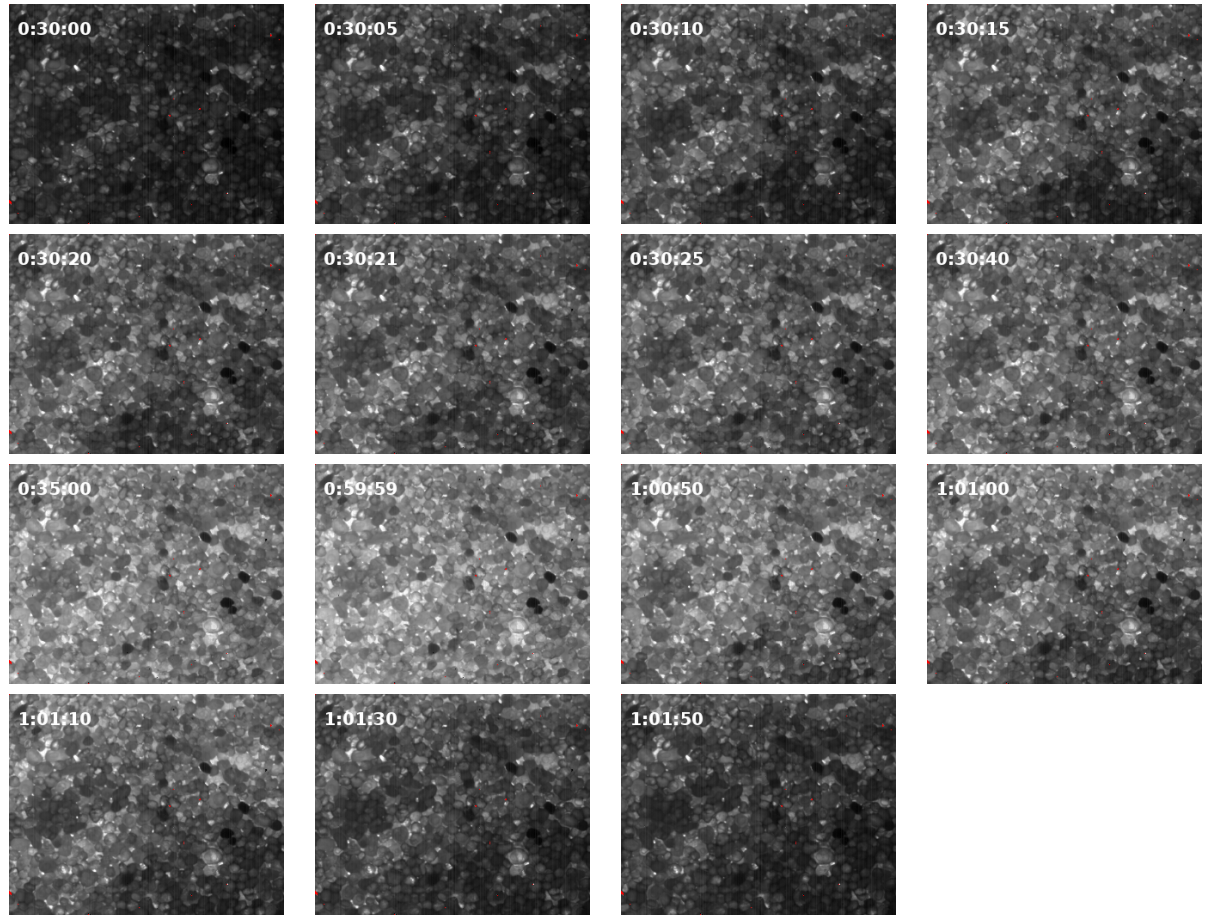


Figure 13.5: Microscopic NIR measurement of changing water distribution during a moving water table experiment with rectangular forcing of $A_{wt} = 50$ mm and $T = 1$ h acquired with 1 s time resolution at 1505 nm. Water is displayed in black, more water leads to less transmitted light due to more absorption. Image dimensions are 20 mm \times 16 mm.

resistance of the low hydraulic conductivity in the upper soil (figure 13.8 (c)). This is a non-equilibrium process. The threshold pressure is smaller for the pre-wetted sample due to the higher conductivity and therefore the CF edge rises higher.

On the other hand, the CF edge of the initially dry sample falls further down in the downwards period. This is explained below after the CF edge dynamics were investigated in more detail.

The half minimum/maximum distance η of CF edge fluctuations for the rectangular forcing was estimated as described in section 12.3.4. The result is shown in figure 13.9. Below a certain threshold in A'_{wt} , η stays constant but positive regardless of the forcing amplitude. Above a certain threshold, however, η linearly raises with A'_{wt} with a slope of 0.718 ± 0.006 . The response remains linear for the whole range of A'_{wt} considered here, which covers amplitudes up to twice the equilibrium CF height. The bend point where the linear increase of η begins is located at $A'_{wt} = (17.3 \pm 0.7)$ mm, the matching $2A'_{wt}$ height corresponds to the pressure offset to drain a pore with a radius of 0.43 mm. This is a reasonable pore size for the medium used here and the offset pressure is likely to correspond to the air entry pressure.

The upper CF boundary had not been risen to the equilibrium height before the water table was lowered again. Hence the water distribution with height $\theta(z)$ does not correspond to

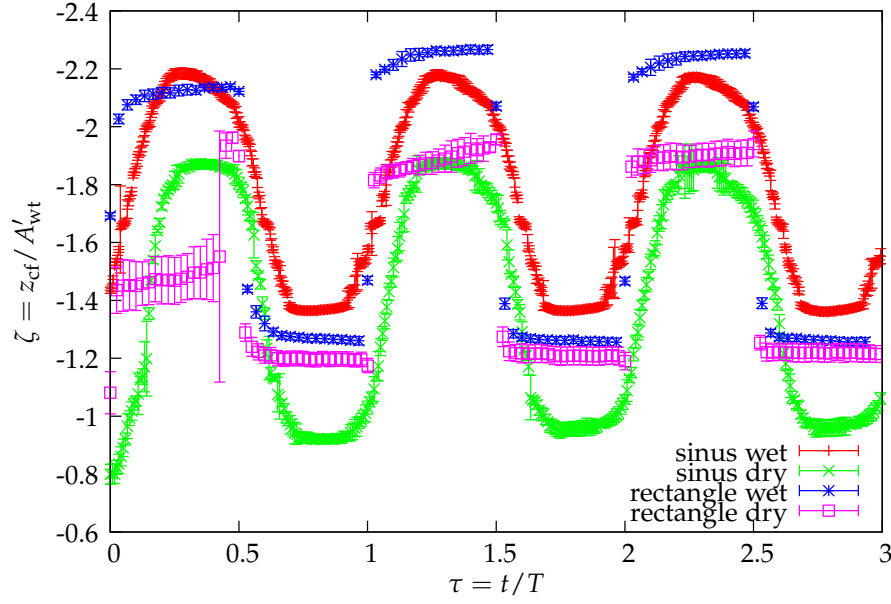


Figure 13.6: Response the the CF edge for pre-wetted and dry experiments. Nominal water table amplitude was $A_{wt} = 50$ mm, period was $T = 6$ h for the sinusoidal and $T = 1$ h for the rectangular experiments. Origin of the ζ axis is the original water table height. The jump of the CF height for the dry rectangular experiment at $\tau = 0.425$ is an artefact which is caused by the threshold and is explained in figure 13.7.

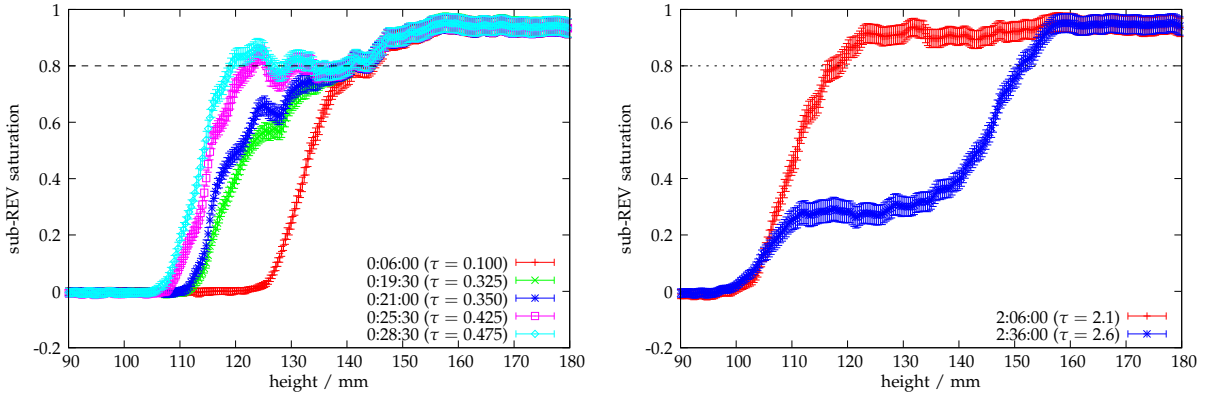


Figure 13.7: Sub-REV saturation profiles which were used for CF edge estimation for the initially dry experiment with $A_{wt} = 50$ mm and $T = 1$ h in the first period (left) and the second period (right). The dashed black horizontal line marks the threshold. Due to the low conductivity in the dry sand, water flow in the first period is slow and the sub-REV saturation of the front raises slowly, which leads to a crossing of the threshold and thereby a jump in the estimated CF edge position. The oversized errorbar in the CF edge location in figure 13.7 at $\tau = 0.425$ stems from the large portion of points touching the threshold at that particular time. Choosing a lower threshold would have made problems in conditions where water is already present above the CF as shown exemplarily for the second period on the right. Too low a threshold interferes with the remaining water when the CF is receding. The sub-REV water content value in that region is even higher for pre-wetted experiments.

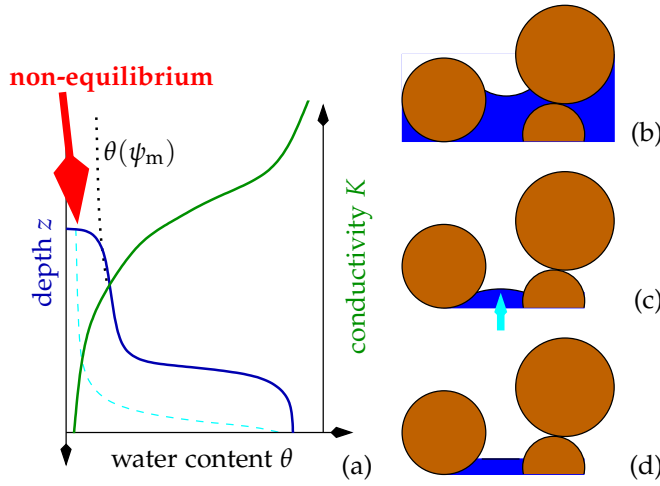


Figure 13.8: Sketch of the non-equilibrium situation found during a fluctuating water table. In equilibrium, the water content distribution corresponds to the water retention curve. However, if the water had not enough time to flow upwards against the resistance of the low conductivity, the real water distribution is “cut-off” (a). Menisci at the water-air interface therefore have a radius larger than the pore radius (b) and a pressure offset is necessary to break them (“pinning”). On the upwards movement, water must be pushed in against the low conductivity and a positive pressure is necessary to overcome the resistance (c). When this positive pressure ceased, menisci are neutral (d) and pinning comes into effect. When the offset has been overcome, water is drained according to the water characteristic started at the cut-off point (small dashed cyan line).

