Günter Hofstetter *Editor*

Computational Engineering

Springer

Computational Engineering

Günter Hofstetter
Editor

Computational Engineering

Springer

Editor
Günter Hofstetter
Institute of Basic Sciences in Engineering Science
University of Innsbruck
Innsbruck, Austria

ISBN 978-3-319-05932-7 ISBN 978-3-319-05933-4 (eBook)
DOI 10.1007/978-3-319-05933-4
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014939849

© Springer International Publishing Switzerland 2014
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Aiming at the design and analysis of complex engineering systems, computational engineering combines engineering sciences, mathematics, and computer science. It comprises the development, application, and validation of computational models as well as the visualization of simulation results. Taking advantage of continuing advances in computer hardware, software technology, and numerical algorithms, computational engineering plays an increasingly important role in the development and operation of engineering products and systems.

The book provides an overview of the broad spectrum of research activities within the framework of the research center *Computational Engineering* at the University of Innsbruck. The topics covered focus on mathematical modeling, numerical simulation, and experimental validation in several fields of engineering sciences. In particular, constitutive models and their implementation into finite element codes, sensitivity and reliability analysis of engineering structures including applications in aerospace engineering and earthquake engineering, multi-phase models and multi-scale models in civil engineering, applications of scientific computing in urban water management and numerical simulations in hydraulic engineering, and—last but not least—the application of a genetic algorithm for the registration of laser scanner point clouds in geoinformation science are presented.

The research center *Computational Engineering* is part of the research focal point *Scientific Computing* at the University of Innsbruck. The latter integrates research activities of the University of Innsbruck in the fields of information technology and e-science. As the success in those scientific disciplines crucially depends on the powerful computer hardware, the financial support by the Austrian Federal Ministry of Science and Research (BMWF) within the framework of the University Infrastructure Program is gratefully acknowledged.

Innsbruck, Austria — Günter Hofstetter
February 2014

Contents

List of Contributors

Stefan Achleitner Unit of Hydraulic Engineering, University of Innsbruck, Innsbruck, Austria

Christoph Adam Unit of Applied Mechanics, University of Innsbruck, Innsbruck, Austria

Matthias Aschaber Unit of Strength of Materials and Structural Analysis, University of Innsbruck, Innsbruck, Austria

Markus Aufleger Unit of Hydraulic Engineering, University of Innsbruck, Innsbruck, Austria

Gregor Burger Unit of Environmental Engineering, University of Innsbruck, Innsbruck, Austria

Wolfgang Fellin Unit of Geotechnical and Tunnel Engineering, University of Innsbruck, Innsbruck, Austria

Roman Gabl Unit of Hydraulic Engineering, University of Innsbruck, Innsbruck, Austria

Peter Gamnitzer Unit of Strength of Materials and Structural Analysis, University of Innsbruck, Innsbruck, Austria

Bernhard Gems Unit of Hydraulic Engineering, University of Innsbruck, Innsbruck, Austria

Thomas Gschnitzer Unit of Hydraulic Engineering, University of Innsbruck, Innsbruck, Austria

Klaus Hanke Unit of Surveying and Geoinformation, University of Innsbruck, Innsbruck, Austria

Günter Hofstetter Unit of Strength of Materials and Structural Analysis, University of Innsbruck, Innsbruck, Austria

Robert Klar Unit of Hydraulic Engineering, University of Innsbruck, Innsbruck, Austria

Manfred Kleidorfer Unit of Environmental Engineering, University of Innsbruck, Innsbruck, Austria

Dimitrios Kolymbas Unit of Geotechnical and Tunnel Engineering, University of Innsbruck, Innsbruck, Austria

Roman Lackner Unit of Material Technology, University of Innsbruck, Innsbruck, Austria

Michael Mair Unit of Environmental Engineering, University of Innsbruck, Innsbruck, Austria

Michael Meister Unit of Environmental Engineering, University of Innsbruck, Innsbruck, Austria

Michael Oberguggenberger Unit of Engineering Mathematics, University of Innsbruck, Innsbruck, Austria

Alexander Ostermann Institute of Mathematics, University of Innsbruck, Innsbruck, Austria

Manuel Plörer Unit of Hydraulic Engineering, University of Innsbruck, Innsbruck, Austria

Wolfgang Rauch Unit of Environmental Engineering, University of Innsbruck, Innsbruck, Austria

Thomas Ring Institute for Mechanics of Materials and Structures, Vienna University of Technology, Vienna, Austria

Stefan Schenk Unit of Surveying and Geoinformation, University of Innsbruck, Innsbruck, Austria

Bernhard Schmelzer Unit of Engineering Mathematics, University of Innsbruck, Innsbruck, Austria

Robert Sitzenfrei Unit of Environmental Engineering, University of Innsbruck, Innsbruck, Austria

Christian Urich Unit of Environmental Engineering, University of Innsbruck, Innsbruck, Austria

Matthias Zeiml Institute for Mechanics of Materials and Structures, Vienna University of Technology, Vienna, Austria

Chapter 1
Constitutive Models in Finite Element Codes

W. Fellin and A. Ostermann

Abstract In finite element simulations the constitutive information is usually handled by a user-supplied subroutine. For a prescribed strain increment, this subroutine provides the finite element code with the corresponding stress increment and the Jacobian, which is required to build the consistent tangent operator. We propose an approach that relieves the user from computing and coding the Jacobian information. Instead, this information is computed automatically together with the stress increment. This approach requires reliable and efficient numerical integration. In particular, adaptivity and automatic error control are highly desirable features. Such integrators are presented in this article. The underlying ideas of the approach are first elucidated at simple one-dimensional problems from geotechnics. However, it is also discussed how this concept can be used in a fully three-dimensional framework. We expect that this new approach will strongly enhance the development of constitutive models and help to identify the most appropriate ones.

1.1 Introduction

Challenging problems in computational inelasticity are usually tackled by an incremental finite element approach: the load is applied in discrete steps and the equilibrium equations are solved after each load increment. In geotechnical applications, however, the actual use of finite element codes is often restricted

W. Fellin (✉)
Unit of Geotechnical and Tunnel Engineering, University of Innsbruck, Technikerstr. 13, 6020 Innsbruck, Austria
e-mail: wolfgang.fellin@uibk.ac.at

A. Ostermann
Department of Mathematics, University of Innsbruck, Technikerstr. 13, 6020 Innsbruck, Austria
e-mail: alexander.ostermann@uibk.ac.at

G. Hofstetter (ed.), *Computational Engineering*, DOI 10.1007/978-3-319-05933-4_1,
© Springer International Publishing Switzerland 2014

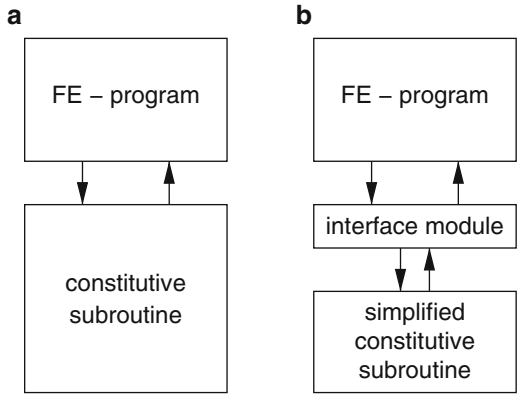

Fig. 1.1 Classical structure of a co-simulation approach (a). The new structure based on an interface module (b)

by the fact that state-of-the-art soil models are not implemented in commercial programs. Although finite element programs offer, in general, an interface for using an own material model, the actual implementation is often a time-consuming task. We believe that developers and users of advanced constitutive models should not be burdened with all the numerical details of such an implementation.[1]

For reasons of flexibility, standard finite element programs use a co-simulation approach: the equilibrium equations are solved with the help of an iterative algorithm in the finite element package, and the necessary constitutive information is obtained through an interface from a user-supplied subroutine which implements the constitutive relation, see Fig. 1.1a. In this way new models can be added to existing and well-established finite element packages.

At the interface, the finite element code proposes a strain increment $\Delta\boldsymbol{\varepsilon}$ for which the constitutive model has to supply the corresponding stress increment $\Delta\mathbf{T}$ as well as the derivative of the stress increment with respect to the strain increment

$$\frac{\partial \Delta \mathbf{T}}{\partial \Delta \boldsymbol{\varepsilon}} . \tag{1.1}$$

This Jacobian information is the constitutive part of the consistent tangent operator. It is well known that any inconsistency with the stress-update algorithm of the constitutive model will spoil the quadratic convergence of Newton's method in the iterative solution of the initial-boundary value problem (see [18]). A consequence will be computational inefficiency. For simple material models, the Jacobian can be found analytically by differentiation of the constitutive equations. For more complicated constitutive models, however, this can be a tedious task and it is sometimes even not feasible. A remedy is numerical differentiation.

[1] By the term *implementation* we understand the whole process of developing the interface module: selecting an appropriate integration scheme, coding the scheme, and testing it at the levels of integration points, elements, and full initial-boundary value problems.

The aim of this article is to convince the reader that the constitutive subroutine should be split up into two parts (see Fig. 1.1b). On the one hand, the former constitutive subroutine gets simplified and just provides the constitutive equation. On the other hand, a new interface module takes care of the integration procedure and provides in addition the Jacobian and, if desired, sensitivity information (for the latter, see [9]). This approach requires general purpose integration schemes that are not tailored to a specific constitutive equation. We propose here fully adaptive second-order integrators. Depending on whether the problem is non-stiff or stiff, an explicit or a semi-implicit integrator is recommended. For index 2 problems that arise in the plastic case, a half-explicit integrator is our method of choice. By controlling the local error, these integrators automatically select an appropriate step size for the sub-stepping procedure.

In the spirit of [18] we first explain the underlying ideas and the basic properties of the integrators at simple one-dimensional problems. In the last part of this paper, however, we extend our approach to the fully 3D case (see Sect. 1.6). We expect that the chosen approach will strongly enhance the development of new constitutive models. Increasing the number of available models will eventually initiate a competitive selection of the most appropriate ones.

1.2 The Merits and Pitfalls of One-Dimensional Considerations

One-dimensional considerations are often simple and easy to follow, in general. Therefore, they are very useful for getting a first understanding of complex approaches and for laying a solid basis for proceeding to the three-dimensional reality. However, one has to bear in mind that a one-dimensional world does not really exist. Every one-dimensional model is just a reduction of a fully three-dimensional one, endowed with certain boundary conditions.

Let us consider a cylindrical sample of cohesive soil like clay. We can simply load this sample on the top with an increasing force F and measure the displacement of the top s, see Fig. 1.2a. Such a test is called uniaxial compression test. When a pressure is applied at the cylinder surface, the test is called triaxial test. We can also enclose the specimen in a rigid hollow cylinder and prohibit thereby the lateral extension of the specimen while compressing it vertically, see Fig. 1.2b. Then it is called confined uniaxial compression test, or oedometer test.

The results of both uniaxial compression tests can be plotted as a relation between the applied force and the measured displacement, see Fig. 1.3. The stiffness increases with displacement in the confined case but decreases in the unconfined one. A maximum possible load will eventually be reached in the unconfined case, which, however, is not the case for a confined sample. Obviously, the evolution of lateral confining stresses will cause this drastic change in behavior from unconfined to confined conditions. The sample will laterally extend in an unconfined

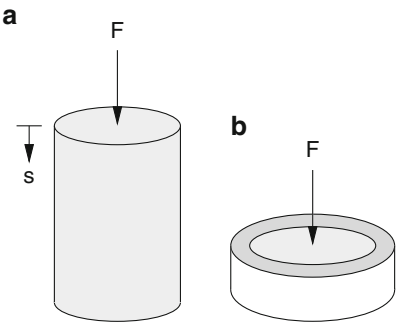

Fig. 1.2 Uniaxial compression tests. (**a**) Unconfined. (**b**) Confined

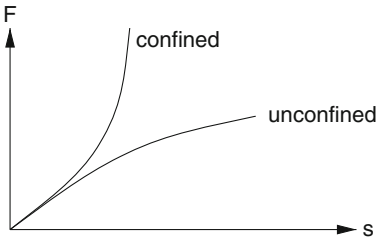

Fig. 1.3 Results of uniaxial compression tests

experiment, which is prohibited in the confined situation. The boundary conditions thus completely change the material behavior that is found in a one-dimensional relationship between forces and displacements.

The simplest constitutive model is linear elasticity. For such a model the stress–strain relationship in the unconfined case reads

$$\sigma = E\varepsilon \tag{1.2}$$

with Young's modulus E. In the confined cases it is

$$\sigma = E_s\varepsilon \tag{1.3}$$

with the stiffness modulus

$$E_s = E\frac{1-\nu}{(1+\nu)(1-2\nu)}\,. \tag{1.4}$$

Here, the effect of lateral confinement enters via Poisson's ratio ν, which is the negative ratio of the lateral to the vertical strain in the unconfined case. Note that the stiffness modulus is higher than Young's modulus. For example, for steel with $\nu = 0.3$ we have $E_s = 1.35E$.

To conclude, it is not possible to derive one-dimensional constitutive models that fully describe the material behavior. However, some essential properties can

be captured. Moreover, in the context of this article, the computational approaches for constitutive models can be explained much simpler.

1.3 Time Integration of Pure Rate Equations

In this section we present constitutive models of the rate type and discuss their numerical integration. We focus on adaptivity and discuss explicit and semi-implicit integration schemes for non-stiff and stiff problems, respectively.

1.3.1 Pure Rate Equation

Some constitutive models, like hypoplasticity for soil [14, 15], are of the rate type. Such models are relations between the objective stress rate $\mathring{\mathbf{T}}$ of the effective Cauchy stress [19], on the one hand, and the effective Cauchy stress $\mathbf{T}$, the Eulerian stretching $\mathbf{D}$, and some additional state or internal variables $\mathbf{Q}$ on the other hand,

$$\mathring{\mathbf{T}} = \mathbf{h}(\mathbf{T}, \mathbf{D}, \mathbf{Q}) \,. \tag{1.5}$$

The additional state variables obey a further set of evolution equations

$$\mathring{\mathbf{Q}} = \mathbf{k}(\mathbf{T}, \mathbf{D}, \mathbf{Q}) \,. \tag{1.6}$$

A one-dimensional version of hypoplasticity for oedometric boundary conditions reads as follows (see [3, 4, 8]):

$$\dot{T} = h(T, D) = C_1 T D + C_2 T |D| \,. \tag{1.7}$$

In the one-dimensional case, the objective time rate of the stress is equal to the total time derivative of the stress, and the stretching D is equal to the total derivative of the logarithmic (Hencky) strain (see [10]), i.e., $D = \dot{\varepsilon}$. An example of the behavior of this model is given in Fig. 1.4a.

A one-dimensional version of (1.5) that is suitable for triaxial boundary conditions is given in [3, 4],

$$\dot{T}_1 = C_1 (T_1 + T_2) D_1 + C_2 (T_1 - T_2) |D_1| \tag{1.8}$$

with T_1 being the vertical stress evolving due to the vertical compression with stretching D_1 under constant lateral stresses $T_2 = T_3$. An example of the behavior of this model is given in Fig. 1.5.