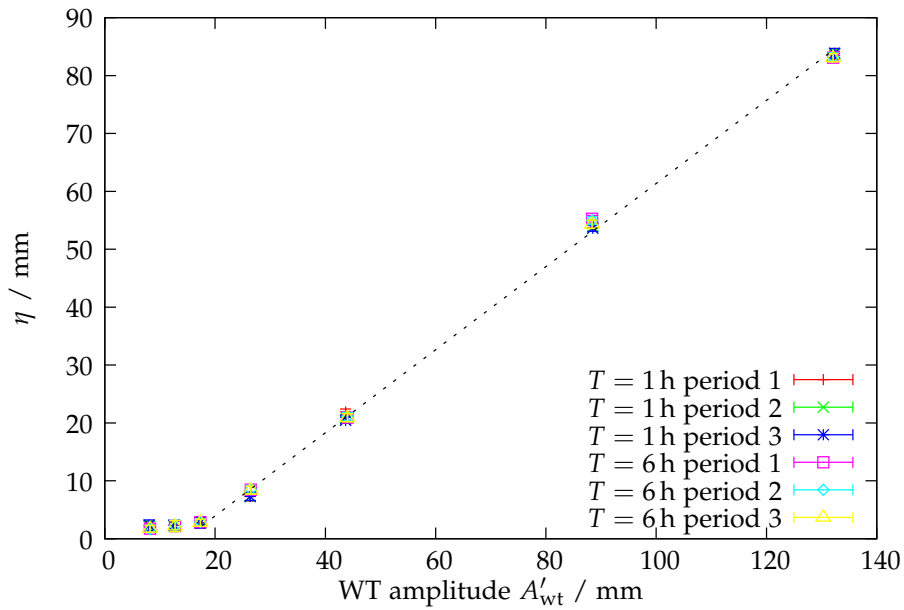


Figure 13.9: Half minimum/maximum distance η of the CF edge for the rectangular forcing and different periods T and amplitudes A_{wt} . The line fitted into the raising points has a slope of 0.718 ± 0.006 . The intersection with the constant points is at $A'_{wt} = (17.3 \pm 0.7) \text{ mm}$, the matching $2A'_{wt}$ height corresponds to a pore radius of 0.43 mm.

the water characteristic but it is “cut off” (figure 13.8 (a)). The time for reaching final static equilibrium is long, especially for a coarse-grained medium. These non-equilibrium menisci (figure 13.8 (b)) would adjust by moving further upwards into the drier medium if the sample would be equilibrated long enough.

The small constant positive distance $\eta = (2.0 \pm 0.2)$ mm found before the linear increase is thought to denote half the small positive pressure required to push the water in against the resistance of the low conductivity in the upper soil (figure 13.8 (c)). When changing the direction of flow, the front moves back until the dynamic positive pressure has ceased (figure 13.8 (d)). Then pinning comes into effect. Here the menisci stay in place until the pressure has dropped low enough that their curvature becomes too small and they are “cracked”. This also leads to the flattening of the CF edge at the upper reversal point for sinus experiments mentioned above. After cracking, water is drained according to the water characteristic, starting at the low cut-off water content (small dashed cyan line).

This also explains the asymmetry in the CF edge response. The downwards amplitude is much lower than the upwards amplitude, because on the downwards half-period the air-entry pressure must be overcome.

A slope of less than one shows that non-equilibrium processes take place. In equilibrium, a slope of 1 would be expected, since the CF should follow the the whole A'_{wt} distance of forcing regardless of the amplitude. The non-equilibrium effects described above lead to a depression of δ_{cf} compared to A'_{wt} . On the upwards movement, a part of the forcing amplitude is necessary to build up the positive pressure required to move the interface into the low-conductive medium. For a higher amplitude, the velocity of the forcing increases and therefore the CF is compressed further, which leads to a larger depression. On the downwards movement, the pinning of the menisci and the drainage along the shifted water characteristic (figure 13.8 (a)) additionally depress the response. With all these effects even a slight decrease of the slope with raising amplitude may be expected. When looking closely at figure 13.9, a very slight decrease of the slope can really be seen but it is not significant.

With these ingredients one can now readily understand that the CF edge of the initially dry sample falls further down in the downwards period, as noticed above (figure 13.6). Due to the much lower conductivity in the initially dry medium, a much higher positive pressure offset is required to push the interface upwards. Hence, on the uppermost position, the cut-off in $\theta(z)$ as sketched in figure 13.8 (a) is found at a lower position (larger depth) and it is more distinct. When reverting the direction of the forcing, the backwards movement starts at a lower position (thin dashed cyan line in figure 13.8 (a) is moved downwards) and consequently, the CF will be lower than in the wet case, where the start point was at a higher position.

The difference between dry and wet becomes smaller for larger amplitudes, since according to the water characteristic the water content further upwards is much smaller. Therefore the difference in conductivity between dry and wet conditions becomes smaller, since the forcing is fast enough that the system has no time to equilibrate and thereby adjust the water content before the front reaches the end point during the fast infiltration.

The thickness of the CF δ_{cf} is defined as the difference between the CF edge and the water table. Figure 13.10 shows the thickness of the CF for sinus experiments and different amplitudes. For low amplitudes which are small compared to the CF thickness, the thickness follows the forcing without significant distortions. The CF edge stays relatively fixed and the thickness change is dominated by the harmonic water table fluctuation. As seen previously, the water table fluctuation is also larger than the responding CF edge fluctuation (see e. g. figure 13.6). If the forcing gets larger than the air entry pressure, however, distortions in the CF response are seen which become more distinct if the forcing gets larger. These are caused by the non-linear dynamics of the water movement. When looking at the water retention curve, it is also obvious

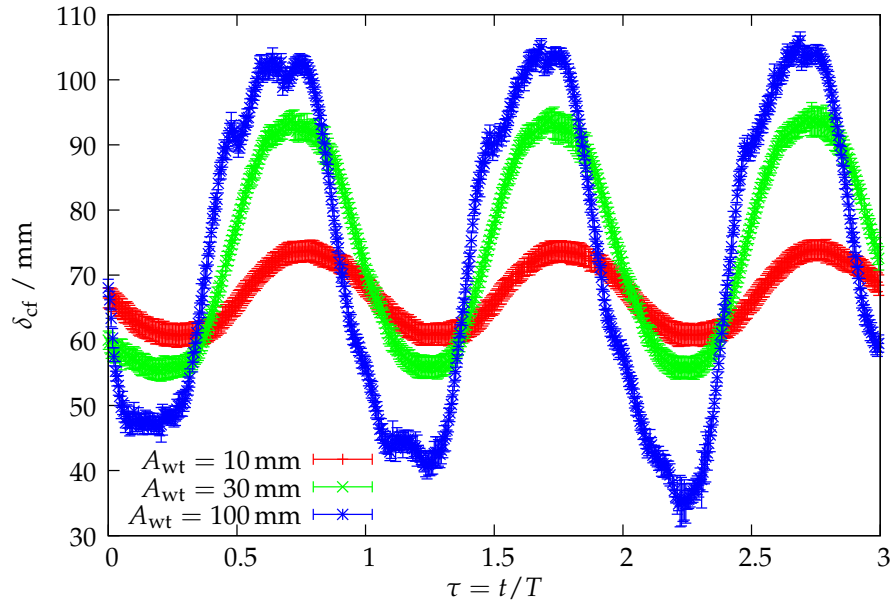


Figure 13.10: CF thickness for the sinusoidal forcing of the initially wet sample, $T = 6$ h and different amplitudes.

that non-linearity becomes more important if the forcing increases. The increase is expected to stop once the forcing has explored all of the water retention curve, and dynamics will be highly non-linear.

For $A_{wt} = 100$ mm, wiggles can be seen in δ_{cf} . These are caused by the interplay of the harmonic WT and non-harmonic CF fluctuations as shown in figure 13.11. At first, with raising WT, the CF edge follows with a delay and the thickness decreases. After a certain pressure offset is overcome, the CF edge follows at constant distance since a dynamic balance between moving interface and positive pressure is established. Here, wiggles are caused by slight velocity differences caused by heterogeneities. When the forcing is reverted, the CF edge is pinned and the thickness increases again. When the CF edge also starts decreasing the increase slows down. With a deflection in the CF response, a temporal depression is found until the decrease of the forcing becomes so steep that it overcomes the speed of the falling CF edge which is limited due to conductivity. Later on, again a dynamic balance adjusts until the forcing is reverted. The interface stays constant until the dynamic positive pressure is established and the cycle restarts. Only very small wiggles are seen during the upwards movement of the WT compared to the downwards movement due to the asymmetry between imbibition and drainage.

To investigate the influence of hysteresis, water table location was plotted versus CF edge location. Results for $A_{wt} = 100$ mm are shown in figure 13.12. The different periods practically overlap nearly everywhere. This indicates that the system stays in the same branches of the hysteresis loop. Due to the many loops which were passed previously, the corresponding drainage and imbibition loops, respectively, overlay. This is also the state expected in a natural system. Only at the upper turning point, differences are seen. These are attributed to entrapped air, which is also seen in the microscopic measurements (figure 13.4). The differences are larger for the sinusoidal forcing, since the system has more time to equilibrate during the phases of constant water table in the rectangular forcing.

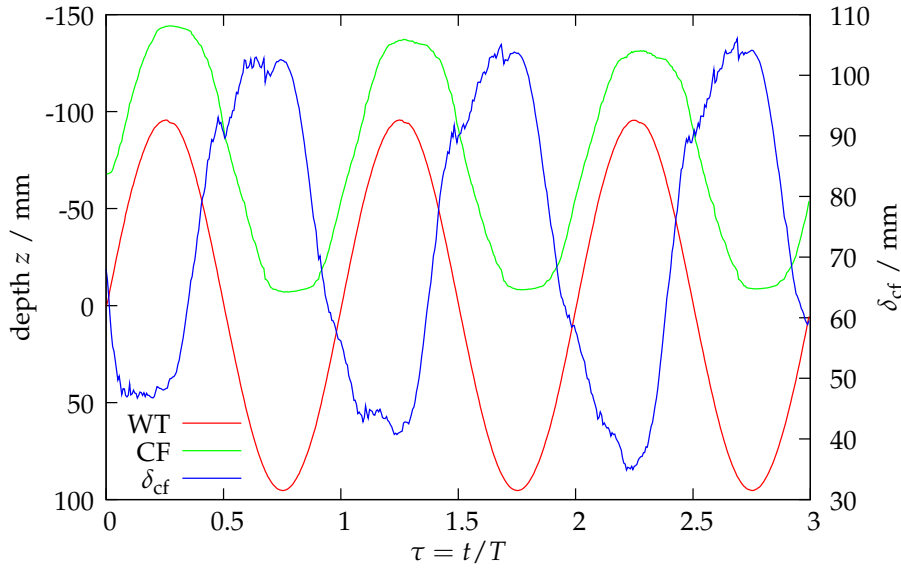


Figure 13.11: Illustration of the source of the “wiggles” in the CF thickness for $A_{wt} = 100$ mm. For more explanations see text.

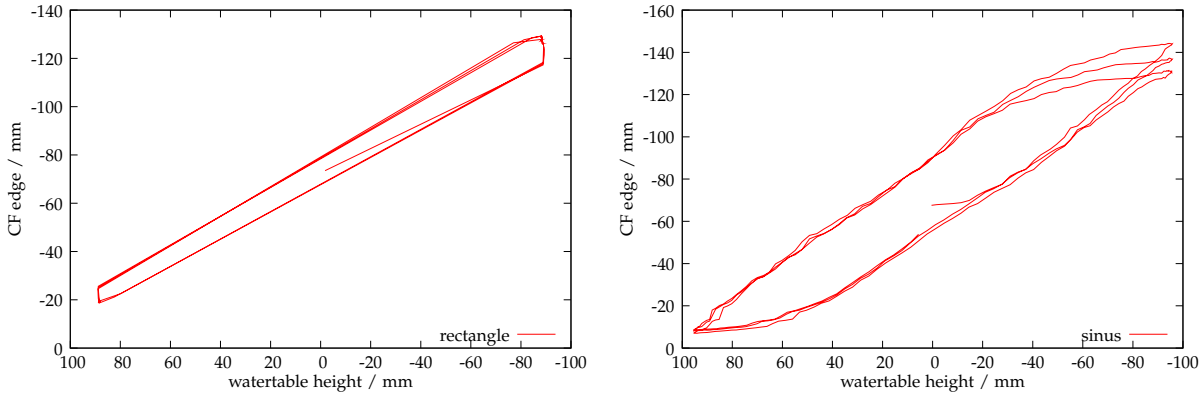


Figure 13.12: Water table location versus CF edge location for the pre-wetted $A_{wt} = 100$ mm measurements.

13.2 Dynamics of fluctuating infiltration

13.2.1 Measurement protocol

The sample was filled according to section 12.2. Porosity was estimated gravimetrically to $\phi = 0.41 \pm 0.05$. Additionally, a layer of (22 ± 1) mm thickness with fine sand (grain size less than 0.25 mm) with a lower saturated hydraulic conductivity was put below the regular medium. This was done to avoid instantaneous equilibration of the water table during infiltration, which is usually not given in natural systems.

Images of the dry sample were taken and the sample was then slowly saturated from below with the water table raising at 30 mm/h. Then, after taking images of the wet sample, the water table was again slowly decreased at 30 mm/h to a middle position. Afterwards the sample was equilibrated for a few hours. A fluctuating water table experiment with $A_{wt} = 50$ mm, $T = 1$ h and 10 periods was run to bring the sample into a more natural initial state. Then infiltration experiments with different fluxes were conducted. Fluxes are given in table 13.3. Temperature

Table 13.3: Fluxes j and period times T of the different infiltration experiments.

no.	j / mm/s	T / h	comment
1	0.48 ± 0.02	0.5	
2	0.053 ± 0.004	1.5	
3	0.106 ± 0.009	1	
4	0.0091 ± 0.0008	7	
5	0.21 ± 0.02	0.5	
6	0.31 ± 0.03	0.5	
7	0.47 ± 0.04	0.5	high resolution
8	0.053 ± 0.004	0.5	high resolution
9	0.42 ± 0.04	0.5	

Table 13.4: Camera settings for the NIR measurements (exposure time t_{exp} at different wavelengths λ and amplifier settings).

λ / nm	t_{exp} / μs	Other settings	
1075	100	V_{in}	2700
1450	250	V_{ref}	1250
1505	90	V_{detcom}	3500
1555	50	T	270 K

was constant within 1 °C during each measurement.

The glass plates of the Hele-Shaw cell were not perfectly parallel, which led to an error in cell thickness. The uncertainty was roughly estimated to $\pm 1/4$ mm, which is already a relative error of 8%. Therefore, also the area estimation of the medium and thus the flux had a relatively large uncertainty.

The NIR camera was used for microscopic measurements of a cut-out of 20 mm \times 16 mm. The viewed area had to be placed relatively near to the left cell border, since the visual camera was required to also measure the part of the cell covered by the NIR camera and the optical setup for the NIR measurement should not interfere the field of vision of the visual camera. The distance between the left image corner of the microscopic measurement and the cell edge was about 10 mm, which is way above the correlation length of the medium. Therefore the influence of the border was still considered neglectable. The macroscopic images do not show any qualitative difference in the dynamics in the region used for the NIR camera to the area in the centre of the cell.

Naturally the NIR light source was used for this experiment instead of the VIS light source which was used for the other fluctuating water table experiments. NIR camera settings are given in table 13.4. Time resolution was 30 s for spectroscopic and 1 s for fast measurements at one wavelength, respectively. For the macroscopic visual camera, ISO 100, an exposure time of 1/8 s and an aperture of 9 was used, time resolution was 2 min for measurement 4 with $j = (0.0091 \pm 0.0008)$ mm/s and 30 s for the other measurements.

13.2.2 System response

Changing water saturation for two different flow rates is shown in figure 13.13. For the high flux $j = (0.48 \pm 0.02)$ mm/s, the infiltration front comes down fast and hits the CF. At the CF edge, air is entrapped. The CF raises because the infiltration flux is larger than the saturated conductivity of the fine sand layer at the bottom of the cell. After the infiltration stopped, the CF edge moves downwards very slowly and does not reach its initial position before the next

period of infiltration starts. Here, the infiltrated water is still ponded at the low-conductive layer. The less water is ponded, the slower the outflow through the layer below since the pressure gradient decreases.

The low flux $j = (0.0091 \pm 0.0008)$ mm/s is similar, but water content of the moving front is much lower. Nevertheless, a water content minimum forms which is much more distinct than for the high flux. The CF edge still goes further down although infiltration started, since resulting from the previous infiltration it is still higher than the equilibrium position.

Figure 13.14 shows vertical water saturation profiles of a small band along the cell. It clearly shows the saturation minimum for all fluxes. This minimum is caused by multi-phase phenomena. The constant imposed flux j leads to a constant flux at all z above the CF edge due to mass conservation. The entrapped air forces the system to stay in the state of water content gradients, since the air cannot move out. Therefore the air pressure in the entrapped air bubbles increases and thereby becomes higher than ambient air pressure. This ceases matric potential gradients and allows water content gradients to persist. Note that this complex situation cannot be described with Richards' equation (2.19), because that equation assumes a continuous air phase at constant pressure. In that description opposing water content gradients would lead to an opposing flux, which contradicts the constant imposed infiltration flux.

The lower the flux, the more pronounced the minimum. The same effect was also observed by Buchner (2009). It disagrees with the immediate intuition that a higher flux would give the air less time to escape. Also, for lower saturation as found at a lower flux, a higher air saturation is available to transport the abundant air. The key to understand this effect is the heterogeneity of the CF and the infiltration flux. Due to heterogeneity, the front hits the CF first at a few points. As soon as a hydraulic connection between the front and the CF is established, high-conductive bridges form and the pressure of the infiltration front decreases drastically since the total resistance dropped. As a positive pressure is necessary to progress the interface, the air in the remaining regions between the front edge and the CF persists. The higher the infiltration rate, the lower the heterogeneity of the front (figure 13.13), because at a higher flux, the interface is pushed so fast into the medium that conductivity differences due to heterogeneity are not as important as at lower fluxes. This leads to a larger portion of entrapped air for a lower infiltration flux.

13.3 Dynamics of fingered infiltration

Infiltration into coarse dry porous media typically leads to unstable fingered flow. The interaction of such an infiltration with the capillary fringe was investigated in an extra experiment. Here, a thin (approx. 3 mm) layer of fine sand was added at the top of the cell for flow homogenisation. Porosity of the regular coarse-grained sand medium was estimated gravimetrically to $\phi = 0.38 \pm 0.05$. After taking dry images, the water table was slowly put into a middle position and equilibrated for about 15 h. Water was infiltrated with $j = (0.031 \pm 0.003)$ mm/s and $T = 2$ h. Three fingers formed sequentially. Since it is not pre-determined where a finger is initiated, the view of the microscopic NIR camera had to be adopted "on the fly" to optimally catch the finger.

The result is shown in figure 13.15 macroscopically and 13.16 microscopically. The first macroscopic image nicely shows the saturation overshoot in the finger tip which is necessary to initiate the finger (Rezanezhad et al., 2006; Roth, 2005, sect. 5.5.2). The microscopic measurement shows that a thin air layer along the CF boundary hinders the finger to directly hit the CF (figure 13.16, second sub-frame). The air is partly moved to the side and, due to the heterogeneity of the finger tip, a first connection builds in the left finger edge which hits the CF first. Here, the same effect as for the infiltration front takes place. After a first conductive bridge