To enlarge the stiffness for small strains the concept of intergranular strain was developed [16]. It can be used as an add-on with all hypoplastic relations. To employ

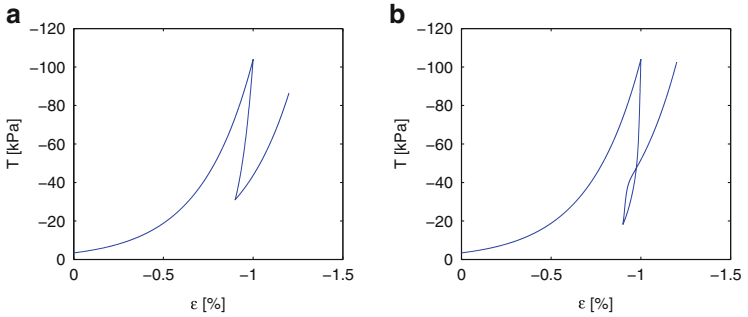

Fig. 1.4 Confined uniaxial compression tests: results of a loading / unloading / reloading cycle with the hypoplastic model and a given strain history: ε (%): 0, −1.0, −0.9, −1.2; material constants: $C_1 = -775$, $C_2 = -433$, $m_r = 5$, $R = 10^{-4}$; initial values: $T_0 = 3.4$ kPa and $\delta_0 = -R$. (**a**) Without intergranular strain (1.7). (**b**) With intergranular strain (1.12)

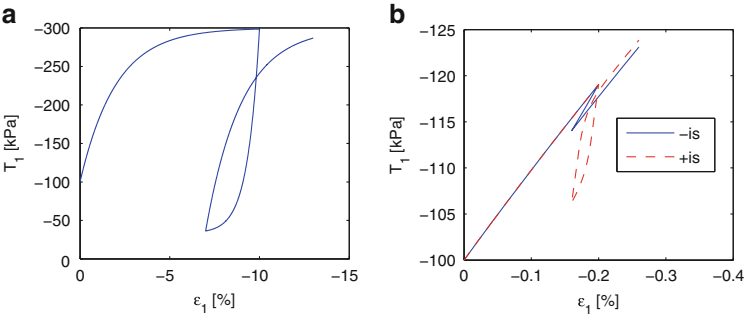

Fig. 1.5 Triaxial compression tests: results of a loading / unloading / reloading cycle with the hypoplastic model (1.8): $C_1 = -50$, $C_2 = -100$, $T_2 = -100$ kPa; initial values: $T_1 = T_2$. (**a**) Large strain cycle without intergranular strain (1.8). (**b**) Small strain cycle without intergranular strain (−is) and with intergranular strain (+is): $m_r = 5$, $R = 10^{-4}$, $\delta_0 = -R$

this concept in our one-dimensional models, we use a general form of hypoplastic relations

$$\dot{T} = LD + N|D|\,, \tag{1.9}$$

e.g., with

$$L = C_1 T \quad \text{and} \quad N = C_2 T \tag{1.10}$$

for oedometric boundary conditions.

The intergranular strain is introduced as additional state variable $Q = \delta$ obeying the evolution equation

$$\dot{\delta} = \begin{cases} \left[1 - \dfrac{|\delta|}{R}\right] D & \text{for } \delta \cdot D > 0\,, \\ D & \text{for } \delta \cdot D \le 0\,. \end{cases} \tag{1.11}$$

Here, R is a material constant which bounds the intergranular strain $-R \le \delta \le R$.

The stress rate is

$$\dot{T} = MD\,, \tag{1.12}$$

with M being a linear combination of N and L

$$M = \begin{cases} (m_R - \rho m_R + \rho)L + \rho N \frac{|\delta|}{\delta} & \text{for } \delta \cdot D > 0\,, \\ m_R L & \text{for } \delta \cdot D \le 0 \end{cases} \tag{1.13}$$

with $\rho = |\delta|/R$, and m_R denoting an additional material constant.

In the case of compression (where $D < 0$) the intergranular strain eventually approaches the limit $\delta = -R$ and $\rho = 1$. For further compression ($\delta \cdot D \ge 0$) the equation of the stress rate, (1.12) with (1.13), reduces to the original hypoplastic relation (1.9)

$$\dot{T} = MD = LD + N\frac{|\delta|}{\delta}D = LD + N|D|\,, \tag{1.14}$$

since $|\delta|/\delta = -1$ and $-D = |D|$ for $D < 0$. An example of the behavior of this model is given in Fig. 1.4b.

The two evolution equations (1.12) and (1.11) constitute a system of coupled differential equations of the form

$$\dot{T} = h(T, \delta, D)\,, \tag{1.15a}$$

$$\dot{\delta} = k(\delta, D)\,. \tag{1.15b}$$

We will use the generic notation

$$\mathbf{y}'(\tau) = \mathbf{f}\big(\mathbf{y}(\tau)\big) \tag{1.16}$$

for the numerical treatment of such systems. The prime denotes differentiation with respect to the independent variable τ. The vector $\mathbf{y}$ collects all state and additional state variables, i.e., for (1.15)

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} T \\ \delta \end{bmatrix}, \qquad \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} h(T, \delta, D) \\ k(\delta, D) \end{bmatrix} = \begin{bmatrix} h(y_1, y_2, D) \\ k(y_2, D) \end{bmatrix}. \tag{1.17}$$

Note that the stretching D is regarded here as a given parameter and not as a variable.

1.3.2 Explicit Integration

The integration of initial value problems[2] of differential equations

$$\mathbf{y}'(\tau) = \mathbf{f}(\mathbf{y}(\tau)), \tag{1.18a}$$

$$\mathbf{y}(0) = \mathbf{y}_0 \tag{1.18b}$$

can be an involved task. Apart from very particular situations, a closed form solution will not be available. Therefore, one has to resort to numerical methods. For a prescribed step size $\Delta\tau_n$ and a given approximation $\mathbf{y}_n$ to $\mathbf{y}(\tau)$ at time $\tau = \tau_n$, a numerical method provides an approximation $\mathbf{y}_{n+1}$ at time $\tau_{n+1} = \tau_n + \Delta\tau_n$. Starting from the given initial value at time $\tau_0 = 0$, the method computes numerical approximations to the solution at discrete times $\tau_1, \tau_2, \tau_3, \ldots$

The integration of the constitutive equations of rate type is just a subtask in a complex finite element simulation. Needless to say that this subtask should be carried out in a reliable and efficient way. Explicit schemes construct the numerical approximation by using *explicit* evaluations of the right-hand side of (1.18) only, in contrast to implicit methods, which also require evaluations at initially unknown states.[3] The simplest representative of explicit schemes for solving (1.18) is the explicit (or forward) Euler method.

1.3.2.1 Explicit Euler Method

In quite a few areas of science and engineering, the explicit Euler method

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \Delta\tau_n \mathbf{f}(\mathbf{y}_n) \tag{1.19}$$

is still employed for the numerical integration of (1.18), sometimes even with a prescribed constant step size $\Delta\tau_n = \Delta\tau$. The use of the explicit Euler scheme is mainly motivated by the simple structure of (1.19), which allows a straightforward implementation. The main shortcomings of this method are the following ones:

1. The explicit Euler method is of first order only. This means that halving the step size will reduce the global errors by a factor of two, only. As a consequence, the method might require a large number of steps to meet the prescribed accuracy requirements.

[2] In the theory of initial value problems, it is common to call the independent variable τ time. We will follow this tradition in our article. In all the applications we have in mind, however, the role of τ is *not* that of a physical time but of a variable that parameterizes the loading and unloading processes.

[3] These states are then to be determined by some iterative process, which might be time consuming.

2. It does not supply an estimate of the committed errors. Such an error estimate, however, is indispensable for selecting the time step size $\Delta\tau_n$ adaptively.
3. It is not suited for stiff problems.

These are the reasons why we propose here other schemes.

1.3.2.2 Richardson Extrapolation of the Explicit Euler Method

Starting from simple Euler steps, we will construct a second-order method by a procedure called extrapolation. Although Richardson extrapolation is a general means to increase the order of numerical approximations, we will illustrate it here just for our particular situation.

Let $\mathbf{y}_n$ be the numerical approximation at time τ_n and $\mathbf{z}(\tau)$ the exact solution of (1.18a) satisfying $\mathbf{z}(\tau_n) = \mathbf{y}_n$, see Fig. 1.6. For our construction, we need to perform an Euler step of length $\Delta\tau_n$

$$\mathbf{u} = \mathbf{y}_n + \Delta\tau_n \mathbf{f}(\mathbf{y}_n), \tag{1.20a}$$

and two consecutive Euler steps of length $\Delta\tau_n/2$

$$\mathbf{v} = \mathbf{y}_n + \frac{\Delta\tau_n}{2}\mathbf{f}(\mathbf{y}_n), \qquad \mathbf{w} = \mathbf{v}_n + \frac{\Delta\tau_n}{2}\mathbf{f}(\mathbf{v}_n). \tag{1.20b}$$

We note that this requires two evaluations of the right-hand side function $\mathbf{f}$.

The Taylor expansion of the solution $\mathbf{z}$ with initial value $\mathbf{z}(\tau_n) = \mathbf{y}_n$ is given by

$$\begin{aligned}\mathbf{z}(\tau_{n+1}) &= \mathbf{z}(\tau_n) + \Delta\tau_n \mathbf{z}'(\tau_n) + \frac{\Delta\tau_n^2}{2}\mathbf{z}''(\tau_n) + \mathcal{O}(\Delta\tau_n^3) \\ &= \mathbf{y}_n + \Delta\tau_n \mathbf{f}(\mathbf{y}_n) + \frac{\Delta\tau_n^2}{2}\mathbf{f}'(\mathbf{y}_n)\mathbf{f}(\mathbf{y}_n) + \mathcal{O}(\Delta\tau_n^3),\end{aligned} \tag{1.21}$$

where we have used the differential equation (1.18a) and its derivative

$$z''(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau}\mathbf{z}'(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau}\mathbf{f}\big(\mathbf{z}(\tau)\big) = \mathbf{f}'\big(\mathbf{z}(\tau)\big)\mathbf{z}'(\tau) = \mathbf{f}'\big(\mathbf{z}(\tau)\big)\mathbf{f}\big(\mathbf{z}(\tau)\big)$$

with $\mathbf{f}'$ denoting the Jacobian of $\mathbf{f}$. The Taylor expansion of $\mathbf{w}$ is given by

$$\begin{aligned}\mathbf{w} &= \mathbf{v}_n + \frac{\Delta\tau_n}{2}\mathbf{f}(\mathbf{v}_n) \\ &= \mathbf{y}_n + \frac{\Delta\tau_n}{2}\mathbf{f}(\mathbf{y}_n) + \frac{\Delta\tau_n}{2}\mathbf{f}\left(\mathbf{y}_n + \frac{\Delta\tau_n}{2}\mathbf{f}(\mathbf{y}_n)\right) \\ &= \mathbf{y}_n + \Delta\tau_n \mathbf{f}(\mathbf{y}_n) + \frac{\Delta\tau_n^2}{4}\mathbf{f}'(\mathbf{y}_n)\mathbf{f}(\mathbf{y}_n) + \mathcal{O}(\Delta\tau_n^3).\end{aligned} \tag{1.22}$$

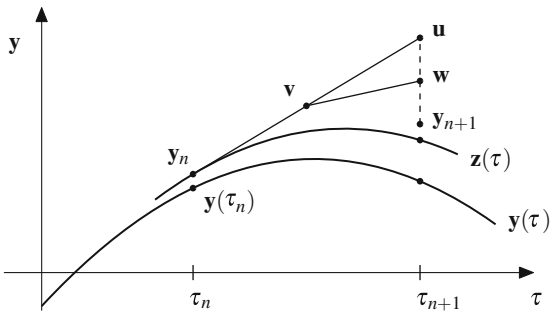

Fig. 1.6 Euler step **u**, half steps **v** and **w**, and the extrapolated value $\mathbf{y}_{n+1}$

Comparing the expansion (1.21) with (1.20a) and (1.22), respectively, shows that both, **u** and **w** are first-order approximations to $\mathbf{z}(\tau_{n+1})$:

$$\mathbf{u} = \mathbf{z}(\tau_{n+1}) - \frac{1}{2}\mathbf{f}'(\mathbf{y}_n)\mathbf{f}(\mathbf{y}_n)\,\Delta\tau_n^2 + \mathscr{O}(\Delta\tau_n^3), \tag{1.23a}$$

$$\mathbf{w} = \mathbf{z}(\tau_{n+1}) - \frac{1}{4}\mathbf{f}'(\mathbf{y}_n)\mathbf{f}(\mathbf{y}_n)\,\Delta\tau_n^2 + \mathscr{O}(\Delta\tau_n^3). \tag{1.23b}$$

Now, the idea is to take the combination

$$\mathbf{y}_{n+1} = 2\mathbf{w} - \mathbf{u}, \tag{1.24}$$

which eliminates the leading error terms in (1.23). The resulting method (1.20), (1.24) is called *Richardson extrapolation* of the explicit Euler method.

The difference to the local solution

$$\mathbf{y}_{n+1} - \mathbf{z}(\tau_{n+1}) = \mathscr{O}(\Delta\tau_n^3) \tag{1.25}$$

is called local error of the method. A standard argument [11] now shows that the resulting method is second-order convergent.

For later reference, we reformulate (1.24) as a two-stage Runge–Kutta method

$$\begin{aligned}
\mathbf{Y}_{n,1} &= \mathbf{y}_n\,,\\
\mathbf{Y}_{n,2} &= \mathbf{y}_n + \Delta\tau_n a_{21}\mathbf{f}(\mathbf{Y}_{n,1})\,,\\
\mathbf{y}_{n+1} &= \mathbf{y}_n + \Delta\tau_n b_1\mathbf{f}(\mathbf{Y}_{n,1}) + \Delta\tau_n b_2\mathbf{f}(\mathbf{Y}_{n,2})\,.
\end{aligned} \tag{1.26}$$

For the choice $a_{21} = \frac{1}{2}$, $b_1 = 0$, and $b_2 = 1$, the above method coincides with Richardson extrapolation of the explicit Euler method, since $\mathbf{v} = \mathbf{Y}_{n,2}$ and

$$\begin{aligned}
2\mathbf{w} - \mathbf{u} &= 2\mathbf{v} + \Delta\tau_n\mathbf{f}(\mathbf{v}) - \mathbf{y}_n - \Delta\tau_n\mathbf{f}(\mathbf{y}_n)\\
&= \mathbf{y}_n + \Delta\tau_n\mathbf{f}(\mathbf{Y}_{n,2})\,.
\end{aligned}$$

We note that this method was already proposed by Runge in 1895 (see [11]).

1.3.3 Adaptivity and Error Control

Apart from its better accuracy, the auxiliary states computed in the extrapolation process can also be used to estimate the local error, that is, the error committed in one step. Although this information cannot be used directly to bound the global error, it is nevertheless very useful to determine an appropriate step size sequence for the integration. Our approach outlined in this section is that of [5, 7, 11].