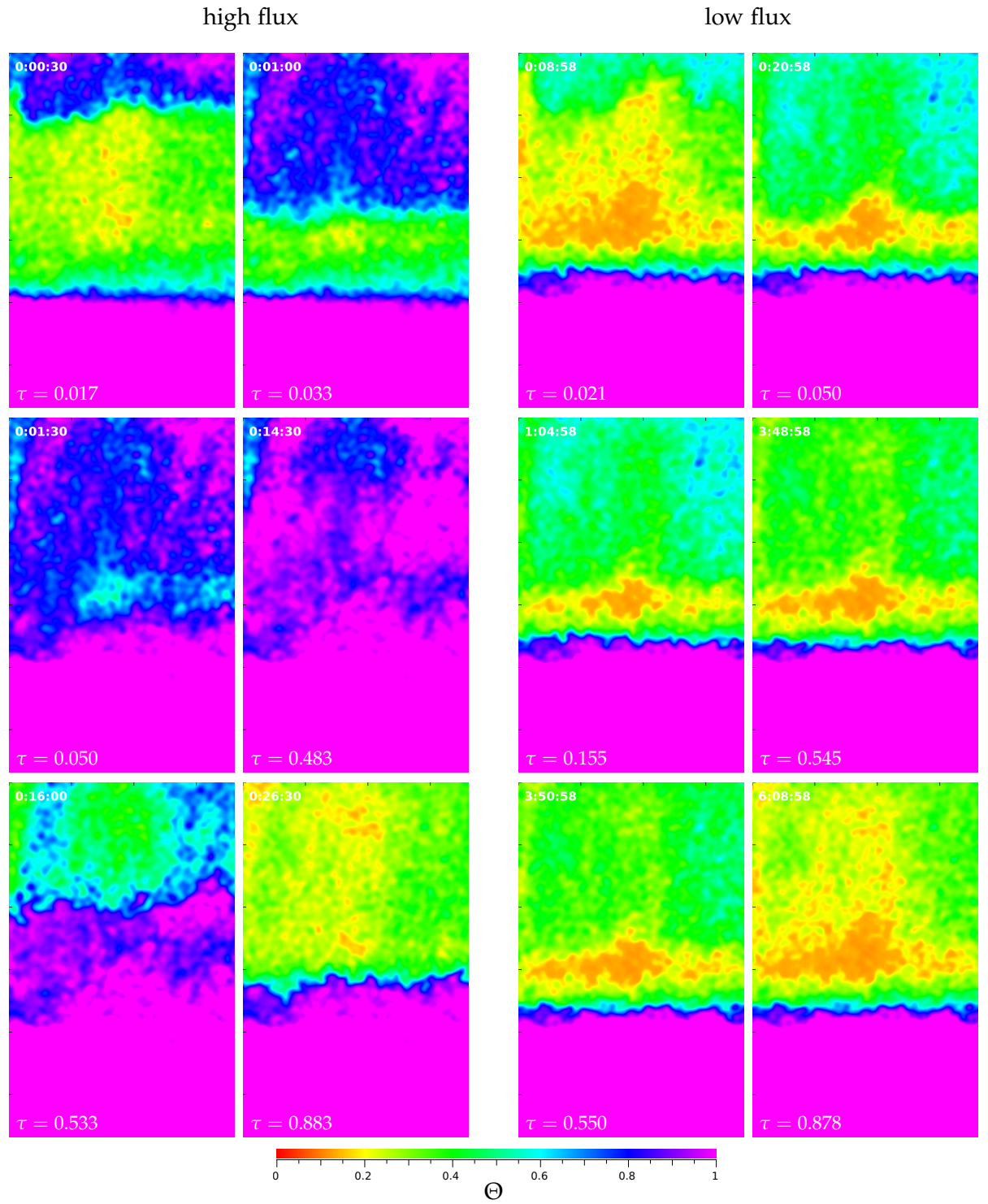


Figure 13.13: Response of the CF to fluctuating infiltration. The left six images show experiment 1 with $T = 0.5$ h, the right six images experiment 4 with $T = 7$ h. Typical propagated relative measurement error σ_{Θ}/Θ is 0.2 %. For more explanations see text.

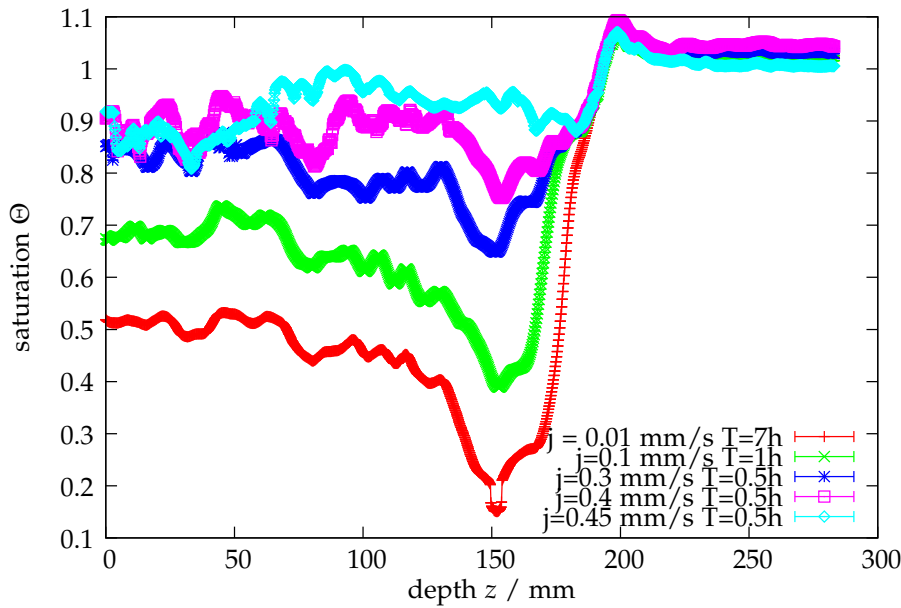


Figure 13.14: Vertical water saturation profile of a 49 mm width band in the middle of the cell directly before the end of the infiltration for different fluxes.

is formed, the pressure decreases (the saturation overshoot in figure 13.15 declines, compare first and second frame) due to the dropped resistance and the interface movement is slowed down dramatically. At longer time scales, the amount of water in the the contact area slowly increases and the contact of the finger with the CF gets closer. During that process, the finger is widening. Part of the air is entrapped between the finger and the CF when water flows around it through more conductive parts.

This shows again the importance of the air phase and its coupling to the water phase near the CF. Even though with fingered flow, air saturation around the finger in the vadose zone is very high, air is entrapped at the edge of the CF due to the dynamic interplay of water and air.

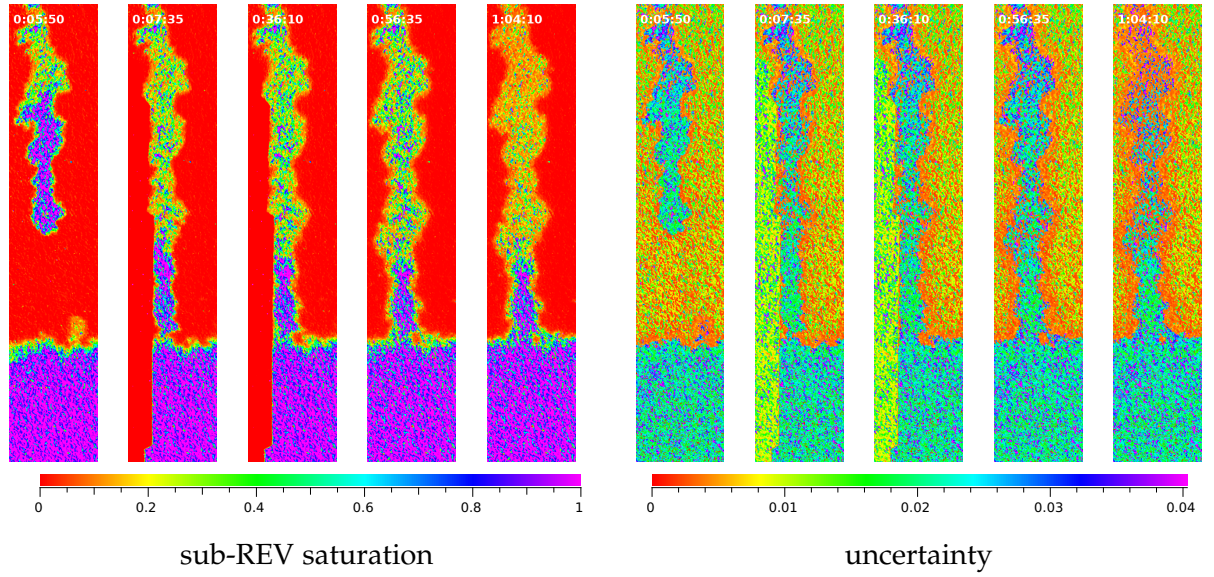


Figure 13.15: Water distribution of unstable infiltration into the capillary fringe as measured with visual light transmission. Experiment time is given at the upper left corner of the images. In images 2 and 3, the red box at the left lower image border is a part of the filter wheel of the NIR camera setup which was moved to optimally catch the finger and thereby interfered with the visual imaging. Typical propagated relative measurement uncertainty was 1%, very few outliers which stem from non-matching grain structure have an error larger than its value. The corresponding absolute uncertainty images are shown on the right.

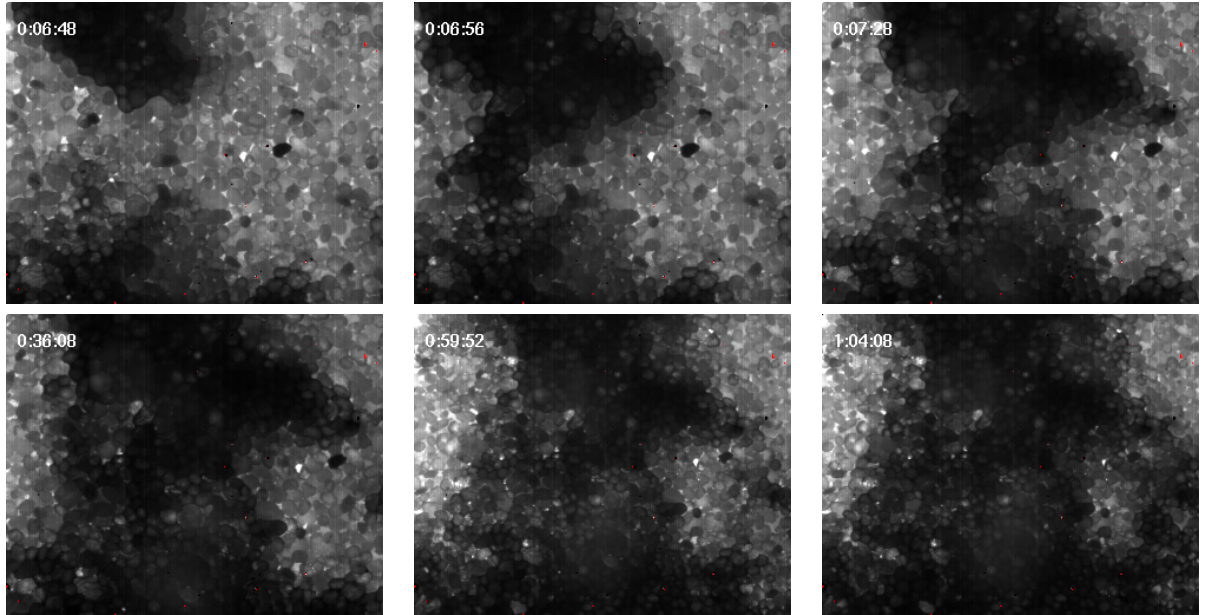


Figure 13.16: Microscopic images of transmitted normalised intensity at 1450 nm during unstable fingered infiltration into the capillary fringe. Water is displayed in black, more water leads to less transmitted light due to more absorption. Note that the view was changed slightly after the second and after the fourth image to optimally follow the dynamics.

14 Conclusions

The response of the CF to fluctuating infiltration, both from a homogeneous infiltration front and unstable fingered flow, and to a fluctuating water table were investigated experimentally in a Hele-Shaw cell with a quasi-2D sand medium using transmission of visual light and NIR imaging spectroscopy. This study gave insight into the fundamental physical processes governing water movement in the capillary fringe under transient boundary conditions.

A water content minimum directly above the CF was observed in infiltration experiments which is more distinct for lower flux. It is caused by the coupled multi-phase character, the positive pressure necessary to progress the front and the heterogeneity of front and CF. In unstable fingered flow, the same effect leads to air entrapment and a widening of the finger directly at the edge of the CF.

The response of the CF to harmonic forcing becomes non-harmonic for forcing amplitudes larger than a height corresponding to half the air entry pressure. Below the air entry pressure, the amplitude of the resulting CF edge fluctuations is small and independent of the forcing amplitude. Above, for the range of amplitudes considered here it increases linearly with the forcing amplitude with a slope of less than 1. These effects are caused by the non-equilibrium imbibition and drainage processes and the corresponding movements of the interfaces which can only be understood at a sub-REV scale. On rapid drainage, a saturation minimum forms above the former CF location since the connection to the receding water front is lost. The progressive air entrapment during successive imbibition and drainage was confirmed by microscopic measurements. The comparison of dynamics in initially dry with initially wetted media showed that the general shape of the CF response is very similar. In the initially dry medium, during infiltration the CF edge is lower due to the much lower conductivity in the dry soil above, and it moves further down on drainage since the corresponding point of the water characteristic starts at the lower position.

The majority of these processes cannot be described with a traditional Richards model, and a multi-phase description coupled with a dynamic non-equilibrium model will be necessary. A major challenge will be the representation of the dynamics at a sub-REV scale.

Summary and sidelook

This work gave insight into the water movement at the two contrasting boundaries of the vadose zone in soils, namely the evaporation at the soil-atmosphere boundary at the top and the capillary fringe at the bottom. Both have a considerable influence on the vadose zone, since they couple it to the surrounding and force its dynamics. While evaporation at the upper boundary is characterised by the properties of the boundary layer and their interplay with the medium which leads to two different regimes in the dynamics, the lower boundary gives rise to non-equilibrium processes and coupled multi-phase phenomena.

Evaporation from an REV-size sample with water vapour partial pressure and temperature as prescribed atmospheric boundary conditions was successfully modelled with a diffusive boundary layer of fixed thickness. Two regimes could be identified, which are caused by the coupling of that layer with the highly non-linear hydraulics of the vadose zone: (i) a boundary-layer dominated regime with a constant evaporative flux which only depends on the boundary layer, (ii) a hydraulically dominated regime where the flux is governed by soil hydraulic properties and which is characterised by a rapid decrease of flux with ongoing drying. These regimes are consistent with experimental findings in the literature. The model was also used to invert soil hydraulic properties from an evaporation experiment. A combination of a statistical Monte-Carlo and a gradient based Levenberg-Marquardt method was necessary for successful inversion. The switching between regimes could be forced by convenient boundary conditions.

In contrast, the capillary fringe has a totally different character. This is already reflected in the much larger vertical extent of the boundary. It has totally different properties than the small diffusive soil-atmosphere boundary layer. The transient forcing and the highly coupled flow caused by the high water contents lead to non-equilibrium processes which cannot be described with the traditional equations. The air phase which is typically neglected in the traditional description plays a crucial role and leads to qualitatively new phenomena like a saturation minimum above the capillary fringe during infiltration and the widening of fingered flow. Sub-REV processes need to be considered to understand the corresponding processes.

Two different boundaries in natural systems were examined in this thesis, one between a solid porous medium and a fluid, and one inside solid porous media between multi-phase and single-phase flow. The third type of boundary which was not within the scope of this thesis is a fluid-fluid interface. Its most prominent representative is the air-sea boundary. Since 71 % of earth's surface is covered by oceans, the boundary also plays a major role in the water cycle, the trace gas budget and for the climate.

The situation for air-sea gas exchange is somewhat similar to the evaporation in regime I, where the boundary layer controls the water flux and the soil hydraulic properties do not influence the evaporation flux. Air-sea gas exchange is controlled by the diffusive boundary layer at the air-sea interface (Jähne and Haußecker, 1998). In the bulk of atmosphere and ocean, turbulent convection governs the transport, which is very efficient on large length scales. Since eddies cannot penetrate the water-air phase boundary, they become successively smaller as the surface is approached. A thin layer at the boundary therefore is dominated by diffusive transport. This thin boundary layer has a much higher transfer resistance than the bulk water and air and is the "bottle neck" for water-air gas exchange. Consequently, to understand and quantify the exchange of atmospheric gases with the ocean, it is crucial to understand the physical processes at the air-sea interface. The similarity with the soil-atmosphere boundary is also manifest in the identical description of the boundary as a small diffusive layer, which in

air-sea gas exchange is typically written as

$$j = k(v_w - v_a) , \quad (14.1)$$

where k is called *transfer velocity*. Note the similarity to equation (8.1) on page 59.

However, there are many important differences to the soil-atmosphere boundary. First of all, the water-air interface is a free surface while the soil surface is rigid. This leads to changes in the boundary layer at the water-air interface and therefore to changes in k depending on the boundary conditions, for example wind speed and waves. Additional effects are for example surface renewal if the surface is taken away by breaking waves, and supplementary transfer by sea spray and white caps. All this makes the interface as such much more complicated. One important point in air-sea gas exchange is therefore to understand the relation of k with all these parameters and processes. In soils, only changing water content may affect the layer thickness. Since the experiment reported in this thesis showed an excellent agreement of the model with the data, this is considered to be of minor importance at least in the experimental setup considered.

Second, for soils it is not possible to relate a boundary layer resistance or soil hydraulic properties obtained from one location to some other location even if the boundary conditions are exactly equal, since the soil structure (geometry) is different (no ergodicity). The structure and properties of the soil must be estimated separately for each location, which is a major problem in hydrology. In contrast, in systems with fully developed turbulence as in atmosphere or ocean, ergodicity is given and it is possible to measure the transfer velocity at one location and then use it for another location with the same boundary conditions. This basically enables to estimate the dependence of the transfer velocity on different parameters and then apply the relation on the ocean as such, e.g. in climate models. Of course there are practical difficulties. For example, surface films and surfactants have an influence on the transfer velocity (McKenna and McGillis, 2004; Frew et al., 1990; Bock et al., 1999), therefore surfactant species and concentrations must be known. However, the relationship between onshore-offshore surfactant gradients and air-sea gas exchange is still unknown.

Finally, regime II of evaporation has no equivalent in air-sea gas exchange, since the latter is always controlled by the water-air boundary layer. This switching between regimes is an interesting feature of soils. It results from the high dynamics and non-linearity of soil hydraulic properties and their influence on the processes. Depending on boundary conditions and the hydraulic state of the system, the governing processes may be qualitatively different.

Appendix

A Statistics

A.1 Random variables

Statistics is important in physics because measurements of physical systems are statistic processes. Every measurement results in a slightly different value. Thus the measurand x is a random variable and the measurement is characterised by a *probability density function* (PDF) $p(x)$ rather than a single value. The integral $\int_x^{x+\Delta x} p(\xi) d\xi$ specifies the probability that the measured values lies within the interval $[x, x + \Delta x]$.

As digital computers can only handle discrete numbers, discrete random variables are needed for all computer-based data processing. Discrete random variables can only have a finite number N of values g_i ($i = 1, \dots, N$) which are measured with the probability p_i . In the following, the continuous expressions will be displayed on the left and the discrete ones on the right.

Per definition the probability to measure any value is 1, therefore the PDF must satisfy the condition

$$\int_{-\infty}^{+\infty} p(\xi) d\xi = 1, \quad \sum_{i=1}^N p_i = 1. \quad (\text{A.1})$$

The basic parameters which describe a random variable are its *mean* μ , also called expected value,

$$\mu = E(x) \quad (\text{A.2})$$

and its moments μ_m of order m ,

$$\mu_m = E((x - \mu)^m), \quad (\text{A.3})$$

where

$$E(g(x)) = \int_{-\infty}^{+\infty} g(\xi) p(\xi) d\xi, \quad E(g(x)) = \sum_{i=1}^N g(x_i) p_i \quad (\text{A.4})$$

with any function g . The moment of order $m = 2$ is called *variance*

$$\sigma^2 = \mu_2 = E((x - \mu)^2) \quad (\text{A.5})$$

and is a measure for the deviation of the of the measured values from the mean value.

The mean can also be estimated using any large number of measurements without explicitly knowing the PDF by

$$\mu = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i. \quad (\text{A.6})$$

In practice it is impossible to have $N = \infty$. The estimation with a finite number of measurements N has an uncertainty which depends on N and the form of the PDF.

A.2 Multi-dimensional random variables

In almost all cases, many quantities are measured in an experiment. If images are considered, there are even millions of measurements, as every pixel is a separate variable. Therefore statistics of multiple random variables is needed.

In general, P random variables are described by a common PDF $p(x_1, \dots, x_P)$. However, if the random variables are statistically independent, the PDF can be written as a product of single PDFs:

$$p(\mathbf{x}) = \prod_{i=1}^P p_{x_i}(x_i) \quad \Leftrightarrow \quad x_i \text{ independent for all } i = 1, \dots, P \quad (\text{A.7})$$

The *covariance* describes the relation of the fluctuations of two random variables. It is defined in extension to the variance (A.5):

$$\sigma_{ij} = [\text{cov}(\mathbf{x})]_{ij} = E((x_i - \mu_i)(x_j - \mu_j)) \quad (\text{A.8})$$

Thus, $\sigma_{ii} = \sigma_i^2$. For P random variables σ_{ij} forms a symmetric $P \times P$ matrix. The correlation coefficient

$$c_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} \text{ with } |c_{ij}| \leq 1 \quad (\text{A.9})$$

relates the covariance with the variances and is a measure for the linear correlation of x_i and x_j . If $c_{ij} = 1$, all points (x_i, x_j) are on a straight line with probability 1. For uncorrelated random variables $c_{ij} = 0$ for $i \neq j$. On contrary, $c_{ij} = 0$ does not necessarily mean that the x_i are independent.

As σ_{ij} is a symmetric matrix, the elementary theorem of linear algebra states that it can always be diagonalised by choosing a set of linear combinations e_i of the x_i as new random variables and e_i are the eigenvalues of σ_{ij} .

A.3 Functions of random variables

Normally, measured values are processed further to obtain the required results. As the measured values are random variables, the result of such a calculation is also a random variable and its PDF must be known in order to know its statistical properties.

The most simple case is a function $f : \mathbb{R} \rightarrow \mathbb{R}$ which transforms one random variable into another: $y = f(x)$. If f is a linear function, the PDF of y is the one of x scaled with the inverse of the factor of proportionality. For a general differentiable function f , the PDF of y is given by

$$p_y(y) = \sum_{s=0}^S \frac{p(x_s)}{|f'(x_s)|} \quad (\text{A.10})$$

where x_s denotes the S leaves of the inverse function of $y = f(x)$.

For non-linear functions f , the mean of y cannot be directly calculated with $E(y) = f(E(x))$, because $E(y) = E(f(x)) = \int_{-\infty}^{+\infty} f(x) p_x(x) dx \neq f(E(x))$. With a Taylor expansion of $f(x)$ to the order of 2, one obtains

$$\mu_y = f(\mu_x) + \frac{f''(\mu_x) \sigma_x^2}{2} + \dots \quad (\text{A.11})$$

This shows that $\mu_y \approx f(\mu_x)$ is only valid if the curvature of f and the variance of x is small, i. e. f can be approximated by a linear function in the interval $[\mu - 3\sigma, \mu + 3\sigma]$.