The term

$$C = \frac{1}{2}\|\mathbf{f}'(\mathbf{y}_n)\mathbf{f}(\mathbf{y}_n)\|$$

is the norm of the leading error term in (1.23a). In order to estimate this term, we take the following difference of the auxiliary states (1.20)

$$\text{EST} = \|\mathbf{w} - \mathbf{u}\| \approx \frac{C}{2}\Delta\tau_n^2. \tag{1.27}$$

The estimated error EST is correct up to the third-order terms in $\Delta\tau_n$, see (1.22). For a user-supplied tolerance TOL, the optimal step size $\Delta\tau_{\text{opt}}$ in the present step would have been

$$\frac{C}{2}\Delta\tau_{\text{opt}}^2 = \text{TOL}\,. \tag{1.28}$$

To compute $\Delta\tau_{\text{opt}}$, we divide (1.28) by (1.27) to eliminate C. This gives an approximation for the optimal step size

$$\Delta\tau_{\text{opt}} \approx \Delta\tau_n \sqrt{\frac{\text{TOL}}{\text{EST}}}\,.$$

With these ingredients, we build up a simple step size control. Starting from the accepted state $\mathbf{y}_n$ and the predicted step size $\Delta\tau_n$, we compute the states $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, and from this the error estimate EST as in (1.27). Further, we compute a new step size according to

$$\Delta\tau_{\text{new}} = \Delta\tau_n \cdot \min\left(\kappa_{\text{I}},\ \max\left(\kappa_{\text{D}},\ 0.9 \cdot \sqrt{\frac{\text{TOL}}{\text{EST}}}\right)\right), \tag{1.29}$$

where the constants κ_{D} and κ_{I} limit the step size change (maximum decrease and increase, respectively). A common choice is $\kappa_{\text{D}} = 0.2$ and $\kappa_{\text{I}} = 2$. If the estimated error EST is smaller than the prescribed tolerance TOL, the step is accepted

$$\mathbf{y}_{n+1} = 2\mathbf{w} - \mathbf{u}$$

and the next step size is chosen as $\Delta\tau_{n+1} = \Delta\tau_{\text{new}}$. If the estimated error EST is larger than TOL, however, we reject the step and redo it with $\Delta\tau_n = \Delta\tau_{\text{new}}$. The factor 0.9 in (1.29) is a safety factor that accounts for the neglected higher-order terms of (1.27).

For making the first step, the method requires a starting step size. The choice is not very critical, as a viable step size will be determined by the code automatically, as described above.

The error in (1.27) can be estimated in any norm. A common choice is the maximum norm

$$\|\mathbf{w} - \mathbf{u}\| = \max_{i=1,\dots,m} \left| \frac{w_i - u_i}{s_i} \right| \tag{1.30}$$

with the scaling factors

$$s_i = a_i + r_i \cdot \max\left(|(\mathbf{y}_n)_i| , |(\mathbf{y}_{n+1})_i| \right).$$

The parameters a_i and r_i are used to fine-tune the error estimate. Taking $a_i = 0$ and $r_i = 1$ results in a relative error estimate. This is important when the absolute value of the corresponding quantity (a stress component, for example) gets considerably larger than 1. On the other hand, this choice is dangerous whenever the solution gets close to zero. In the latter case, the absolute error should be controlled.

Let AERR_i be the lowest resolution of component i requested by the user. Then the choice

$$a_i = \frac{\text{AERR}_i}{\text{TOL}} \tag{1.31}$$

has the following implication. Whenever the first $\log_{10}(\text{TOL}^{-1})$ digits of the solution are correct or the absolute error $|w_i - v_i|$ is less than AERR_i for all i the step is accepted. Otherwise, the error control will enforce the integrator to reject the step. A one-dimensional example illustrating the effect of AERR on the step size selection is given in Table 1.1.

1.3.4 Implicit Integration

In contrast to explicit methods, implicit schemes require the evaluation of the right-hand side of (1.18a) at states that are still to be computed during the step. An archetypical example is the implicit (or backward) Euler method

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \Delta\tau_n \mathbf{f}(\mathbf{y}_{n+1}). \tag{1.32}$$

In order to determine the sought-after state $\mathbf{y}_{n+1}$, one has to solve a system of nonlinear equations. In principle, this can be done by fixed-point iteration or

Table 1.1 One-dimensional example ($m = 1$) of error control with TOL $= 10^{-3}$ and $r = 1$: the step is accepted or rejected depending on the size of AERR

v	w	AERR = 10	AERR = 1
100,005	100,000	Accepted	Accepted
10,005	10,000	Accepted	Accepted
1,005	1,000	Accepted	*Rejected*
105	100	Accepted	*Rejected*

by some Newton-type iterations. Note, however, that fixed-point iteration only converges if the (local) Lipschitz constant of $\mathbf{f}$ times the step size $\Delta\tau_n$ is smaller than one. This is not always the case in applications, but can be enforced by choosing a small time step size. However, such a choice makes the integrator less efficient.

There are situations in which implicit methods perform better, sometimes tremendously better, than explicit ones. Such problems are called *stiff*. In this situation, Newton-type iterations have to be employed to determine the state $\mathbf{y}_{n+1}$.

1.3.4.1 Implicit Euler Method

In order to investigate some properties of the implicit Euler method, we use Taylor expansion of the right-hand side of (1.32). This yields

$$\begin{aligned}\mathbf{y}_{n+1} &= \mathbf{y}_n + \Delta\tau_n \mathbf{f}(\mathbf{y}_{n+1}) \\ &= \mathbf{y}_n + \Delta\tau_n \mathbf{f}\big(\mathbf{y}_n + \Delta\tau_n \mathbf{f}(\mathbf{y}_{n+1})\big) \\ &= \mathbf{y}_n + \Delta\tau_n \mathbf{f}(\mathbf{y}_n) + \Delta\tau_n^2 \mathbf{f}'(\mathbf{y}_n)\mathbf{f}(\mathbf{y}_{n+1}) + \mathscr{O}(\Delta\tau_n^3) \\ &= \mathbf{y}_n + \Delta\tau_n \mathbf{f}(\mathbf{y}_n) + \Delta\tau_n^2 \mathbf{f}'(\mathbf{y}_n)\mathbf{f}(\mathbf{y}_n) + \mathscr{O}(\Delta\tau_n^3).\end{aligned} \tag{1.33}$$

Comparing this expansion with (1.20a) and (1.21) shows at once that the implicit Euler method has the *same* accuracy as the explicit one; the leading error term is just of opposite sign. This shows that one cannot expect higher accuracy by using the implicit Euler method. The main difference to the explicit Euler scheme is its better stability properties for stiff problems. For a detailed discussion, we refer to the literature (see, e.g., the monograph [12]).

1.3.5 Semi-Implicit Integration

Due to the lack of stability, explicit integrators are forced to use unreasonably small time steps when integrating stiff problems. As a consequence, they become computationally inefficient. A remedy would be to resort to implicit methods. However, the required solution of nonlinear systems of equations by Newton-type methods can be a time-consuming task, which again will harm the overall efficiency.

A viable compromise between small time steps and expensive iterations are so-called semi-implicit methods. The semi-implicit Euler method

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \Delta\tau_n \mathbf{f}(\mathbf{y}_n) + \Delta\tau_n \mathbf{f}'(\mathbf{y}_n)(\mathbf{y}_{n+1} - \mathbf{y}_n) \tag{1.34}$$

can be seen as the result of one Newton iteration applied to (1.32). The method can be written equivalently as

$$\big[\mathbf{I} - \Delta\tau_n \mathbf{f}'(\mathbf{y}_n)\big](\mathbf{y}_{n+1} - \mathbf{y}_n) = \Delta\tau_n \mathbf{f}(\mathbf{y}_n)\,. \tag{1.35}$$

In contrast to the implicit Euler method, it only requires the solution of linear systems of equations.

1.3.5.1 Richardson Extrapolation of the Semi-Implicit Euler Method

Again, one can use Richardson extrapolation to improve the order and accuracy of the semi-implicit Euler scheme. We start off from the basic integration step

$$\big[\mathbf{I} - \Delta\tau_n \mathbf{f}'(\mathbf{y}_n)\big](\mathbf{u} - \mathbf{y}_n) = \Delta\tau_n \mathbf{f}(\mathbf{y}_n) \tag{1.36a}$$

which defines the state $\mathbf{u}$. Next, we construct the auxiliary states $\mathbf{v}$ and $\mathbf{w}$ by performing two consecutive steps with half the step size

$$\Big[\mathbf{I} - \frac{\Delta\tau_n}{2} \mathbf{f}'(\mathbf{y}_n)\Big](\mathbf{v} - \mathbf{y}_n) = \frac{\Delta\tau_n}{2} \mathbf{f}(\mathbf{y}_n), \tag{1.36b}$$

$$\Big[\mathbf{I} - \frac{\Delta\tau_n}{2} \mathbf{f}'(\mathbf{y}_n)\Big](\mathbf{w} - \mathbf{v}) = \frac{\Delta\tau_n}{2} \mathbf{f}(\mathbf{v})\,. \tag{1.36c}$$

In order to save computational time, we have used the same Jacobian $\mathbf{f}'(\mathbf{y}_n)$ in all three steps. Note that

$$\big[\mathbf{I} - \Delta\tau \mathbf{f}'(\mathbf{y})\big]^{-1} = \mathbf{I} + \Delta\tau \mathbf{f}'(\mathbf{y}) + \mathcal{O}(\Delta\tau^2).$$

Taylor expansion now shows that

$$\begin{aligned}
\mathbf{u} &= \mathbf{y}_n + \Delta\tau_n \mathbf{f}(\mathbf{y}_n) + \Delta\tau_n^2 \mathbf{f}'(\mathbf{y}_n)\mathbf{f}(\mathbf{y}_n) + \mathcal{O}(\Delta\tau_n^3)\,,\\
\mathbf{v} &= \mathbf{y}_n + \frac{\Delta\tau_n}{2} \mathbf{f}(\mathbf{y}_n) + \frac{\Delta\tau_n^2}{4} \mathbf{f}'(\mathbf{y}_n)\mathbf{f}(\mathbf{y}_n) + \mathcal{O}(\Delta\tau_n^3)\,,\\
\mathbf{w} &= \mathbf{v} + \frac{\Delta\tau_n}{2} \mathbf{f}(\mathbf{v}) + \frac{\Delta\tau_n^2}{4} \mathbf{f}'(\mathbf{y}_n)\mathbf{f}(\mathbf{v}) + \mathcal{O}(\Delta\tau_n^3)\\
&= \mathbf{y}_n + \Delta\tau_n \mathbf{f}(\mathbf{y}_n) + \frac{3\Delta\tau_n^2}{4} \mathbf{f}'(\mathbf{y}_n)\mathbf{f}(\mathbf{y}_n) + \mathcal{O}(\Delta\tau_n^3)\,.
\end{aligned} \tag{1.37}$$

These expansions reveal at once that the extrapolated value

$$\mathbf{y}_{n+1} = 2\mathbf{w} - \mathbf{u} \tag{1.38}$$

approximates the (local) solution $\mathbf{z}(\tau)$ with initial value $\mathbf{z}(\tau_n) = \mathbf{y}_n$ by one order higher than the semi-implicit Euler method

$$\mathbf{y}_{n+1} - \mathbf{z}_{n+1} = \mathscr{O}(\Delta\tau_n^3)\,.$$

The resulting method (1.36), (1.38) is consequently a second-order method. It is called *Richardson extrapolation* of the semi-implicit Euler method.

Error control and step size selection for this method are performed in exactly the same way as for the Richardson extrapolation of the explicit Euler method.

1.3.6 Examples

We consider first a one-dimensional compression test with the one-dimensional hypoplastic model (1.7) for loading ($D < 0$)

$$\dot{T} = h(T, D) = (C_1 - C_2)TD = KTD = KT\dot{\varepsilon} \tag{1.39}$$

with the constant $K = -2{,}000$ and the initial stress $T(0) = T_0 = -100\,\text{kPa}$. Starting from that initial stress and the initial strain $\varepsilon(0) = 0$, the analytic time integration of (1.39) yields

$$T(t) = T_0 e^{KDt} = T_0 e^{K\varepsilon}\,. \tag{1.40}$$

The stress at the end of a strain increment $\Delta\varepsilon = \varepsilon - \varepsilon(0) = -1.1513\times10^{-3}$ is

$$T^{\text{an}} = T_0 e^{K\varepsilon} = -100 \cdot e^{2\cdot1.1513} = -1{,}000\,\text{kPa}\,. \tag{1.41}$$

For comparing the numerical results T^{num} with the analytical solution (1.41), the relative error

$$\text{err}\,T = \frac{T^{\text{num}} - T^{\text{an}}}{T^{\text{an}}} \tag{1.42}$$

is used. The required step sizes for various numerical integration schemes to achieve the same relative error in stress are summarized in Table 1.2.

Next we investigate the behavior of two adaptive methods for solving the equation of intergranular strains (1.11). For $\delta \cdot D > 0$ this is known to be a numerically stiff differential equation. The analytic solution for $\delta(0) = 0$ is given by

$$\delta(t) = R\left(1 - \exp\frac{-Dt}{R}\right) \tag{1.43}$$

Table 1.2 Numerical time integration of (1.39): required steps to achieve the same relative error err $T = \pm 1.3 \times 10^{-3}$

Numerical method	Required steps	Rejected steps
Explicit Euler	2,000	–
Implicit Euler	2,000	–
Semi-implicit integration	2,000	–
Adaptive explicit Richardson	40	3
Adaptive semi-implicit Richardson	42	3

The error tolerances of the adaptive methods are: TOL $= 10^{-3}$, AERR $= 10^{-6}$. Rejected steps occur in these methods since the initial step size is chosen to be the whole strain increment, i.e., an integration with one step was initially tried

for $\delta > 0$ and $D > 0$, and

$$\delta(t) = -R\left(1 - \exp\frac{Dt}{R}\right) \tag{1.44}$$

for $\delta < 0$ and $D < 0$. Evaluating (1.44) for a strain of $\varepsilon = 8 \times 10^{-3}$ yields an intergranular strain δ that deviates from R by less than 10^{-16}, i.e., numerically, it holds $\delta = -R$. Integrating (1.11) up to the same strain takes 51 steps with the adaptive explicit Richardson method and only 16 steps with the adaptive semi-implicit Richardson method. The relative errors are -4.7×10^{-3} and 2.2×10^{-10}, respectively. As both methods start with the whole strain as initial step size, some steps are rejected at the starting point, namely 5 for the explicit and 8 for the semi-implicit method. The chosen error tolerances are the same as in the previous example.

1.4 Extensions to Elasto-Plastic Models

In this section, we extend our investigation to elasto-plastic models. In mathematical formulation, we obtain differential-algebraic systems of index 2. For their numerical integration, we propose a half-explicit Runge–Kutta method of order two.

1.4.1 Nonlinear Elasticity, Nonlinear Isotropic Hardening

Elasto-plastic models are based on the assumption of an additive decomposition of the total strain

$$\varepsilon = \varepsilon_e + \varepsilon_p\,, \tag{1.45}$$

where ε_e denotes the elastic and ε_p the plastic strain (see, e.g., [18]). The stress response in a linear elastic model is

$$T = E\varepsilon_e = E(\varepsilon - \varepsilon_p) \tag{1.46}$$

with the elastic modulus E. This relation, written in rate form, is

$$\dot{T} = E(D - D_p)\,, \tag{1.47}$$

where D_p denotes the plastic stretching. A yield function f_Y is defined to distinguish between elastic and plastic behavior. For an ideal plastic model this function has the form

$$f_Y(T) = |T| - T_Y \leq 0 \tag{1.48}$$

with the yield stress T_Y. If $T < T_Y$, i.e. $f(T) < 0$, the response is purely elastic and the plastic stretching is zero, $D_p = 0$. If $f(T) = 0$ two cases are possible: loading or unloading. Loading takes place if an elastic trial step ($D_p = 0$) would enlarge $|T|$. This can be formulated in our one-dimensional model simple as $D \cdot T > 0$. For loading the plastic stretching evolves according to

$$D_p = \gamma \frac{\partial f_Y}{\partial T} = \gamma \operatorname{sign} T \tag{1.49}$$

with $\gamma > 0$, while the stress remains on the yield surface, i.e. $T(t) = T_Y$ or

$$\dot{f}_Y(T) = 0\,. \tag{1.50}$$

In the case of unloading, here $D \cdot T < 0$, the response is purely elastic. Note that (1.49) generally holds if $\gamma = 0$ is used in elastic steps.