The variance of y is given by

$$\sigma_y^2 = |p'(\mu_x)|^2 \sigma_x^2 + \dots \quad (\text{A.12})$$

For multi-dimensional random variables, consider a P -dimensional random vector $\mathbf{x} \in \mathbb{R}^P$ and a function $f : \mathbb{R}^P \rightarrow \mathbb{R}^P$. By expanding f to a Taylor series, one obtains

$$E(y_p) \approx f(E(x_p)) + \sum_{i=1}^P \sum_{j=1}^P \mathbf{H}(f_p(E(\mathbf{x})))_{ij} \text{cov}(\mathbf{x})_{ij} , \quad (\text{A.13})$$

where $\mathbf{H}(f_p)_{ij} = \frac{\partial^2 f_p}{\partial x_i \partial x_j}$ is the symmetric Hessian matrix of f_p (assuming that the f_p are continuously differentiable), and

$$\text{cov}(\mathbf{y}) = \mathbf{J} \text{cov}(\mathbf{x}) \mathbf{J}^T , \quad (\text{A.14})$$

where $\mathbf{J}_{ij} = \frac{\partial f_i}{\partial x_j}$ is the Jacobian matrix of f .

B Fourier transform

It is often desirable to decompose a signal into periodical patterns. A periodical signal is described by a wave vector \mathbf{k} with length $|\mathbf{k}| = k = 1/\lambda$ as well as the phase ϕ (figure B.1). k is the wave number and λ the wave length. The direction of \mathbf{k} points orthogonal to lines of constant values. However, it is more convenient to use complex numbers. Then, any periodic signal can be described as $\hat{g} \exp(2\pi i \mathbf{k} \cdot \mathbf{x})$ with a complex \hat{g} . Then, only one complex variable is required to describe amplitude and phase.

In the discrete case, the wave number is an integer value ν . It states how many wave lengths fit into an interval of length N .

The decomposition of an arbitrary signal into its periodic patterns can be viewed as a basis transform. The vector space of periodic patterns is called Fourier space.

For a discrete signal $\mathbf{g} = (g_0, \dots, g_{N-1}) \in \mathbb{R}^N$, the orthonormal basis vectors are

$$\mathbf{b}_\nu = \frac{1}{\sqrt{N}} \left(w_N^0, w_N^\nu, \dots, w_N^{(N-1)\nu} \right) \quad (\text{B.1})$$

with

$$w_N := \exp\left(\frac{2\pi i}{N}\right). \quad (\text{B.2})$$

The discrete Fourier transform (DFT) is defined as

$$\hat{g}_\nu = \mathbf{b}_\nu \cdot \mathbf{g} = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} g_n \exp\left(-\frac{2\pi i n \nu}{N}\right), \quad 0 \leq \nu < N, \quad (\text{B.3})$$

where \cdot denotes the standard scalar product. The inverse transform is given by

$$g_n = \mathbf{b}_n^{-1} \cdot \hat{\mathbf{g}} = \frac{1}{\sqrt{N}} \sum_{\nu=0}^{N-1} \hat{g}_\nu \exp\left(+\frac{2\pi i n \nu}{N}\right), \quad 0 \leq n < N. \quad (\text{B.4})$$

The relation is also written as $g \circ \bullet \hat{g}$

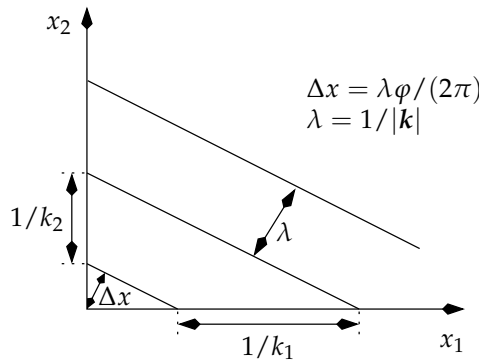


Figure B.1: Description of a periodic pattern using the wave vector \mathbf{k} and phase ϕ .

Note that the definition of the discrete Fourier transform restricts both the spacial and the Fourier domain to a finite number of values. Thereby, signals are assumed to be periodic, i. e. the boundaries of a discrete signal are joined like a ring in 1D or a torus in 2D:

$$\begin{aligned} \text{Fourier domain: } \forall k \in \mathbb{Z} : \hat{g}_{i+kN} &= \hat{g}_i \\ \text{space domain: } \forall k \in \mathbb{Z} : g_{i+kN} &= g_i \end{aligned} \quad (\text{B.5})$$

Note that the popular FFTW library does not normalise the Fourier transform with \sqrt{N} , thus it calculates

$$\hat{g}'_\nu = \hat{g}_\nu \sqrt{N} = \sum_{n=0}^{N-1} g_n \exp\left(-\frac{2\pi i n \nu}{N}\right). \quad (\text{B.6})$$

Accordingly, for the inverse transform, a factor of $1/N$ must be explicitly included.

For continuous signals, g is a continuous quadratic integrable function $g : \mathbb{R}^W \rightarrow \mathbb{C}$, $W \in \mathbb{N}$. The basis of continuous Fourier space is

$$B_k(\mathbf{x}) = e^{2\pi i \mathbf{k} \cdot \mathbf{x}} \quad (\text{B.7})$$

and the continuous Fourier transformation is defined as

$$\hat{g}(\mathbf{k}) = B_k \cdot g = \int_{-\infty}^{+\infty} g(\mathbf{x}) \exp(-2\pi i \mathbf{k} \cdot \mathbf{x}) \, d\mathbf{x}, \quad (\text{B.8})$$

and the inverse transformation by

$$g(\mathbf{x}) = B_x^{-1} \cdot \hat{g} = \int_{-\infty}^{+\infty} \hat{g}(\mathbf{k}) \exp(+2\pi i \mathbf{k} \cdot \mathbf{x}) \, d\mathbf{k}. \quad (\text{B.9})$$

The relation is again written as $g(\mathbf{x}) \circ\!\!\!\rightarrow \hat{g}(\mathbf{k})$

For the more-dimensional case the Fourier transform is analogous to the one-dimensional case for each dimension.

In general, variables in Fourier space are complex numbers. In contrast, measured signals are always real functions. Real functions become hermitian functions in Fourier space. Thus, there is not more information present in the complex Fourier space, and the transformation back always yields real functions.

For practical applications of the DFT, it is important to know the spatial discretisation Δx to compare different signals in Fourier space. The different signals may have different discretisation intervals or some of them may be continuous. The relation can be understood if the Fourier integral, equation (B.8), is approximated by a sum and discretised both in position space with $x = n\Delta x$ and in Fourier space with $k = \nu\Delta k$, $\Delta x \Delta k = 1/N$:

$$\begin{aligned} \hat{g}(\nu\Delta k) &= \int_{-\infty}^{+\infty} g(x) \exp(-2\pi i \nu \Delta k x) \, dx \\ &\approx \sum_{n=1}^N g_n \exp(-2\pi i \nu \Delta k \Delta x) \Delta x \\ &= \sqrt{N} \Delta x \frac{1}{\sqrt{N}} \sum_{n=1}^N g_n \exp\left(-\frac{2\pi i n \nu}{N}\right) \\ &= \sqrt{N} \Delta x \hat{g}'_\nu \end{aligned} \quad (\text{B.10})$$

Thus, a Fourier transform calculated by DFT must be multiplied with $\sqrt{N}\Delta x = 1/(\sqrt{N}\Delta k)$ to be referred to the unit interval of wave numbers. Without this normalisation the signal is referred to the interval $\sqrt{N}\Delta k$ and therefore different for every discretisation.

An analogous normalisation must be done for 2D and higher dimensional signals.

Convolution

An important operation in signal analysis is the convolution

$$(g * h) := \int_{-\infty}^{\infty} h(\xi)g(x - \xi) d\xi \quad (\text{continuous}), \quad (\text{B.11})$$

$$(g * h)_i := \sum_{j=1}^N h_j g_{i-j} \quad (\text{discrete}). \quad (\text{B.12})$$

Often, h is zero except for a small area. It is called *convolution mask* or *point spread function* (PSF). For all x , the result is a kind of weighted average of $g(x)$ in the neighbourhood of x . Such a process is also called *filter*.

The convolution is symmetric: $h * g = g * h$, and associative: $f * (g * h) = (f * g) * h$, and distributive over addition: $(f + g) * h = f * h + g * h$. In the Fourier domain, a convolution becomes a complex multiplication:

$$\begin{aligned} g(x) * h(x) &\longleftrightarrow \hat{g}(k)\hat{h}(k) \\ g * h &\longleftrightarrow N\hat{g}\hat{h} \end{aligned} \quad (\text{B.13})$$

The Fourier-transformed PSF is called *optical transfer function* (OTF).

Nomenclature

Notation

The mathematical structure of symbols is indicated by their typographical appearance:

- a scalar
- \mathbf{a} vector or image
- \mathbf{A} tensor

Symbols

- j water flux [m s^{-1}], page 6
- k wave vector [m^{-1}], page 147
- x position [m]
- ψ_i potential energy density of fluid i [J m^{-3}], page 6
- ψ_m matric potential [J m^{-3}], page 6
- K hydraulic conductivity [$\text{kg}^{-1} \text{m}^3 \text{s}$], page 6
- θ volumetric water content [-], page 6
- t time [s]
- z depth [m]

Abbreviations

- CF capillary fringe, page 83
- IOTF inverse optical transfer function, page 110
- LTM light transmission method, page 88
- OTF optical transfer function, page 149
- PDF probability density function, page 144
- PSF point spread function, page 149
- REV representative elementary volume, page 5
- TDR time domain reflectometry
- WT water table

Bibliography

- Allen, R. G., Pereira, L. S., Raes, D., and Smith, M. *Crop evapotranspiration – Guidelines for computing crop water requirements*. FAO Food and Agriculture Organization of the United Nations, Rome, 1998. ISBN 92-5-104219-5.
- Bakker, G., van der Ploeg, M. J., de Rooij, G. H., Hoogendam, C. W., Gooren, H. P. A., Huiskes, C., Koopal, L. K., and Kruidhof, H. New Polymer Tensiometers: Measuring Matric Pressures Down the the Wilting Point. *Vadose Zone J.*, 6:196–202, 2007. doi: 10.2136/vzj2006.0110.
- Bänninger, D. *Remote soil texture recognition*. Ph.D. thesis, ETH Zürich, Zürich, 2004.
- Bänninger, D., Lehmann, P., Flühler, H., and Tölke, J. Effect of Water Saturation on Radiative Transfer. *Vadose Zone J.*, 4:1152–1160, 2005. doi:10.2136/vzj2004.0109.
- Bastian, P. Numerik partieller Differentialgleichungen. 2008. URL <http://conan.iwr.uni-heidelberg.de/teaching/scripts/numpde-article.pdf>. Vorlesungsskript.
- Bayer, A. *X-ray attenuation techniques to explore the dynamics of water in porous media*. Ph.D. thesis, Heidelberg University, Heidelberg, 2005. URL <http://www.ub.uni-heidelberg.de/archiv/5629>.
- Berkowitz, B., Silliman, S. E., and Dunn, A. M. Impact of the Capillary Fringe on Local Flow, Chemical Migration, and Microbiology. *Vadose Zone J.*, pages 534–548, 2004.
- Biggs, D. S. C. and Andrews, M. Acceleration of iterative image restoration algorithms. *Applied Optics*, 36(8):1766–1775, 1997.
- Bock, E., Hara, T., Frew, N., and McGillis, W. Relationship between air-sea gas transfer and short wind waves. *J. Geophys. Res.*, 104:25821–25831, 1999.
- Bronstein, I. N., Semendjajew, K. A., Musiol, G., and Mühlig, H. *Taschenbuch der Mathematik*. Verlag Harri Deutsch, Thun, Frankfurt am Main, 5 edition, 2001. ISBN 3-8171-2005-2.
- Brooks, R. H. and Corey, A. T. Properties of porous media affecting fluid flow. *J. Irrigation and Drainage Div., Proc. Am. Soc. Civil Eng. (IR2)*, 92:61–88, 1966.
- Buchner, J. S. *Solute Transport in Porous Media: Theory and Experiment*. Diploma thesis, Heidelberg University, Heidelberg, 2009.
- Chaplin, M. Water Structure and Science. Website, 2009. URL <http://www.lsbu.ac.uk/water/index.html>.
- Constantz, J. Confirmation of Rate-Dependent Behavior in Water Retention During Drainage in Nonswelling Porous Materials. *Water Resour. Res.*, 29(4):1331–1334, 1993.
- Corey, A. T. *Mechanics of Immiscible Fluids in Porous Media*. Water Resources Publications, Highlands Ranch, 3rd edition, 1994.