Isotropic hardening can be modeled with an additional state variable $\alpha \geq 0$ which increases during plastic flow and modifies the yield condition to

$$f_Y(T, \alpha) = |T| - (T_Y + K\alpha) \leq 0 \tag{1.51}$$

with the plastic modulus K. The evolution of α can simply be

$$\dot{\alpha} = |D_p|\,, \tag{1.52}$$

which is a linear hardening model. We will use a nonlinear hardening model with an evolution equation similar to that of the intergranular strain

$$\dot{\alpha} = \left[1 - \frac{\alpha}{R}\right] |D_p|\,. \tag{1.53}$$

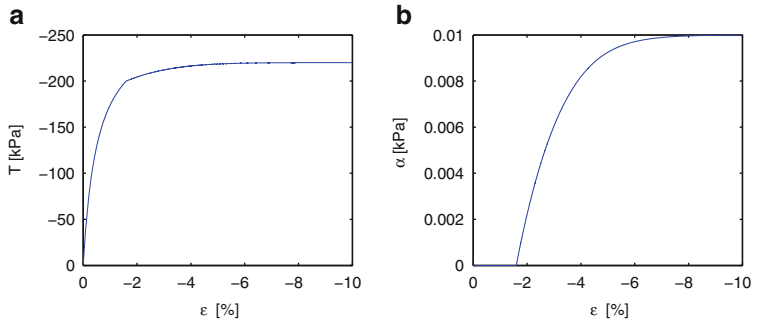

Fig. 1.7 Unconfined uniaxial compression tests: loading ($D < 1$) with an elasto-plastic model: nonlinear elasticity (1.55) and nonlinear isotropic hardening (1.53). The material constants are: $E = 50,000$ kPa, $K = 2,000$ kPa, $T_Y = 200$ kPa, $R_f = 0.75$, $R = 0.01$. (**a**) Stress–strain relationship. (**b**) Evolution of the hardening parameter α

In the case of loading and $f_Y(T, \alpha) = 0$ (plastic flow) the stress must remain on the growing yield surface, i.e.,

$$\dot{f}_Y(T, \alpha) = 0\,, \tag{1.54}$$

and the plastic flow is again defined by (1.49), see [18].

Some elasto-plastic models in geotechnical engineering employ nonlinear elastic relations to model a nonlinear stress–strain response in the first loading. For example, the hardening soil model [17] makes use of the hyperbolic relation of Duncan and Chang [2]. Such a relation can be formulated in a one-dimensional model as

$$\dot{T} = E\left(1 - R_f \frac{T}{T_Y}\right)^2 (D - D_p) \tag{1.55}$$

with T_Y / R_f being the asymptote of the hyperbolic stress–strain relation. The behavior of such a model combined with nonlinear hardening is shown in Fig. 1.7. The unloading / reloading behavior of soil does not show the same strong nonlinearity as the first loading. To model this, linear elasticity is used in such cases, with a Young's modulus $E_{\text{ur}} > E$, which is used when the actual stress is lower than the stress ever been applied, i.e. $|T| < \max |T(t)|$.

For the numerical treatment of the constitutive model we again rewrite the evolution equations in generic form. In case of $f_Y = 0$ and further loading, we obtain a so-called index 2 problem, a combination of differential equations and nonlinear constraints:

$$\mathbf{y}' = \mathbf{f}(\mathbf{y}, \mathbf{z})\,, \tag{1.56a}$$

$$\mathbf{0} = \mathbf{g}(\mathbf{y})\,. \tag{1.56b}$$

Let us illustrate this general form with the help of the example that we just discussed. Using (1.55), (1.49), and (1.53) we set

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} T \\ \alpha \end{bmatrix}, \qquad z = \gamma\,, \tag{1.57}$$

$$\mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} E\left(1 - R_f \frac{y_1}{T_Y}\right)^2 (D - \gamma \operatorname{sign} y_1) \\ \left[1 - \frac{y_2}{R}\right]\gamma \end{bmatrix}, \tag{1.58}$$

and from (1.51) we obtain

$$g = f_Y(y_1, y_2) = |y_1| - (T_Y + K y_2)\,. \tag{1.59}$$

The elastic case (first loading as long as $f_Y < 0$) can be integrated using (1.16) from the pure rate models with $\mathbf{f}$ defined in (1.58) and $\gamma = 0$. For unloading and reloading we can again use (1.58) with $\gamma = 0$ but have to assign $R_f = 0$ and $E = E_{\text{ur}}$.

1.4.2 Event Location

When switching from a nonlinear elastic to a plastic region, one has to determine the transition point with sufficient accuracy in order not to spoil the overall accuracy. This is done by event location. Let g be a state-dependent function that changes sign at the sought-after transition point. Then one has to integrate the elastic problem

$$\mathbf{y}'(\tau) = \mathbf{f}\big(\mathbf{y}(\tau)\big) \tag{1.60}$$

as long as $g\big(\mathbf{y}(\tau)\big) < 0$. A simple strategy consists in integrating (1.60) until the numerical solution indicates a sign change of g. Assume that $g(\mathbf{y}_n) < 0$, but $g(\mathbf{y}_{n+1}) > 0$, which indicates that the transition takes place in the time interval $[\tau_n, \tau_{n+1}]$. The simplest possibility to determine the transition time and its state is using an interpolation procedure. Recall that we are given $\mathbf{y}_n$ and $\mathbf{y}_{n+1}$. These two values, together with the derivative $\mathbf{f}(\mathbf{y}_n)$ determine uniquely a quadratic interpolation polynomial.

Let $\theta = (\tau - \tau_n)/\Delta\tau_n$. Then this Hermitian interpolation polynomial has the form

$$\mathbf{p}(\theta) = \mathbf{a} + \theta\mathbf{b} + \theta(\theta - 1)\mathbf{c}, \quad 0 \le \theta \le 1, \tag{1.61}$$

where the coefficients $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are determined by the interpolation conditions

$$\begin{aligned} \mathbf{p}(0) &= \mathbf{a} = \mathbf{y}_n\,, \\ \mathbf{p}(1) &= \mathbf{a} + \mathbf{b} = \mathbf{y}_{n+1}\,, & \mathbf{b} &= \mathbf{y}_{n+1} - \mathbf{y}_n\,, \\ \mathbf{p}'(0) &= \mathbf{b} - \mathbf{c} = \Delta\tau_n \mathbf{f}(\mathbf{y}_n)\,, & \mathbf{c} &= \mathbf{y}_{n+1} - \mathbf{y}_n - \Delta\tau_n \mathbf{f}(\mathbf{y}_n)\,. \end{aligned}$$

Inserting these values into (1.61) gives

$$\mathbf{p}(\theta) = \mathbf{y}_n + \theta\big(\mathbf{y}_{n+1} - \mathbf{y}_n\big) + \theta(\theta - 1)\big(\mathbf{y}_{n+1} - \mathbf{y}_n - \Delta\tau_n \mathbf{f}(\mathbf{y}_n)\big)\,, \tag{1.62}$$

which is a second-order approximation to $\mathbf{y}(\tau_n + \theta\Delta\tau_n)$, if the underlying integration scheme is of second order. Any root finding algorithm can be used to determine a root $\theta_\star$ of $g\big(\mathbf{p}(\theta)\big)$ in $[0, 1]$. Since the evaluation of (1.62) is cheap, the bisection algorithm is efficient and reliable. The detected root finally provides us with a second-order approximation to the transition time $\tau_\star = \tau_n + \theta_\star\Delta\tau_n$ and the sought-after state $\mathbf{y}_\star = \mathbf{p}(\theta_\star)$ with $g(\mathbf{y}_\star) = 0$.

1.4.3 Time Integration of Index 2 Problems

We have seen in Sect. 1.4.1 that the following combination of differential and nonlinear equations

$$\mathbf{y}'(\tau) = \mathbf{f}\big(\mathbf{y}(\tau), \mathbf{z}(\tau)\big)\,, \tag{1.63a}$$

$$\mathbf{0} = \mathbf{g}\big(\mathbf{y}(\tau)\big) \tag{1.63b}$$

arises in the description of elasto-plastic models. We always assume that the initial value $\mathbf{y}(0) = \mathbf{y}_0$ is *consistent* with the problem, in particular that the constraint $\mathbf{g}(\mathbf{y}_0) = \mathbf{0}$ is satisfied. The above problem is a differential equation for the state variables $\mathbf{y}$ that, in addition, have to fulfil a constraint. This is made possible by the nonlinear control variable $\mathbf{z}$.

Differentiating the constraint (1.63b) with respect to τ and inserting the differential equation yields a nonlinear system of equations

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\mathbf{g}(\mathbf{y}(\tau)) = \frac{\partial \mathbf{g}}{\partial \mathbf{y}}\big(\mathbf{y}(\tau)\big)\mathbf{y}'(\tau) = \frac{\partial \mathbf{g}}{\partial \mathbf{y}}\big(\mathbf{y}(\tau)\big)\mathbf{f}\big(\mathbf{y}(\tau), \mathbf{z}(\tau)\big) = \mathbf{0}\,, \tag{1.64}$$

which can be solved (locally) uniquely for $\mathbf{z}(\tau)$, if the so-called index 2 condition

$$\det\left[\frac{\partial \mathbf{g}}{\partial \mathbf{y}}(\mathbf{y})\frac{\partial \mathbf{f}}{\partial \mathbf{z}}(\mathbf{y}, \mathbf{z})\right] \neq 0 \tag{1.65}$$

is satisfied. In such a situation, (1.63) is called an *index 2 problem*.

In the literature (see, e.g., [12, Chap. 7]), several methods for solving (1.63) are proposed. Obviously, explicit methods fail to preserve the constraint (1.63b). Therefore, we consider half-explicit schemes (see [12, Chap. 7.6]).

1.4.3.1 Half-Explicit Euler Method for Index 2 Problems

Starting off from a given step size $\Delta\tau_n$ and a consistent approximation $\mathbf{y}_n$ to the solution of (1.63) at $\tau = \tau_n$, we define the sought-after approximation $\mathbf{y}_{n+1}$ at time τ_{n+1} by an explicit Euler step

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \Delta\tau_n \mathbf{f}(\mathbf{y}_n, \mathbf{z}_{n+1}) \tag{1.66a}$$

with a control $\mathbf{z}_{n+1}$ still to be determined. In order to guarantee that the numerical solution obeys the constraint (1.63b), we require that

$$\mathbf{g}\big(\mathbf{y}_{n+1}\big) = \mathbf{0}. \tag{1.66b}$$

Inserting (1.66a) into (1.66b) gives

$$\mathbf{G}(\mathbf{z}_{n+1}) = \mathbf{g}\big(\mathbf{y}_n + \Delta\tau_n \mathbf{f}(\mathbf{y}_n, \mathbf{z}_{n+1})\big) = \mathbf{0}, \tag{1.67}$$

which has, due to (1.65), a locally unique solution $\mathbf{z}_{n+1}$. We propose to solve the system $\mathbf{G}(\mathbf{z}) = \mathbf{0}$ by some Newton-type iteration. Finally, the sought-after approximation $\mathbf{y}_{n+1}$ is obtained from (1.66a).

1.4.3.2 An Adaptive Second-Order Method

In the same way as before, a second-order method with the option to control the local error can be constructed from the above half-explicit Euler scheme. However, it is simpler to generalize the underlying idea of method (1.66) directly to Runge–Kutta schemes, see [12, Chap. 7.6]. Starting from (1.26), we define the method

$$\begin{aligned}
\mathbf{Y}_{n,1} &= \mathbf{y}_n\,, \\
\mathbf{Y}_{n,2} &= \mathbf{y}_n + \Delta\tau_n a_{21} \mathbf{f}(\mathbf{Y}_{n,1}, \mathbf{Z}_{n,1})\,, \\
\mathbf{0} &= \mathbf{g}(\mathbf{Y}_{n,2})\,, \\
\mathbf{y}_{n+1} &= \mathbf{y}_n + \Delta\tau_n b_1 \mathbf{f}(\mathbf{Y}_{n,1}, \mathbf{Z}_{n,1}) + \Delta\tau_n b_2 \mathbf{f}(\mathbf{Y}_{n,2}, \mathbf{Z}_{n,2})\,, \\
\mathbf{0} &= \mathbf{g}(\mathbf{y}_{n+1})\,.
\end{aligned} \tag{1.68}$$

Again, we make the choice $a_{21} = \frac{1}{2}$, $b_1 = 0$, and $b_2 = 1$, which results in a second-order method. With this choice, the vector $2\mathbf{Y}_{n,2} - \mathbf{Y}_{n,1}$ is a first-order approximation to the exact solution (it is actually a slight perturbation of (1.66)). Therefore, we can use

$$\mathrm{EST} = \|\mathbf{y}_{n+1} - 2\mathbf{Y}_{n,2} + \mathbf{Y}_{n,1}\|$$

as an error estimate. The step size selection is then done as explained in Sect. 1.3.3.

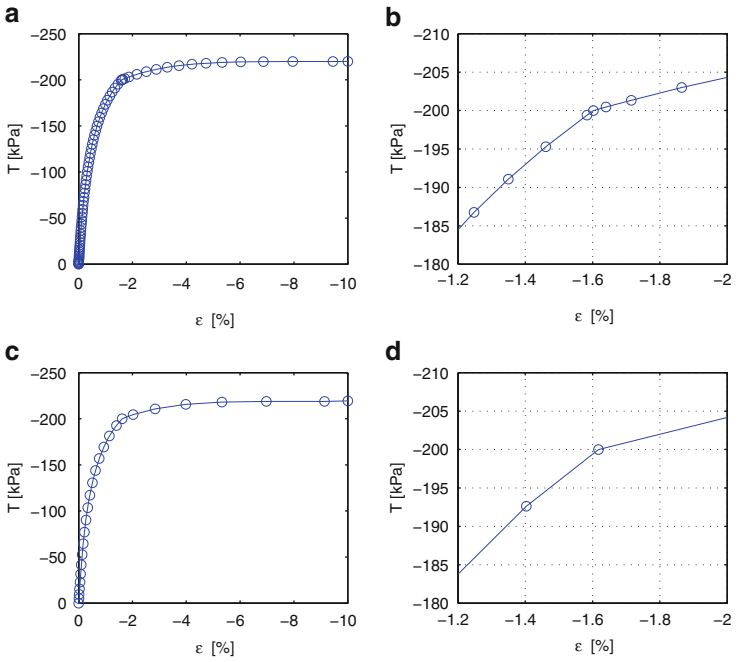

Fig. 1.8 Numerical integration of an elasto-plastic model: nonlinear elasticity (1.55) and nonlinear isotropic hardening (1.53) with the adaptive half-explicit Runge–Kutta method of order two. Time steps are denoted by circles. Material constants: $E = 50,000$ kPa, $K = 2,000$ kPa, $T_Y = 200$ kPa, $R_f = 0.75$, $R = 0.01$; numerical parameter: AERR $= 10^{-6}$, which is also used as tolerance for the Newton iterations. (**a**) TOL $= 10^{-3}$. (**b**) TOL $= 10^{-3}$. (**c**) TOL $= 10^{-2}$. (**d**) TOL $= 10^{-2}$

1.4.4 Example

We apply the above described method to the elasto-plastic model of Sect. 1.4.1. The result of the numerical integration with the half-explicit Runge–Kutta method is shown in Fig. 1.8. The adaptivity of the step size can be seen in Fig. 1.8a, c, where the step size increases until the end of the test. The setting of TOL clearly influences the total amount of taken steps. Dense output is used for the event location $f_Y = 0$. The cutting of the last elastic step can be clearly seen in Fig. 1.8b.

1.5 Consistent Tangent Operator

In the following we discuss some aspects of the consistent tangent operator, and we recall its purpose in finite element methods. For the sake of simplicity, we restrict our presentation to a finite element discretization with one element only. More elements

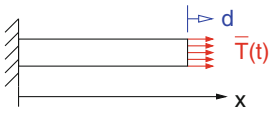

Fig. 1.9 Discretization with one finite element

are treated in [8, 18]. We consider a bar (see Fig. 1.9) with one global degree of freedom, namely the nodal displacement d at the end of the bar. There the bar is loaded by a boundary stress $\overline{T}$.

In the finite element literature a common way of writing the equilibrium condition is

$$F^{\text{int}} - F^{\text{ext}} = 0\,, \tag{1.69}$$

where the external force vector F^{ext} is a function of the load $\overline{T}$, and the internal force vector F^{int} is a function of the internal stress T. In the case of one element with linear shape functions this reduces to

$$\overline{T} - T = 0\,. \tag{1.70}$$

1.5.1 Incremental Loading

In nonlinear finite element calculations the load is applied in increments at discrete times $t_{n+1} = t_n + \Delta t_n$. Let us start with the equilibrated body at time $t = t_n$ with given internal stress T_n. Thus

$$F^{\text{int}}(T_n) - F^{\text{ext}}(\overline{T}_n) = T_n - \overline{T}_n = 0 \tag{1.71}$$

holds.

During the time increment the load is changed from $\overline{T}_n$ by the load increment to $\overline{T}_{n+1} = \overline{T}_n + \Delta\overline{T}_n$. Our task is to find the updated displacement $d_{n+1} = d_n + \Delta d_n$ and the stress T_{n+1} such that (1) the body is equilibrated at time t_{n+1}

$$T_{n+1} - \overline{T}_{n+1} = 0\,, \tag{1.72}$$

and (2) the stress update is compatible with the constitutive model.

The standard solution strategy is an iterative one: the equilibrium equation is solved with the help of a finite element package, and the constitutive model with a solver for ordinary differential equations. The relevant information is passed between these two solvers. The process is iterated until convergence,

1.5.2 The Equilibrium Iteration

We choose a displacement increment Δd_n. Starting from the equilibrated body with the known displacement d_n, we calculate the nodal displacement at the end of the increment

$$d_{n+1} = d_n + \Delta d_n \,. \tag{1.73}$$

The strain in the element at the end of the increment follows from geometrical relations, e.g., the logarithmic strain is

$$\varepsilon_{n+1} = \ln\left(\frac{l_0 + d_{n+1}}{l_0}\right) \tag{1.74}$$

with l_0 being the initial length of the bar. The element is deformed by the strain increment

$$\Delta\varepsilon_n = \varepsilon_{n+1} - \varepsilon_n \,, \tag{1.75}$$

where ε_n is known. The stress increment ΔT_n follows from a time integration of the constitutive model $\dot{T}_n = h(T_n, D_n)$ with the stretching $D_n = \Delta\varepsilon_n/\Delta t_n$. The new stress is given by

$$T_{n+1} = T_n + \Delta T_n \,. \tag{1.76}$$

Inserting this new stress in the equilibrium equation (1.72) generally results in a nonzero right-hand side, the so-called residual R

$$T_{n+1} - \overline{T}_{n+1} = R(d_{n+1}) \,. \tag{1.77}$$

Note that T_{n+1} is a nonlinear function of d_{n+1} due to the nonlinear constitutive model. Thus R is also a nonlinear function of d_{n+1}. We will use Newton's method to find the correct displacement increment, i.e., a zero of R.

1.5.3 Newton's Method

Denote the first guess of the new displacement with d^0_{n+1}. This displacement will cause a stress T^0_{n+1} which is computed with the help of (1.74)–(1.76). Inserting the result into (1.77) gives the first residual $R(d^0_{n+1})$, see Fig. 1.10. Note that the function R can be evaluated at prescribed points, and an analytic formula, however, is not available, in general.

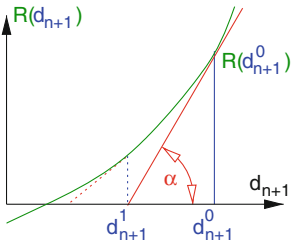

Fig. 1.10 Newton's method

In order to improve our guess, we replace the function R by its tangent at d^0_{n+1}. The slope of this tangent is given by

$$R'(d^0_{n+1}) = \tan\alpha\,, \tag{1.78}$$

where α is the angle of intersection with the axis. The tangent intersects the axis at d^1_{n+1} (see Fig. 1.10), which determines an improved iterate. In order to get an explicit formula for d^1_{n+1}, we consider the triangle with angle α in Fig. 1.10. From the relation

$$\tan\alpha = \frac{R(d^0_{n+1})}{d^0_{n+1} - d^1_{n+1}}\,, \tag{1.79}$$

we easily find an explicit representation of the next iterate

$$d^1_{n+1} = d^0_{n+1} - R'(d^0_{n+1})^{-1} R(d^0_{n+1})\,. \tag{1.80}$$

The derivative of the residual $R'(d^k_{n+1})$ is called *consistent tangent* of the kth equilibrium iteration, as it is consistent with Newton's method. If any other gradient is used, the quadratic convergence of the method is lost.

1.5.4 Consistent Tangent

To calculate the consistent tangent in the kth equilibrium iteration, we have to differentiate the equilibrium equation (1.77) with respect to d_{n+1}, knowing that the external force $F^{\text{ext}} = \overline{T}_{n+1}$ is independent of d_{n+1}. We evaluate this derivative at d^k_{n+1} and get

$$\left.\frac{\mathrm{d}\,R(d_{n+1})}{\mathrm{d}\,d_{n+1}}\right|_{d^k_{n+1}} = R'(d^k_{n+1}) = \underbrace{\frac{\partial T^k_{n+1}}{\partial \varepsilon^k_{n+1}}}_{\text{material}} \cdot \underbrace{\frac{\partial \varepsilon^k_{n+1}}{\partial d^k_{n+1}}}_{\text{geometry}}\,. \tag{1.81}$$

The derivative of the stress with respect to the strain is a material information, which has to be provided by the constitutive relation. The derivative of the strain with respect to the nodal displacement is a geometrical information, which has to be provided by the finite element program. For the logarithmic strain (1.74), this quantity is

$$\frac{\partial \varepsilon_{n+1}^{k}}{\partial d_{n+1}^{k}} = \frac{1}{l_0 + d_{n+1}} . \tag{1.82}$$

In a finite element framework, (1.81) is an element consistent tangent which has to be assembled to a global consistent tangent stiffness [8, 18].

1.5.5 The Jacobian

The material information on the element level that is required to build the consistent tangent operator (1.81) is the so-called Jacobian

$$\frac{\partial T_{n+1}}{\partial \varepsilon_{n+1}} = \frac{\partial (T_{n+1} - T_n)}{\partial (\varepsilon_{n+1} - \varepsilon_n)} = \frac{\partial \Delta T}{\partial \Delta \varepsilon} . \tag{1.83}$$

The Jacobian has to be provided by the subroutine that supplies the constitutive model for the finite element code.

We consider the one-dimensional compression test with the one-dimensional hypoplastic model for loading (1.39) to explain how the Jacobian can be calculated for a constitutive model of the rate type.

1.5.6 Analytical Solution

In this simple example the Jacobian can be found analytically by differentiation. The analytic result of the time integration of the material model is stated in (1.40). From this we deduce that the stress increment for a given strain increment $\Delta\varepsilon = D\Delta t$ is given by

$$\Delta T = T(\Delta t) - T_0 = T_0 \left(e^{K \Delta \varepsilon} - 1 \right) . \tag{1.84}$$

The analytic Jacobian is the derivative of the stress increment with respect to the strain increment, i.e.

$$\frac{\partial \Delta T}{\partial \Delta \varepsilon} = K T_0 e^{K \Delta \varepsilon} = K T_0 e^{K D \Delta t} . \tag{1.85}$$

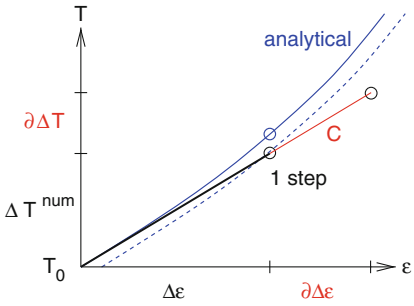

Fig. 1.11 Jacobian with numerical scheme

1.5.7 Numerical Time Integration

The numerical time integration of (1.39) with one explicit Euler step results in the stress increment $\Delta T^{\text{num}} = KT_0\Delta\varepsilon$ (straight line with slope KT_0 in Fig. 1.11), which is different from the analytic solution (1.40) (solid curved line in Fig. 1.11). Calculating the tangent of the constitutive model at the end of the time step will result in the gradient of the dashed line $K(T_0 + \Delta T^{\text{num}})$. The Jacobian required by the finite element code, however, is the gradient of the numerical scheme (straight line) $C = \frac{\partial \Delta T}{\partial \Delta \varepsilon} = KT_0$. We see clearly that the Jacobian and the constitutive tangent are different, even in one-dimensional cases.

1.5.8 Variational Equation

The Jacobian depends on the numerical time integration scheme, i.e., we have to differentiate the numerically computed stress with respect to the strain increment. This can be performed with the help of the variational equation of the constitutive model.

We differentiate the constitutive model $\dot{T} = h(T, D)$ with respect to the stretching D using the chain rule

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial D} = \frac{\partial h}{\partial T}\frac{\partial T}{\partial D} + \frac{\partial h}{\partial D}\,. \tag{1.86}$$

Replacing the right-hand side by (1.39) we arrive at

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial D} = KD\frac{\partial T}{\partial D} + KT\,. \tag{1.87}$$

Next we denote $\partial T/\partial D$ by C and end up with

$$\frac{\mathrm{d}}{\mathrm{d}t}C = KDC + KT, \tag{1.88}$$

which is a differential equation for C.

Equations (1.39) and (1.88) form a coupled system of differential equations. They have to be solved together with the initial conditions

$$T(0) = T_0 \tag{1.89a}$$

$$C(0) = \frac{\partial T}{\partial D}(0) = \frac{\partial T_0}{\partial D} = 0\,. \tag{1.89b}$$

The last identity follows from the fact that $T(0) = T_0$ is the initial condition and therefore independent of D.

1.5.9 Analytic Solution of Stress Update and Jacobian

In this simple case, we can solve (1.39) and (1.88), (1.89) analytically

$$T(t) = T_0 e^{KDt}\,, \tag{1.90}$$

$$C(t) = KT_0 t e^{KDt}\,. \tag{1.91}$$

At the end of the time increment Δt we get the stress update

$$T(\Delta t) = T_0 e^{KD\Delta t} = T_0 e^{K\Delta\varepsilon} = \Delta T + T_0 \tag{1.92}$$

and

$$C(\Delta t) = KT_0 \Delta t e^{KD\Delta t} = \Delta t K T_0 e^{K\Delta\varepsilon}\,. \tag{1.93}$$

Comparing this with (1.85), we see that

$$\frac{C(\Delta t)}{\Delta t} = \frac{\partial \Delta T}{\partial \Delta\varepsilon}\,, \tag{1.94}$$

i.e., we have found the Jacobian.

1.5.10 Numerical Approximation of the Jacobian

Working out the Jacobian analytically for a complex constitutive model can be a tedious task or sometimes even not feasible. Thus we want to use a numerical approximation. For a small variation ϑ we define the approximation B by

$$\frac{\mathrm{d}}{\mathrm{d}\,t} B = \frac{1}{\vartheta}\Big(h(T + \vartheta B, D + \vartheta) - h(T, D)\Big)\,. \tag{1.95}$$

Taylor expansion shows that B satisfies the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}B = \frac{\partial h}{\partial T}\,B + \frac{\partial h}{\partial D} + \mathscr{O}(\vartheta)\,,$$

which is a good approximation to the variational equation for small ϑ.

For our example, substituting (1.39) yields

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}B &= \frac{1}{\vartheta}\Big(KT\cdot(D+\vartheta) + K\vartheta B\cdot(D+\vartheta) - KTD\Big)\\ &= KB\cdot(D+\vartheta) + KT\,.\end{aligned} \tag{1.96}$$

Comparing (1.96) with (1.88) we see that $B = C$ for $\vartheta \to 0$. If ϑ is sufficiently small, $B(\Delta t)/\Delta t$ will be thus a good approximation to the Jacobian.

1.5.11 Example

We investigate a simple but illustrative example to study some numerical aspects. We assume small strains and use the one-dimensional hypoplastic model (1.7) to simulate loading in a one-dimensional compression test (1.39) with the constant $K = -2{,}000$ and the initial stress $T_0 = -100$ kPa. The loading is $\overline{T} = -1{,}000$ kPa. We are searching an ε such that the integration of (1.39) yields an internal stress T which equals the external load $\overline{T}$. It is therefore convenient to consider T also as a function of ε. The initial-boundary value problem (1.70) to be solved reads then

$$T(\varepsilon) - \overline{T} = 0\,. \tag{1.97}$$

The analytic solution of problem (1.97), (1.39) is given by

$$\varepsilon^{\mathrm{an}} = \frac{1}{K}\ln\frac{\overline{T}}{T_0} = -1.1513\times 10^{-3}\,. \tag{1.98}$$

This is obtained from the analytic solution of the time integration of the material model (1.40) with $\varepsilon = Dt$ and the chosen constants.

Equation (1.97) can be solved numerically with standard Newton's method

$$\varepsilon^{i+1} = \varepsilon^{i} - \left(\frac{\mathrm{d}T^{i}}{\mathrm{d}\varepsilon^{i}}\right)^{-1} R^{i}\,, \tag{1.99}$$

where

$$R^{i} = T(\varepsilon^{i}) - \overline{T} \tag{1.100}$$

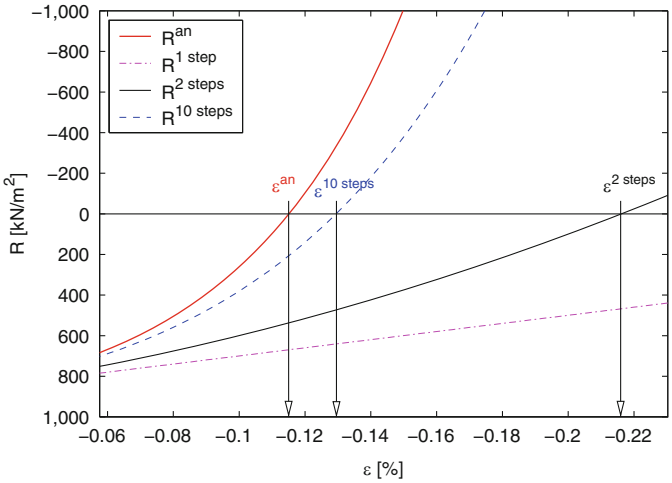

Fig. 1.12 Dependence of the residual R on the numerical method. Here the explicit Euler method was chosen with the number of steps as a parameter. The results are compared with the analytic one

denotes the residual in iteration i. The consistent tangent $\mathrm{d}T^i/\mathrm{d}\varepsilon^i$ is calculated together with the numerical stress $T^i = T(\varepsilon^i)$ by integrating (1.39), (1.88), (1.89) with explicit methods (Sect. 1.3.2). As starting value of the iteration, $\varepsilon^0 = 10^{-3}$ was chosen.