- Dane, J. H. and Wierenga, P. J. Effect of hysteresis on the prediction of infiltration, redistribution and drainage of water in a layered soil. *J. Hydrol.*, 25:229–242, 1975.
- Davidson, J. M., Nielson, D. R., and Biggar, J. W. The dependence of soil water uptake and release upon the applied pressure increment. *Soil Sci. Soc. Am. Proc.*, 30:298–303, 1966.
- de Gennes, P.-G., Brochard-Wyart, F., and Quéré, D. *Capillarity and Wetting Phenomena*. Springer, New York, 2004. ISBN 0-387-00592-7.
- Demtröder, W. *Experimentalphysik 3*. Springer, Berlin Heidelberg New York, 2nd edition, 2000. ISBN 3-0-66790-3.
- Durner, W. and Flühler, H. *Soil Hydraulic Properties*. John Wiley and Sons, 2005.
- Eching, S. O., Hopmans, J. W., and Wendroth, O. Unsaturated hydraulic conductivity from transient multistep outflow and soil water pressure data. *Soil Sci. Soc. Am. J.*, 58:687–695, 1994.
- Eisenberg, D. and Kauzmann, W. *The structure and properties of water*. Clarendon Press, Oxford, 1969.
- Enderby, J. A. The domain model of hysteresis. Part 2: Interacting domains. *Trans. Faraday Soc.*, 52:106–120, 1956. doi:10.1039/TF9565200106.
- Everett, D. H. and Whitton, W. I. A general approach to hysteresis. *Trans. Faraday Soc.*, 48:749–757, 1952.
- Foken, T. *Angewandte Meteorologie*. Springer, Berlin, Heidelberg, New York, 2003. ISBN 3-540-00322-3.
- Frew, N., Goldman, J., Dennett, M., and Johnson, A. Impact of phytoplankton-generated surfactants on air-sea gas exchange. *J. Geophys. Res.*, 95:3337–3351, 1990.
- Friedman, S. P. Dynamic contact angle explanation of flow rate-dependent saturation-pressure relationships during transient liquid flow in unsaturated porous media. *J. Adhesion Sci. Technol.*, 13(12):1495–1518, 1999.
- Gardner, W. R. and Miklich, F. J. Unsaturated conductivity and diffusivity measurements by a constant flux method. *Soil Sci. Soc. Am. J.*, 93:271–274, 1962.
- Gillham, R. W. The capillary fringe and its effect on water-table response. *J. Hydrol.*, 67:307–324, 1984.
- Gillham, R. W., Klute, A., and Heermann, D. F. Hydraulic properties of a porous medium: Measurement and empirical representation. *Soil Sci. Soc. Am. J.*, 40:203–207, 1976.
- Glass, R. J., Steenhuis, T. S., and Parlange, J.-Y. Mechanism for finger persistence in homogeneous unsaturated porous media: theory and verification. *Soil Sci.*, 148(1):60–70, 1989.
- Gonzalez, R. C. and Woods, R. E. *Digital image processing*. Addison-Wesley, Reading, Mass., 1993. ISBN 0-201-60078-1, 978-0-201-60078-0.
- Großmann, C. and Roos, H.-G. *Numerische Behandlung partieller Differentialgleichungen*. Teubner, Wiesbaden, 3rd edition, 2005. ISBN 3-519-22089-X.
- Haken, H. and Wolf, H. C. *Molekülphysik und Quantenchemie*. Springer, Berlin Heidelberg, 5 edition, 2005. ISBN 3-540-43551-4.

- Hanisch, R. J., White, R., and Gilliland, R. *The Richardson-Lucy Algorithm and Adaptions*, pages 318–337. Academic Press, San Diego, 2nd edition, 1997. ISBN 0-12-380222-9.
- Hanks, R. J., Klute, A., and Bresler, E. A numeric method for estimating infiltration, redistribution, drainage, and evaporation of water from soil. *Water Resour. Res.*, 5:1064–1069, 1969.
- Heimann, F. *An unfitted discontinuous galerking method for two-phase flow*. diploma thesis, Heidelberg University, 2009.
- Heitler, W. *The Quantum Theory of Radiation*. Dover, Oxford, 3rd edition, 1984.
- Heow, D. *Pore-Scale Investigation of Two-Phase Flow Phenomena in Porous Media*. Diploma thesis, Heidelberg University, Heidelberg, 2008.
- Hirakawa, K. and Parks, T. W. Adaptive homogeneity-directed demosaicing algorithm. *IEEE Transactions on Image Processing*, 14(3):360–369, 2005. doi:10.1109/TIP.2004.838691.
- Hoa, N. T. A New Method Allowing the Measurement of Rapid Variations of the Water Content in Sandy Porous Media. *Water Resour. Res.*, 17(1):41–48, 1981.
- Hopmans, J. and Šimůnek, J. Review of inverse estimation of soil hydraulic properties. In: M. van Genuchten, F. Leij, and L. Wu (editors), *Proceedings of the International Workshop Characterization and Measurement of Hydraulic Properties of Unsaturated Porous Media*, pages 643–659. University of California, Riverside, CA, 1999.
- Ippisch, O. Numerik von Transportprozessen in porösen Medien. 2008. URL http://conan.iwr.uni-heidelberg.de/teaching/numpormed_ss2009/numpormed.pdf. Vorlesungsskript.
- Ippisch, O., Vogel, H., and Bastian, P. Validity limits for the van Genuchten-Mualem model and implications for parameter estimation and numerical simulation. *Adv. Water Resour.*, 29:1780–1789, 2006. doi:10.1016/j.advwatres.2005.12.011.
- Jackson, J. D. *Classical electrodynamics*. John Wiley & Sons, New York, 3rd edition, 1999. ISBN 978-0-471-30932-1.
- Jähne, B. *Digital image processing*. Springer, Berlin, Heidelberg, New York, 5 edition, 2005. ISBN 3-540-24035-7.
- Jähne, B. and Haußecker, H. Air-water gas exchange. *Annu. Rev. Fluid Mech.*, 30:443–468, 1998.
- Jaynes, D. B. Comparison of soil-water hysteresis models. *J. Hydrol.*, 75:287–299, 1984.
- Jin, Y. and Jury, W. A. Characterizing the Dependence of Gas Diffusion Coefficient on Soil Properties. *Soil Sci. Soc. Am. J.*, 60:66–71, 1996.
- Jury, W. A., Gardner, W. R., and Gardner, W. H. *Soil Physics*. John Wiley & Sons, New York, 5. edition edition, 1991.
- Klute, A. and Heermann, D. F. Soil water profile development under a periodic boundary condition. *Soil Sci.*, 117:265–271, 1974.
- Landau, L. D. and Lifschitz, E. M. *Lehrbuch der theoretischen Physik VI – Hydrodynamik*. Akademie-Verlag, Berlin, 1991. ISBN 3-05-500070-6.
- Lehmann, P., Stauffer, F., Hinz, C., Dury, O., and Flühler, H. Effect of hysteresis on water flow in a sand column with a fluctuating capillary fringe. *J. Contam. Hydrol.*, 33:81–100, 1998.

- Lennartz, F. *Einfluß von instationären Fließzuständen auf die Wassergehalt-Wasserpotentialbeziehung: eine kritische Beleuchtung des herkömmlichen Konzeptes.* Ph.D. thesis, Kiel University, Kiel, 1992.
- Lide, D. R. (editor). *Handbook of Chemistry and Physics.* CRC Press, Boca Raton, 76th edition, 1995.
- Lückenhaus, M. *Grundlagen des Wiener-Filters und seine Anwendung in der Bildanalyse.* diploma thesis, Technical University München, 1997.
- Lucy, L. B. An iterative technique for the rectification of observed distributions. *Astron. J.*, 79(6):745–754, 1974.
- Magain, P., Courbin, F., and Sohy, S. Deconvolution with Correct Sampling. *The Astronomical J.*, 494(1):472–477, 1998. doi:10.1086/305187. URL <http://www.journals.uchicago.edu/doi/abs/10.1086/305187>.
- McKenna, S. and McGillis, W. The role of free-surface turbulence and surfactants in airewater gas transfer. *Int. J. Heat and Mass Transfer*, 47:539–553, 2004.
- Mualem, Y. Modified Approach to Capillary Hysteresis Based on a Similarity Hypothesis. *Water Resour. Res.*, 9(5):1324–1331, 1973.
- Mualem, Y. A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resour. Res.*, 12:513–522, 1976.
- Mualem, Y. Extention of the Similarity Hypothesis Used for Modeling the Soil Water Characteristics. *Water Resour. Res.*, 13(4):773–780, 1977.
- Mualem, Y. A modified dependent-domain theory of hysteresis. *Soil Sci.*, 137:243–291, 1984.
- Mualem, Y. and Dagan, G. A Dependent Domain Model of Capillary Hysteresis. *Water Resour. Res.*, 11(3):452–460, 1975.
- Mualem, Y. and Miller, E. E. A hysteresis model based on an explicit domain-dependence function. *Soil Sci. Soc. Am. J.*, 43:1067–1073, 1979.
- Murray, F. W. On the Computation of Saturation Vapor Pressure. *Journal of Applied Meteorology*, 6:203–204, 1967.
- Nielsen, P. and Perrochet, P. Watertable dynamics under capillary fringes: experiments and modelling. *Adv. Water Res.*, 23(5):503–515, 2000.
- Niemet, M. R., Rockhold, M. L., Weisbrod, N., and Selker, J. S. Relationships between gas-liquid interfacial surface area, liquid saturation, and light transmission in variably saturated porous media. *Water Resour. Res.*, 38(8):1135–1146, 2002. doi:10.1029/2001WR000785. URL <http://www.agu.org/journals/wr/wr0208/2001WR000785/2001WR000785.pdf>.
- Niemet, M. R. and Selker, J. S. A new method for quantification of liquid saturation in 2D translucent porous media using light transmission. *Adv. Water Res.*, 24:651–666, 2001.
- Palmer, K. F. and Williams, D. Optical properties of water in the near infrared. *J. Opt. Soc. Am.*, 64(8):1107–1110, 1974. doi:10.1364/JOSA.64.001107. URL <http://www.opticsinfobase.org/abstract.cfm?URI=josa-64-8-1107>.