It is important to realize that the residual R as function of the strain ε depends on the time integration method, because ε depends on the time integration method, see Fig. 1.12.

The residual obtained with analytical time integration is

$$R^{\mathrm{an}} = T^{\mathrm{an}}(\varepsilon) - \overline{T} = T_0 e^{K\varepsilon} - \overline{T}\,. \tag{1.101}$$

The residual calculated with the numerical time integration is a composition of the integrator, e.g., for one and two steps explicit Euler steps

$$R^{1\text{ step}} = T_0 + T_0 K\varepsilon - \overline{T}\,, \tag{1.102}$$

$$R^{2\text{ steps}} = T_0 + T_0 K\varepsilon/2 + (T_0 + T_0 K\varepsilon/2)K\varepsilon/2 - \overline{T}\,. \tag{1.103}$$

If the step size sequence for the time integration is kept fixed in all Newton iterations, we stay on the same numerical approximation R^{num} and thus quadratic convergence to the zero of R^{num} is achieved, see Fig. 1.13.

However, we change the problem, i.e., switch from R^{num} to a neighboring $\tilde{R}^{\mathrm{num}}$, whenever we change the step size sequence from one Newton iteration to the other. The change of the step size sequence can be due to the error control of the time integrator. Such a behavior is illustrated in Fig. 1.13 by the sequence ε^0, ε^1, $\tilde{\varepsilon}^2$, etc.

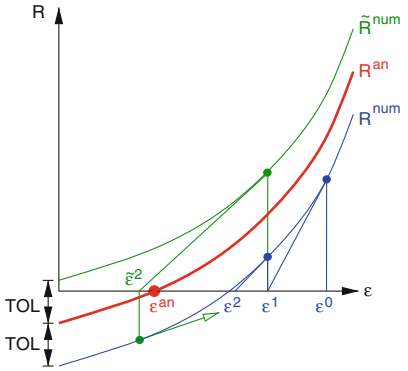

Fig. 1.13 Newton iterations with varying time integration schemes

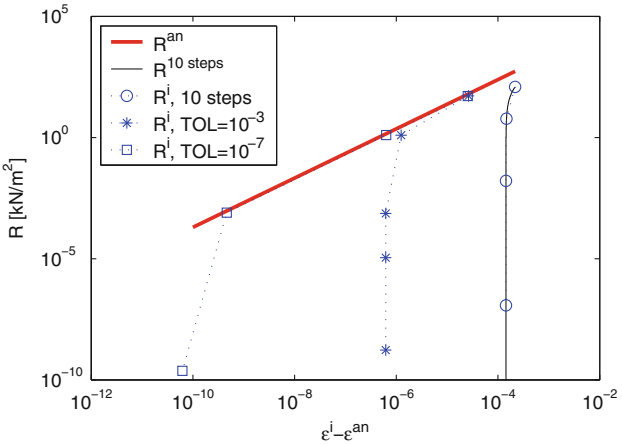

Fig. 1.14 Loss of quadratic convergence using different step size sequence

The iteration might thus become quite irregular whenever the numerical solution is very close to the exact one. However, this is *no problem at all* since such a numerical solution is accurate enough anyway and should therefore be accepted. In any case, the solution obtained with a fixed step size sequence is not better than the adaptive one. It is only a zero of R^{num} which, in general, is different from ε^{an}.

We study this in detail for the Newton iterations in our example, see Fig. 1.14. The analytic function of the residual (1.101) appears as a straight line R^{an} in this double logarithmic diagram. The composition of 10 Euler steps results in the curve $R^{10\ \text{steps}}$. The circles on this curve denote the results of each Newton iteration (R^i, 10 Steps). The convergence is quadratic (the vertical distance between the circles is approximately doubled in each iteration), but the error is large, compare Table 1.3.

Using an adaptive time integration scheme, we end up with a much more accurate solution, see Table 1.3 and Fig. 1.14: $R^i, \text{TOL} = 10^{-3}$ and $R^i, \text{TOL} = 10^{-7}$ (both

Table 1.3 Newton iteration of (1.97) with numerical time integration of (1.39), err $\varepsilon = \frac{\varepsilon^{\text{num}} - \varepsilon^{\text{an}}}{\varepsilon^{\text{an}}}$

10 Euler steps		Extrapolated Euler, TOL $= 10^{-10}$	
It. No.	R^i	It. No.	R^i
1	380.8	1	261.1
2	-124.6	2	-52.07
3	-6.040	3	-1.267
4	-1.629×10^{-2}	4	-8.024×10^{-4}
5	-1.195×10^{-7}	5	-3.218×10^{-10}
6	-5.573×10^{-18}	6	4.629×10^{-20}
err ε	-0.124	err ε	-5.400×10^{-11}

with AERR $= 10^{-6}$). The Newton iteration shows quadratic convergence when the numerical residual R^i is near the analytic residual R^{an}. We loose the quadratic convergence when the residual is of the same order as the tolerance of the time integration. However, the solution ε is then near the analytic one and the iteration can be stopped.

1.6 Fully Three-Dimensional Formulation

The methods of the previous sections can easily be generalized to two and three-dimensional problems. We have worked out this approach in several articles [5,7,8]. However, in order to make the present article self-contained, we repeat here the main line of argumentation.

The equilibrium equations together with the constitutive model form a coupled system of equations. A steady-state solution of this system is usually obtained by operator splitting: the equilibrium equations are solved with the help of a finite element package, and the constitutive model with a solver for ordinary differential equations. To distinguish between the stress in the finite element program and that in the time integration of the constitutive model, we denote these stresses with $\boldsymbol{\sigma}$ and $\mathbf{T}$, respectively.

1.6.1 Consistent Tangent Operator, Jacobian

We start from an equilibrium at time t_n and apply a strain increment $\Delta\boldsymbol{\varepsilon}$ over the time window Δt. For the given initial stress tensor $\mathbf{T}(0) = \boldsymbol{\sigma}(t_n)$ and the strain increment, the constitutive subroutine has to provide the new stress tensor $\boldsymbol{\sigma}(t_n + \Delta t) = \mathbf{T}(\Delta t)$ at time $t_{n+1} = t_n + \Delta t$ as well as its derivative with respect to the strain increment

$$\frac{\partial \Delta\boldsymbol{\sigma}}{\partial \Delta\boldsymbol{\varepsilon}} = \frac{\partial \boldsymbol{\sigma}(t_n + \Delta t)}{\partial \Delta\boldsymbol{\varepsilon}}. \tag{1.104}$$

Due to the incremental form of the solution procedure, the temporal rate of the strain tensor is not known as a function of time. Only its mean value over the chosen time window Δt

$$\mathbf{D} = \frac{\Delta \boldsymbol{\varepsilon}}{\Delta t} \tag{1.105}$$

is available for use in the constitutive model. Assuming that the finite element program only needs the co-rotational parts, e.g. [1], we have to solve the following system of differential equations for $0 \le t \le \Delta t$

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{T} &= \mathbf{h}(\mathbf{T}, \mathbf{D}, \mathbf{Q}) \,, \qquad & \mathbf{T}(0) &= \boldsymbol{\sigma}(t_n) \,, \\ \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{Q} &= \mathbf{k}(\mathbf{T}, \mathbf{D}, \mathbf{Q}) \,, \qquad & \mathbf{Q}(0) &= \mathbf{Q}_0 \,. \end{aligned} \tag{1.106}$$

Here, $\mathbf{Q}$ denotes the additional state variables, and $\mathbf{Q}_0$ are their values at time t_n.

Differentiation of (1.106) with respect to $\mathbf{D}$ yields the variational equations

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathbf{T}}{\partial \mathbf{D}} &= \frac{\partial \mathbf{h}}{\partial \mathbf{T}} \cdot \frac{\partial \mathbf{T}}{\partial \mathbf{D}} + \frac{\partial \mathbf{h}}{\partial \mathbf{Q}} \cdot \frac{\partial \mathbf{Q}}{\partial \mathbf{D}} + \frac{\partial \mathbf{h}}{\partial \mathbf{D}} \,, \qquad & \frac{\partial \mathbf{T}}{\partial \mathbf{D}}(0) &= \mathbf{0} \,, \\ \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathbf{Q}}{\partial \mathbf{D}} &= \frac{\partial \mathbf{k}}{\partial \mathbf{T}} \cdot \frac{\partial \mathbf{T}}{\partial \mathbf{D}} + \frac{\partial \mathbf{k}}{\partial \mathbf{Q}} \cdot \frac{\partial \mathbf{Q}}{\partial \mathbf{D}} + \frac{\partial \mathbf{k}}{\partial \mathbf{D}} \,, \qquad & \frac{\partial \mathbf{Q}}{\partial \mathbf{D}}(0) &= \mathbf{0} \,. \end{aligned} \tag{1.107}$$

Let $\Delta \boldsymbol{\sigma} = \boldsymbol{\sigma}(t_n + \Delta t) - \boldsymbol{\sigma}(t_n)$. In order to get

$$\frac{\partial \Delta \boldsymbol{\sigma}}{\partial \Delta \boldsymbol{\varepsilon}} = \frac{\partial \boldsymbol{\sigma}(t_n + \Delta t)}{\partial \Delta \boldsymbol{\varepsilon}} = \frac{1}{\Delta t} \cdot \frac{\partial \mathbf{T}}{\partial \mathbf{D}}(\Delta t) \,, \tag{1.108}$$

system (1.107) has to be solved simultaneously with system (1.106). Due to the complicated structure of our constitutive model, the calculation (and implementation) of the expressions appearing on the right-hand side of (1.107) might be a tedious task. We therefore strongly recommend to replace (1.107) by the following approximation which is obtained by numerical differentiation

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{B}_{ij} &= \frac{1}{\vartheta}\Big(\mathbf{h}(\mathbf{T} + \vartheta \mathbf{B}_{ij}, \mathbf{D} + \vartheta \mathbf{V}_{ij}, \mathbf{Q} + \vartheta \mathbf{G}_{ij}) - \mathbf{h}(\mathbf{T}, \mathbf{D}, \mathbf{Q})\Big) \,, \\ \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{G}_{ij} &= \frac{1}{\vartheta}\Big(\mathbf{k}(\mathbf{T} + \vartheta \mathbf{B}_{ij}, \mathbf{D} + \vartheta \mathbf{V}_{ij}, \mathbf{Q} + \vartheta \mathbf{G}_{ij}) - \mathbf{k}(\mathbf{T}, \mathbf{D}, \mathbf{Q})\Big) \end{aligned} \tag{1.109}$$

with $\mathbf{B}_{ij}(0) = \mathbf{0}$ and $\mathbf{G}_{ij}(0) = \mathbf{0}$ for $1 \le i \le j \le 3$, see (1.95). Here, $\mathbf{V}_{ij}$ denotes the standard basis tensor

$$\mathbf{V}_{ij} = (\delta_{ik}\delta_{j\ell})_{k,\ell=1}^{3} \tag{1.110}$$

with $\delta_{ik} = 1$ if $i = k$ and $\delta_{ik} = 0$ else. A Taylor series expansion of the right-hand side of (1.109) shows that

$$\mathbf{B}_{ij} = \frac{\partial \mathbf{T}}{\partial D_{ij}} + \mathscr{O}(\vartheta)\,. \tag{1.111}$$

Thus the six tensors $\mathbf{B}_{ij}$ are good approximations to the Jacobian for ϑ suitably chosen [7]. We propose to solve (1.106) and (1.109) simultaneously with the same numerical method (as described subsequently). This guarantees the consistency of the derivatives.

1.6.2 Adaptive Time Integration

We have to solve the coupled system (1.106) and (1.109) with Richardson extrapolation of the explicit Euler scheme. Collecting all the variables of our problem in a super-vector

$$\begin{aligned}\mathbf{y} = \big[&T_{11}, T_{22}, T_{33}, T_{12}, T_{13}, T_{23},\\ &(B_{11})_{11}, (B_{11})_{22}, (B_{11})_{33}, (B_{11})_{12}, (B_{11})_{13}, (B_{11})_{23},\\ &(B_{22})_{11}, \ldots, (B_{22})_{23}, (B_{33})_{11}, \ldots, (B_{23})_{23},\\ &Q_1, \ldots, Q_m, (G_{11})_1, \ldots, (G_{11})_m, (G_{22})_1, \ldots, (G_{23})_m\big]^{\mathsf{T}}\end{aligned} \tag{1.112}$$

and denoting the right-hand sides of (1.106) and (1.109) by $\mathbf{f}$, we obtain the initial value problem

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\mathbf{y}(t) = \mathbf{f}(\mathbf{y}(t))\,, \qquad \mathbf{y}(0) = \mathbf{y}_0 \quad \text{given}\,. \tag{1.113}$$

The integration of this system is performed as explained in the Sect. 1.3.

1.6.3 Application to Hypoplasticity

The use of explicit and semi-implicit adaptive integration schemes in combination with a numerical computation of the Jacobian was investigated in detail in [5, 6]. There we used hypoplasticity with intergranular strain (see "Appendix: Hypoplastic Models") in a finite element framework. As the evolution equations for the intergranular strain constitute a numerically stiff problem, the semi-implicit method turned out to be superior in element tests like the drained and undrained triaxial tests (see Figs. 1.15 and 1.16). This is in line with the one-dimensional investigation in Sect. 1.3.6.

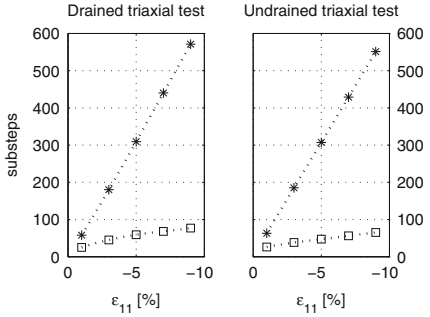

Fig. 1.15 Total number of time steps against vertical strain ε_{11} in drained and undrained triaxial test: *crosses and squares* mark the explicit and the semi-implicit method, respectively

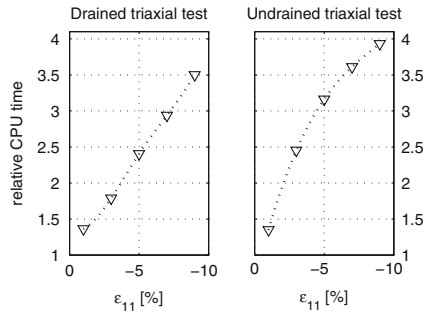

Fig. 1.16 Ratio of the required CPU time as a function of the vertical strain ε_{11}. The ratio is defined as the CPU time required by the explicit integration divided by that of the semi-implicit one

However, the situation changes, when typical problems from geotechnical applications are considered. In such finite element calculations, large deformations typically take place in small regions only, and small deformations have to be expected in the rest of the computational domain. Due to this fact, only few integration steps have to be taken in most of the elements. The savings (in terms of steps) of the semi-implicit method are then too small to counterbalance the higher computational cost.

In summary, the adaptive explicit method turned out to be the best choice for integrating hypoplasticity with intergranular strain in geotechnical applications. Switching to semi-implicit integration in numerically stiff regions is worth thinking about. However, as these regions are typically small and any switch algorithm will take some extra time, the effect on the overall performance is assumed to be small.

1.6.4 Application to Elasto-Plasticity

Here we show the applicability of the numerical time integration strategies to an extended von Mises elasto-plastic model. For the sake of simplicity we will formulate the model in principal stresses, which is sufficient as we will calculate the stress response of an unconfined compression test. We extend the linear elastic perfectly plastic von Mises model by nonlinear elasticity in a Duncan–Chang

formulation and nonlinear isotropic hardening with an evolution equation like the hypoplastic intergranular strain.

The yield function reads

$$f_Y(\mathbf{T}, \alpha) = T^\star - (T_Y + K\alpha) \tag{1.114}$$

with

$$T^\star = \sqrt{3J_2} = \frac{1}{\sqrt{2}}\sqrt{(T_1 - T_2)^2 + (T_2 - T_3)^2 + (T_3 - T_1)^2}\,, \tag{1.115}$$

where J_2 is the second invariant of the deviatoric stresses. The time rate of the stress is given by

$$\dot{\mathbf{T}} = \left(1 - R_f \frac{T^\star}{T_Y}\right)^2 \mathbf{C}^{-1}(\mathbf{D} - \mathbf{D}_p)\,, \tag{1.116}$$

where we have used the following vector notation for stress and stretching

$$\mathbf{T} = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix}, \qquad \mathbf{D} = \begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix}. \tag{1.117}$$

The elastic stiffness matrix in (1.116) is

$$\mathbf{C}^{-1} = \frac{E}{(1+\nu)(1-2\nu)} \begin{bmatrix} 1-\nu & \nu & \nu \\ \nu & 1-\nu & \nu \\ \nu & \nu & 1-\nu \end{bmatrix}, \tag{1.118}$$

and the plastic stretching has the form

$$\mathbf{D}_p = \gamma \frac{\partial f_Y}{\partial \mathbf{T}} \tag{1.119}$$

with

$$\frac{\partial f_Y}{\partial \mathbf{T}} = \frac{1}{2(T_Y + K\alpha)} \begin{bmatrix} 2T_1 - T_2 - T_3 \\ 2T_2 - T_3 - T_1 \\ 2T_3 - T_1 - T_2 \end{bmatrix}. \tag{1.120}$$

For $f_Y < 0$, the multiplier has the value $\gamma = 0$. For $f_Y = 0$ and further loading, it has to be calculated from the condition

$$\dot{f_Y} = \left(\frac{\partial f_Y}{\partial \mathbf{T}}\right)^{\mathsf{T}} \dot{\mathbf{T}} + \frac{\partial f_Y}{\partial \alpha}\dot{\alpha} = 0 \tag{1.121}$$

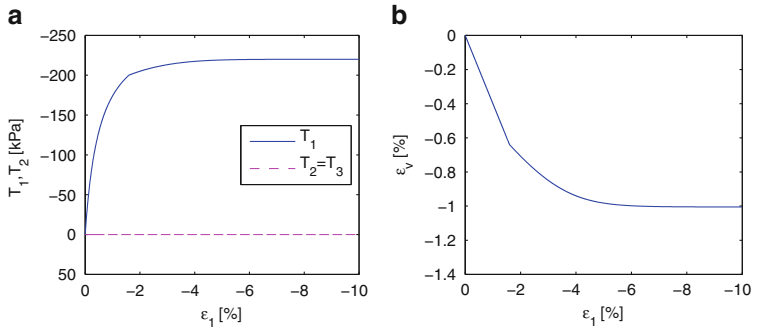

Fig. 1.17 Unconfined uniaxial compression test with extended von Mises plasticity. Material constants: $E = 50{,}000$ kPa, $\nu = 0.3$, $K = 2{,}000$ kPa, $T_Y = 200$ kPa, $R_f = 0.75$, $R = 0.01$. (**a**) Vertical stress T_1, horizontal stresses T_2 and T_3. (**b**) Volumetric strain

for the given stretching $\mathbf{D}$. The time rate of the hardening variable is given by

$$\dot{\alpha} = \left[1 - \frac{\alpha}{R}\right] \|\mathbf{D}_p\| \,. \tag{1.122}$$

Setting

$$\mathbf{y} = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ \alpha \end{bmatrix} \,, \qquad z = \gamma \,, \tag{1.123}$$

and

$$g = f_Y(\mathbf{y}) \tag{1.124}$$

gives the material equations in the form of a standard index 2 problem (1.63). In this form, they can be integrated with the above proposed methods. The results of such an integration with the half-explicit Euler method is shown in Fig. 1.17. The evolution of the internal variables and the yield function is shown in Fig. 1.18. In the region where $f_Y = 0$ an index 2 problem has been solved.

1.7 Conclusion

Finite element programs need the consistent tangent operator to achieve quadratic convergence in the equilibrium iterations. The constitutive part of this operator is the Jacobian, which has to be provided by the user. In this article, we have proposed

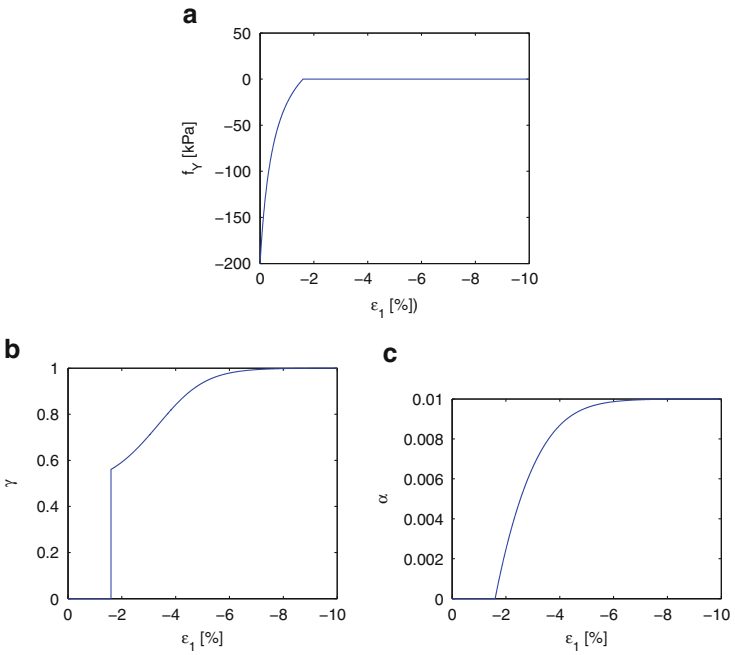

Fig. 1.18 Yield function and internal variables of the calculation in Fig. 1.17. (**a**) Yield function. (**b**) Plastic multiplier. (**c**) Hardening variable

an approach that relieves the user from computing and coding this information for every newly developed constitutive relation. We have shown a way how this Jacobian information can be computed together with the stress increments. The method is based on numerical integration of the variational equation, which themselves are set up automatically by numerical differentiation. The whole approach requires adaptive and efficient time integration. For this purpose, we have presented three integrators, each of them covering a different situation: an explicit method for non-stiff problems, a semi-implicit method for stiff problems, and a half-explicit method for index 2 problems, which appear in plastic problems. Several numerical examples illustrate our approach.

Appendix: Hypoplastic Models

For the sake of completeness, we outline the used hypoplastic model and the parameters used for our calculations. Tensors of second order are denoted with bold letters (e.g., $\mathbf{D}, \mathbf{T}, \boldsymbol{\delta}, \mathbf{N}$) and tensors of fourth order with calligraphic letters (e.g., $\mathscr{L}, \mathscr{M}$). Different kinds of tensorial multiplication are used: $\mathbf{TD} = T_{ij}D_{kl}$, $\mathbf{T} : \mathbf{D} = T_{ij}D_{ij}$, $\mathscr{L} : \mathbf{D} = L_{ijkl}D_{kl}$, $\mathbf{T} \cdot \mathbf{D} = T_{ij}D_{jk}$. The Euclidian norm of a tensor

is $\|\mathbf{D}\| = \sqrt{D_{ij}D_{ij}}$. Unit tensors of second and fourth orders are denoted by $\mathbf{I}$ and $\mathscr{I}$, respectively.

A.1 Basic Model

The basic hypoplastic model was proposed in [20]:

$$\mathring{\mathbf{T}} = \mathscr{L}(\mathbf{T}, e) : \mathbf{D} + \mathbf{N}(\mathbf{T}, e)\|\mathbf{D}\| \tag{1.125}$$

with the linear term

$$\mathscr{L} = f_s \frac{1}{\hat{\mathbf{T}} : \hat{\mathbf{T}}} \left(F^2 \mathscr{I} + a^2 \hat{\mathbf{T}}\hat{\mathbf{T}} \right) \tag{1.126}$$

and the nonlinear term

$$\mathbf{N} = f_s f_d \frac{aF}{\hat{\mathbf{T}} : \hat{\mathbf{T}}} \left(\hat{\mathbf{T}} + \hat{\mathbf{T}}^* \right). \tag{1.127}$$

The employed stress variables are defined as follows

$$\hat{\mathbf{T}} = \frac{\mathbf{T}}{\operatorname{tr}\mathbf{T}}, \qquad \hat{\mathbf{T}}^* = \hat{\mathbf{T}} - \frac{1}{3}\mathbf{I}.$$

The factors for pressure and density dependency (barotropy and pyknotropy) are given by

$$a = \frac{\sqrt{3}(3 - \sin\varphi_c)}{2\sqrt{2}\sin\varphi_c}, \qquad f_d = \left(\frac{e - e_d}{e_c - e_d} \right)^{\alpha},$$

$$f_s = \frac{h_s}{n} \left(\frac{e_i}{e} \right)^{\beta} \frac{1 + e_i}{e_i} \left(\frac{-\operatorname{tr}\mathbf{T}}{h_s} \right)^{1-n} \left[3 + a^2 - a\sqrt{3} \left(\frac{e_{i0} - e_{d0}}{e_{c0} - e_{d0}} \right)^{\alpha} \right]^{-1}.$$

The factor F for adapting the deviatoric yield surface to that of Matsuoka–Nakai is

$$F = \sqrt{\frac{1}{8}\tan^2\psi + \frac{2 - \tan^2\psi}{2 + \sqrt{2}\tan\psi\cos 3\theta}} - \frac{1}{2\sqrt{2}}\tan\psi$$

with

$$\tan\psi = \sqrt{3}\left\|\hat{\mathbf{T}}^*\right\| \quad \text{and} \quad \cos 3\theta = -\sqrt{6}\frac{\operatorname{tr}(\hat{\mathbf{T}}^* \cdot \hat{\mathbf{T}}^* \cdot \hat{\mathbf{T}}^*)}{\left[\hat{\mathbf{T}}^* : \hat{\mathbf{T}}^*\right]^{3/2}}.$$

The void ratios are assumed to fulfill the compression model

$$\frac{e_i}{e_{i0}} = \frac{e_c}{e_{c0}} = \frac{e_d}{e_{d0}} = \exp\left[-\left(\frac{-\operatorname{tr}\mathbf{T}}{h_s}\right)^n\right]. \tag{1.128}$$

This hypoplastic relation has eight parameters: the critical friction angle φ_c, the granular hardness h_s, the void ratios e_{i0}, e_{c0}, and e_{d0}, and the exponents n, α, and β. They can be determined easily from simple index and element tests [13].

Since the mass is assumed to remain constant, the evolution of the void ratio e is described by

$$\dot{e} = (1+e)\operatorname{tr}\mathbf{D}. \tag{1.129}$$

A.2 Extended Hypoplastic Model

The here used extended version of hypoplasticity with intergranular strain was proposed in [16]. The general stress–strain relation is written as

$$\mathring{\mathbf{T}} = \mathscr{M} : \mathbf{D}, \tag{1.130}$$

where $\mathscr{M}$ is a fourth-order tensor that represents the stiffness. It depends on the hypoplastic tensors $\mathscr{L}(\mathbf{T}, e)$ and $\mathbf{N}(\mathbf{T}, e)$ and is defined as follows:

$$\begin{aligned}\mathscr{M} = {} & \left[\rho^{\chi} m_T + (1-\rho^{\chi}) m_R\right]\mathscr{L} + \\ & \begin{cases} \rho^{\chi}(1-m_T)\mathscr{L} : \hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}} + \rho^{\chi}\mathbf{N}\hat{\boldsymbol{\delta}} & \text{for } \hat{\boldsymbol{\delta}} : \mathbf{D} > 0, \\ \rho^{\chi}(m_R - m_T)\mathscr{L} : \hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}} & \text{for } \hat{\boldsymbol{\delta}} : \mathbf{D} \le 0, \end{cases}\end{aligned} \tag{1.131}$$

where $\boldsymbol{\delta}$ is the intergranular strain, and m_R, m_T, χ, and R denote material parameters. The normalized magnitude of $\boldsymbol{\delta}$ is defined as

$$\rho = \frac{\|\boldsymbol{\delta}\|}{R}. \tag{1.132}$$

Further, the direction of the intergranular strain $\boldsymbol{\delta}$ is set

$$\hat{\boldsymbol{\delta}} = \begin{cases} \boldsymbol{\delta}/\|\boldsymbol{\delta}\| & \text{for } \boldsymbol{\delta} \neq \mathbf{0}, \\ \mathbf{0} & \text{for } \boldsymbol{\delta} = \mathbf{0}. \end{cases} \tag{1.133}$$

The evolution equation of the intergranular strain tensor $\boldsymbol{\delta}$ is postulated as

$$\mathring{\boldsymbol{\delta}} = \begin{cases} (\mathscr{I} - \hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}\rho^{\beta_r}) : \mathbf{D} & \text{for } \hat{\boldsymbol{\delta}} : \mathbf{D} > 0, \\ \mathbf{D} & \text{for } \hat{\boldsymbol{\delta}} : \mathbf{D} \le 0, \end{cases} \tag{1.134}$$

Table 1.4 Parameters for the basic model

$\varphi_c(^\circ)$	h_s(kPa)	n	e_{d0}	e_{c0}	e_{i0}	α	β
33	1×10^6	0.25	0.55	0.95	1.05	0.25	1.50

Table 1.5 Parameters for the extended model

R	m_R	m_T	β_r	χ
1×10^{-4}	5.0	2.0	0.5	6.0

where $\mathring{\boldsymbol{\delta}}$ is the objective rate of intergranular strain and the exponent β_r is a material parameter.

For a monotonic continuation of straining with $\mathbf{D} \sim \hat{\boldsymbol{\delta}}$, the stiffness is

$$\mathscr{M} = \mathscr{L} + \mathbf{N}\hat{\boldsymbol{\delta}} \,. \tag{1.135}$$

Note that $\mathbf{D} = \hat{\boldsymbol{\delta}}\|\mathbf{D}\|$ and $\mathbf{N}\hat{\boldsymbol{\delta}} : \mathbf{D} = \mathbf{N}\|\mathbf{D}\|$ in this case. Thus we obtain the basic hypoplastic equation (1.125).