- Parker, J. C., Kool, J. B., and van Genuchten, M. T. Determining soil hydraulic properties from one-step outflow experiments by parameter estimation: II. Experimental studies. *Soil Sci. Soc. Am. J.*, 49:1354–1359, 1985.
- Peters, A. and Durner, W. Simplified evaporation method for determining soil hydraulic properties. *J. Hydrol.*, 356:147–162, 2008.
- Philip, J. R. and de Vries, D. A. Moisture Movement in Porous Materials under Temperature Gradients. *Trans. Am. Geophys. Union (EOS)*, 38:222–232, 1957.
- Plagge, R. *Bestimmung der ungesättigten hydraulischen Leitfähigkeit im Boden*. Ph.D. thesis, Institut für Ökologie, TU, Berlin, 1991.
- Plagge, R., Haeupl, P., and Renger, M. Transient effects on the hydraulic properties of porous media. In: M. T. van Genuchten, F. J. Leif, and L. Wu (editors), *Proceedings of the International Workshop on Characterization and Measurement of the Hydraulic Properties of Unsaturated Porous Media*, pages 905–912. University of California, Riverside, CA, USA, 1999.
- Poulovassilis, A. Hysteresis of pore water, an application of the concept of independent domains. *Soil Sci.*, 93(6):405–412, 1962.
- Poulovassilis, A. and Childs, E. E. The hysteresis of pore water: The nonindependence of domain. *Soil Sci.*, 112:301–312, 1971.
- Prat, M. Recent advances in pore-scale models for drying of porous media. *Chemical Engineering Journal*, 86(1-2):153–164, 2002. ISSN 1385-8947. doi:10.1016/S1385-8947(01)00283-2. URL <http://www.sciencedirect.com/science/article/B6TFJ-44SK3MM-D/2/7b05431be257ac2b63e517347139a0a9>.
- Rad, N. S. and Tumay, M. T. Factors affecting sand specimen preparation by raining. *Geotechnical Testing J.*, 10(1):31–37, 1987.
- Rawlins, S. L. and Campbell, G. S. Water potential: Thermocouple Psychrometry. In: A. Klute (editor), *Methods of Soil Analysis, Part 1. Physical and Mineralogical Methods*, Agronomy Series 9, pages 597–618. American Society of Agronomy, Madison, WI, 2nd edition, 1986.
- Rezanezhad, F., Vogel, H.-J., and Roth, K. Experimental study of fingered flow through initially dry sand. *Hydrology and Earth System Sciences Discussions*, 3(4):2595–2620, 2006. ISSN 1812-2108. URL <http://www.hydrol-earth-syst-sci-discuss.net/3/2595/2006/>.
- Richards, L. A. Capillary conduction of liquids through porous mediums. *Physics*, 1:318–333, 1931.
- Richardson, W. H. Bayesian-Based Iterative Method of Image Restoration. *J. Opt. Soc. Am.*, 62(1):55–59, 1972.
- Romano, N. and Santini, A. Determining soil hydraulic functions from evaporation experiments by a parameter estimation approach: Experimental verifications and numerical studies. *Water Resour. Res.*, 35:3343–3359, 1999.
- Ronen, D., Scher, H., and Blunt, M. Field observations of a capillary fringe before and after a rainy season. *J. Contam. Hydr.*, 44:103–118, 2000.
- Ross, P. J. and Smetten, K. R. A simple treatment of physical nonequilibrium water flow in soils. *Soil Sci. Soc. Am. J.*, 64:1926–1930, 2000.

- Roth, K. Soil Physics Lecture Notes. Heidelberg, 2005. URL http://www.iup.uni-heidelberg.de/institut/forschung/groups/ts/soil_physics/students/lecture_notes05.
- Scherer, G. W. Theory of Drying. *J. American Ceramic Soc.*, 73(1), 1990. doi:10.1111/j.1151-2916.1990.tb05082.x.
- Schneider, K. *Novel evaporation experiment to determine soil hydraulic properties*. diploma thesis, Heidelberg University, 2005.
- Schultze, B., Ippisch, O., Huwe, B., and Durner, W. Dynamic Nonequilibrium During Unsaturated Water Flow. In: M. T. van Genuchten, F. J. Leij, and L. Wu (editors), *Proc. Int. Workshop on Characterization and Measurement of the Hydraulic Properties of Unsaturated Porous Media*. Riverside, CA., USA, 1999.
- Segelstein, D. *The complex refractive index of water*. Master's thesis, University of Missouri, Kansas City, 1981.
- Sherwood, T. K. The Drying of Solids—III Mechanism of the Drying of Pulp and Paper. *Ind. Eng. Chem.*, 22(2):132–136, 1930. doi:10.1021/ie50242a009. URL <http://pubs.acs.org/doi/abs/10.1021/ie50242a009>.
- Shokri, N., Lehmann, P., Vontobel, P., and Or, D. Drying front and water content dynamics during evaporation from sand delineated by neutron radiography. *Water Resour. Res.*, 44, 2008. doi:10.1029/2007WR006385.
- Šimůnek, J., Jarvis, N. J., van Genuchten, M. T., and Gardenas, A. Review and comparison of models for describing non-equilibrium and preferential flow and transport in the vadose zone. *J. Hydrol.*, 272:14–35, 2003.
- Šimůnek, J., Wendroth, O., and van Genuchten, M. T. Parameter Estimation Analysis of the Evaporation Method for Determining Soil Hydraulic Properties. *Soil Sci. Soc. Am. J.*, 62:894–905, 1998.
- Smiles, D. E., Vachaud, G., and Vauclin, M. A Test of the Uniqueness of the Soil Moisture Characteristic During Transient, Nonhysteretic Flow of Water in a Rigid Soil. *Soil Sci. Soc. Am. J.*, 35:534–539, 1971.
- Stauffer, F. *Einfluss der kapillaren Zone auf instationäre Drainagevorgänge*. Ph.D. thesis, Eidgenössische technische Hochschule, Zürich, 1977.
- Tamari, S., Bruckler, L., Halbertsma, J., and Chadoeuf, J. A Simple Method for Determining Soil Hydraulic Properties in the Laboratory. *Soil Sci. Soc. Am. J.*, 57:642–651, 1993.
- Tidwell, V. C. and Glass, R. J. X ray and visible light transmission for laboratory measurement of two-dimensional saturation fields in thin-slab systems. *Water Resour. Res.*, 30(11):2873–2882, 1994.
- Toorman, A. F., Wierenga, P. J., and Hills, R. G. Parameter estimation of hydraulic properties from one-step outflow data. *Water Resour. Res.*, 28(11):3021–3028, 1992.
- Topp, G. C. Soil-water hysteresis: The domain model theory extended to pore interaction conditions. *Soil Sci. Soc. Am. Proc.*, 35:219–225, 1971.

- Topp, G. C., Klute, A., and Peters, D. B. Comparison of water content-pressure head data obtained by equilibrium, steady state, and unsteady state methods. *Soil Sci. Soc. Am. Proc.*, 31:312–314, 1967.
- Topp, G. C. and Miller, E. E. Hysteretic moisture characteristics and hydraulic conductivities for glass-bead media. *Soil Sci. Soc. Am. Proc.*, 30:156–162, 1966.
- Tsimpanogiannis, I. N., Yortsos, Y. C., Poulou, S., Kanellopoulos, N., and Stubos, A. K. Scaling theory of drying in porous media. *Phys. Rev. E*, 59(5):4353–4365, 1999.
- Ustohal, P., Stauffer, F., and Dracos, T. Measurement and modeling of hydraulic characteristics of unsaturated porous media with mixed wettability. *J. Contam. Hydrol.*, 33:5–37, 1998.
- Vachaud, G., Vauclin, M., and Wakil, M. A study of the uniqueness of the soil moisture characteristic during desorption by vertical drainage. *Soil Sci. Soc. Am. Proc.*, 36:531–532, 1972.
- van Dam, J. C., Stricker, J. N. M., and Droogers, P. Inverse method to determine soil hydraulic functions from multistep outflow experiments. *Soil Sci. Soc. Am. J.*, 58:647–652, 1994.
- van Genuchten, M. T. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.*, 44:892–898, 1980.
- Vrugt, J., Stauffer, P., Wöhling, T., Robinson, B., and Vessilinov, V. Inverse modeling of subsurface flow and transport properties: A review with new developments. *Vadose Zone J.*, 7(2):843–864, 2008.
- Warren, S. Optical constants of ice from the ultraviolet to the microwave. *Appl. Optics*, 23(8):1206–1225, 1984. doi:10.1364/AO.23.001206.
- Wendroth, O., Ehlers, W., Hopmans, J. W., Kage, H., Halbertsma, J., and Wösten, J. H. M. Reevaluation of the Evaporation Method for Determining Hydraulic Functions in Unsaturated Soils. *Soil Sci. Soc. Am. J.*, 57:1436–1443, 1993.
- Wildenschild, D., Hopmans, J. W., and Šimůnek, J. Flow Rate Dependence of Soil Hydraulic Characteristics. *Soil Sci. Soc. Am. J.*, 65:35–48, 2001.
- Zinth, W. and Zinth, U. *Optik*. Oldenbourg, München, 2nd edition, 2009. ISBN 978-3-486-58801-9.
- Zurmühl, T. and Durner, W. Determination of parameters for bimodal hydraulic functions by inverse modeling. *Soil Sci. Soc. Am. J.*, 62:874–880, 1998.

Acknowledgements

First of all, I want to thank Kurt Roth for giving me the opportunity to write this thesis, the friendly atmosphere in the group, freedom in my work, interesting and revealing discussions and ideas, and for the key idea to understand the water content minimum during infiltration. I thank Jens B. and Steffen for the successful and friendly cooperation, Bernd Jähne for the nice and instructive collaboration, Uwe Schimpf and Roland Rocholz for answering all my questions about image analysis and cameras, and Michael Erz for helping with the camera calibration at the HCI. I am grateful to Olaf Ippisch for providing the $\mu\phi$ model, for implementing the soil-atmosphere boundary and water vapour transport into it, for helping with numerical difficulties and for interesting discussions. Thanks also Michael Gisi for the FTIR measurements, Jens Tschritter for the spectroscopic measurement of the light source, Peter Bastian for interesting discussions, and the workshop for their excellent work on the experimental setup. Furthermore, I thank Ute, Ulrike, Steffen, Rüdiger, Patrick, Philipp, Pan, Moritz, Mirjam, Mathias, Martin, Jens W., Jens B., Holger, Gaby, Felix, Doreen, David, Cornelius, Anne and Angelika for the enjoyable atmosphere in the soil physics group, coffee breaks at the Botanik and interesting discussions, Steffen, Jens B., Andreas, Felix, Gaby and Korinna for proof-reading, the colleagues from the DyCap research group for an interesting experience, and finally Korinna for all the things which cannot all be put into words and therefore should be kept in silence.