A.3 Material Parameters

The parameters used in all calculations are listed in Tables 1.4 and 1.5.

References

1. Abaqus: User's manual, version 5.8, volume 3. HKS Inc., Hibbit, Karlson & Sorenson, Rhode Island, USA (1998)
2. Duncan, J.M., Chang, C.Y.: Nonlinear analysis of stress and strain in soils. J. Soil Mech. Found. Div. ASCE **96**(SM5), 1629–1653 (1970)
3. Fellin, W.: Hypoplastizität für Einsteiger. Bautechnik **77**(1), 10–14 (2000)
4. Fellin, W.: Hypoplasticity for beginners. University of Innsbruck (2002). ftp.uibk.ac.at/pub/uni-innsbruck/igt/publications/_fellin/hypo_beginner.pdf
5. Fellin, W., Mittendorfer, M., Ostermann, A.: Adaptive integration of constitutive rate equations. Comput. Geotechnics **36**, 698–708 (2009)
6. Fellin, W., Mittendorfer, M., Ostermann, A.: Adaptive integration of hypoplasticity. In: Benz, T., Nordal, S. (eds.) Numerical Methods in Geotechnical Engineering (NUMGE 2010), pp. 15–20. CRC Press/Balkema, London (2010)
7. Fellin, W., Ostermann, A.: Consistent tangent operators for constitutive rate equations. Int. J. Numer. Anal. Methods Geomech. **26**, 1213–1233 (2002)
8. Fellin, W., Ostermann, A.: Using constitutive models of the rate type in implicit finite-element calculations: error-controlled stress update and consistent tangent operator. In: Kolymbas, D. (ed.) Advanced Mathematical and Computational Geomechanics. Lecture Notes in Applied and Computational Mechanics. vol. 13, pp. 211–237. Springer, Heidelberg (2003)
9. Fellin, W., Ostermann, A.: Parameter sensitivity in finite element analysis with constitutive models of the rate type. Int. J. Numer. Anal. Methods Geomech. **30**, 91–112 (2006)

10. Gurtin, M., Spear, K.: On the relationship between the logarithmic strain rate and the stretching tensor. Int. J. Solids Struct. **19**(5), 437–444 (1983)
11. Hairer, E., Nørsett, S., Wanner., G.: Solving Ordinary Differential Equations I. Nonstiff Problems, 2nd edn. Springer, Berlin (1993)
12. Hairer, E., Wanner., G.: Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems. Springer, Berlin (1991)
13. Herle, I.: Hypoplastizität und Granulometrie einfacher Korngerüste. Veröffentlichung des Institutes für Bodenmechanik und Felsmechanik, vol. 142. Universität Fridericiana in Karlsruhe (1997)
14. Kolymbas, D.: A generalized hypoelastic constitutive law. In: Proc. XI Int. Conf. Soil Mechanics and Foundation Engineering, San Francisco, vol. 5, p. 2626. Balkema, Rotterdam (1985)
15. Kolymbas, D.: Introduction to Hypoplasticity. No. 1 in Advances in Geotechnical Engineering and Tunnelling. Balkema, Rotterdam (2000)
16. Niemunis, A., Herle, I.: Hypoplastic model for cohesionless soils with elastic strain range. Mech. Cohesive-Frictional Mater. **2**(4), 279–299 (1997)
17. Schanz, T., Vermeer, P.A., Bonnier, P.G.: The hardening soil model: formulation and verification. In: Brinkgreve, R.B.J. (ed.) Beyond 2000 in Computational Geotechnics, pp. 281–296. A. A. Balkema Publishers, Rotterdam (1999)
18. Simo, J., Huges, T.: Computational Inelasticity. Springer, New York (1998)
19. Truesdell, C., Noll, W.: The Non-Linear Field Theories of Mechanics. Springer, Berlin, Heidelberg (1965)
20. von Wolffersdorff, P.A.: A hypoplastic relation for granular materials with a predefined limit state surface. Mech. Cohesive-Frictional Mater. **1**, 251–271 (1996)

Chapter 2
Barodesy: The Next Generation of Hypoplastic Constitutive Models for Soils

D. Kolymbas

Abstract Barodesy is, like hypoplasticity, a frame for an evolution equation where the stress rate is expressed as tensorial function of stress, stretching and other parameters like void ratio. This equation being non-linear and non-integrable allows to express the path-dependent evolution of stress with deformation. The specific feature of barodesy is that it is based on two very simple theorems on asymptotic behavior of sand. The first theorem states that proportional strain paths starting from the stress-free state lead to proportional stress paths. Barodesy shows that this can be easily modeled with an exponential mapping. The second theorem refers to proportional strain paths starting form a non-vanishing stress state. They lead asymptotically to proportional stress paths that would have been obtained starting at the stress free state. Barodesy models this by adding a simple term in the constitutive relation, and this is now the complete new constitutive relation. The so obtained mathematical relation allows to embed in a simple and elegant way many known principles of soil mechanics, allowing additionally for some asymptotic effects due to cyclic loading. The striking simplicity of the new model not only facilitates its application in numerical applications but also offers a frame for understanding the behavior of soil and granular matter, in general. Moreover, it offers a good starting point for further investigations towards open problems such as rate sensitivity and behavior at small strains.

2.1 Introduction

Barodesy is a completely new frame of constitutive models for soils. In this article the structure of the new theory is outlined; the presentation of special applications and the results of simulations are left for a forthcoming paper. The present article

D. Kolymbas (✉)
Unit of Geotechnical and Tunnel Engineering, University of Innsbruck, Technikerstr. 13, A6020 Innsbruck, Austria
e-mail: Dimitrios.Kolymbas@uibk.ac.at

G. Hofstetter (ed.), *Computational Engineering*, DOI 10.1007/978-3-319-05933-4_2,

© Springer International Publishing Switzerland 2014

refers to sand, barodesy, however, holds also for clay, as shown in the PhD thesis of Gertraud Medicus (in preparation). Clay, being also a particulate material consisting of minute particles has a behaviour very similar to sand. However, there are some differences that mainly arise from the fact that the stiffness of sand in monotonic compression is much higher than that of virgin consolidated clay.

2.2 Empirical Basis of Barodesy

Of basic importance for the following is the notion of a proportional path. Proportional stress and strain paths are characterized by constant ratios of the principal values $\sigma_1 : \sigma_2 : \sigma_3$ and $\varepsilon_1 : \varepsilon_2 : \varepsilon_3$, respectively.

There are two basic experimental findings for sand:

1. Starting from the stress-free state, proportional strain paths lead to proportional stress paths.
2. Starting from a non-vanishing stress state and applying a proportional strain path leads asymptotically to the proportional stress path that would be obtained starting from the stress-free state.

The two rules stated above are inferred by Goldscheider from his test results obtained with rectilinear extensions of sand [5]. These tests have been carried out in a so-called true triaxial apparatus. This apparatus allows to apply rectilinear extensions (i.e. motions without rotation of the principal axes of deformation) independently in all three directions of space.

Besides these rules, the generally observed lack of an elastic regime in soils contradicts a basic ingredient of the theory of plasticity.

2.3 Early Quests for Alternatives to Plasticity Theory

The theory of plasticity, based on the notions of yield surface, flow rule, consistency, decomposition of strain into elastic and plastic parts, etc. was for a long time the only mathematical tool to describe irreversible deformation. Thus, also soil mechanics has been developed along the principles of this theory. The first consistent model for soil, the Cam-Clay model is a particular plasticity theory adapted to clay. However, several researchers have called to depart from plasticity. In 1973 Palmer and Pearce published a paper titled "Plasticity theory without yield surfaces" [18]. Some sentences of this paper deserve being quoted here:

> It was quite natural that the idea of a yield surface should assume such importance in a theory built on experience with metals, since in most metals yield occurs at a fairly well-defined stress level. ...
> In soil mechanics the status of the yield surface concept is quite different, both in theory and experiment...

> . . . strain measurements in clay depend on direct observation of boundary displacements, so that only quite large strain increments are reliably measurable, creep and pore-pressure diffusion confuse results. . .
> . . . yield surface motions during strain-hardening are often too complex for the results to be helpful in constructing usable stress–strain relations.
> Might it be possible to resolve this (dilemma) by constructing a different kind of plasticity model, in which the yield surface concept had been dropped or relegated to a minor role?
> . . . it might be useful to idealise clay as a material in which the yield surface has shrunk to a point, so that all deformations are plastic and *any* changes of stress from the current state will produce plastic strain increments.

Palmer and Pearce present in their paper a concept for a plasticity theory without yield surfaces. This concept is based on two postulates by Ilyushin which are, in a sense, precursors of Goldscheider's theorems:

Isotropy postulate: *If the strain path is rotated in strain space, then the corresponding stress path is rotated by the same amount.* This postulate has nothing to do with isotropy, since it considers rotations in the strain and stress spaces, not in the natural space. It is controversial and certainly not valid in the full stress and strain spaces. It is only approximately valid in the deviatoric subspace: This postulate implies that the deviatoric directions of proportional strain and stress paths coincide. This is, however, not true, according to experimental results by Goldscheider [5].

Delay postulate: *The stress at some instant in a loading history does not depend on the whole previous history, but only on the last part of it.* This is a postulate of fading memory and is similar to Goldscheider's second theorem.

Based on Ilyushin's postulates, Palmer and Pearce present the following concept:

> The deviatoric stress has two components. The magnitude of the first component is a function of the octahedral shear strain, and its direction coincides with the principal strain vector (referring strain to an isotropically-consolidated initial state). The magnitude of the second component is constant, and its direction coincides with the current strain rate . . .
> Reversal of the strain path would reverse the second component but not the first . . .

The very last sentence strongly resembles to a basic concept of hypoplasticity and barodesy, to which presumably the authors would have concluded, had they used rate equations instead of finite ones.

2.4 Barodesy and Hypoplasticity

Constitutive models can't be *derived* from general principles, because they have to describe specific features of particular materials. Thus, besides intuition, trial and error is a basic tool in developing constitutive models. In hypoplasticity, trial and error has been guided by general principles of objectivity and representation

theorems for tensor-valued functions. In barodesy, the amount of trial and error has been further reduced in favour of reasoning on asymptotic behaviour of granulates. Asymptotic states are attractive not only from conceptual reasons but also from the experimental viewpoint: If we consider long monotonic deformations, initial disturbances, related, for example, to sample preparation, fade out and do no more influence the measurements.

Barodesy can be seen as a hypoplastic implementation of the Critical States Concept. Previous attempts to incorporate Critical States into hypoplasticity have been published e.g. by Bauer [1], Gudehus [6], Herle and Kolymbas [7], Masin [16], Niemunis [17], and Wu et al. [21]. As in the original proposal by the author [9, 10], they are composed of two parts, the one being linear and the other non-linear in **D**. Barodesy is also composed of two parts, neither of which is linear in the stretching tensor **D**. The response envelopes of hypoplastic versions are ellipses, whereas in barodesy they have a similar form but are not ellipses. As these features are not essential, the author believes that the mathematical structure of barodesy is appealing and capable of useful extensions. Clearly, all mathematical models (including elastoplastic ones) succeeding to describe the same object, e.g. soil, must include a common mathematical kernel, which is however still hidden.

2.4.1 About the Name "Barodesy"

One should not be fast in introducing new names, as too many neologisms create confusion. However, sometimes new names are needed to denote ideas that are really new. There is an abundance of elastic and plastic concepts equipped with prefixes such as hypo-, para-, hyper- etc. Therefore, the author suggests to avoid using the words elasticity and plasticity (to the extend the latter is associated with notions such as yield surface, elastic regime etc., originally created for metals), since they are not the only framework to describe granular materials such as soil. It should be admitted that soil behaviour can—in principle—also be described in the framework of the theory of elastoplasticity. The author believes, however, that yield surfaces and the other concepts of plasticity theory may prejudice our perception and sometimes obscure soil mechanics, which suffers from the long lasting fragmentation in constitutive modelling [11]. Hypoplasticities have been developed, independently of each other, in California [3], Karlsruhe and Grenoble [2]. The Grenoble and Karlsruhe branches are inherently related. The California and the Karlsruhe/Grenoble perceptions of hypoplasticity have nothing in common but the name, the first one being designed in the frame of elastoplasticiy. It should be stressed that there is no use in seeking rigorous definitions of what *is* hypoplasticity. Taking that a constitutive relation is a mathematical expression being continuously developed, any attempt to provide a strict definition and distinctive characteristics ends up in a sterile exercise of dogmatism. As for the Karlsruhe branch, many different versions have emanated since the publication of the first proposal by

the author in 1977.[1] This proposal was motivated by the quest to describe the mechanical behaviour of soil on the basis of Rational Mechanics without any recourse to the formalism of elastoplasticity.

The new approach presented in this paper pays tribute to a basic idea of Gudehus, who guided the research team in Karlsruhe in the years 1973–2006: Asymptotic states, as represented by proportional strain paths, are attractors and play a paramount role in mechanics of granulates. In this paper is shown that *almost the entire constitutive relation for granulates can be derived from the consideration of proportional paths.* The here presented theory, which yields a variety of more or less realistic predictions, is based on a few reasonable assumptions. Therefore it claims generality and deserves a new name. The name barodesy has been coined motivated by the fact that granular materials gain their stiffness ($\delta\varepsilon\sigma\iota\varsigma$ = bond, hence stiffening, hardening) from externally applied pressure ($\beta\alpha\rho o\varsigma$). Thus, the names "barodesy" and "barodetic" are proposed for granular materials to distinguish them from what traditionally is denoted as "elastic" or "plastic".

2.5 Symbols and Notation

The notation in the "Non-Linear Field Theories of Mechanics" [19] is mainly followed in this article. Compared to the notation of tensors with indices, the symbolic notation facilitates insight into the prevailing relationships.

$\mathbf{T}$:	Cauchy-stress. Its principal components are denoted with $\sigma_1, \sigma_2, \sigma_3$.
$\mathbf{D}$:	Stretching tensor, i.e. the symmetric part of the velocity gradient $\nabla\mathbf{v}$. It can be set approximately equal to the strain rate, $D_{ij} \approx \dot{\varepsilon}_{ij}$.
e:	Void ratio, i.e. the ratio V_p/V_s, where V_p and V_s are the volumes of pores and solids (grains), respectively.
exponent 0:	Denotes normalization of a tensor $\mathbf{A}$, i.e. $\mathbf{A}^0 := \mathbf{A}/\lvert\mathbf{A}\rvert$, with $\lvert\mathbf{A}\rvert := \sqrt{\mathrm{tr}\mathbf{A}^2}$.
σ:	$\lvert\mathbf{T}\rvert$
$\dot{\varepsilon}$:	$\lvert\mathbf{D}\rvert$
ζ:	$\mathrm{tr}\mathbf{D}^0$
$\dot{\mathbf{T}}$:	Time rate of stress. In the general case, $\dot{\mathbf{T}}$ should be replaced by a co-rotational stress rate $\mathring{\mathbf{T}}$. For rectilinear extensions it is $\mathring{\mathbf{T}} \equiv \dot{\mathbf{T}}$.
c_1, c_2, c_3, c_4:	Material constants.

[1] The first versions were not yet named "hypoplastic".

2.6 Proportional Paths A

Let us first consider proportional strain paths starting from the stress-free state. Such paths can be volume-decreasing (we will call them "consolidations"), characterized by $\mathrm{tr}\mathbf{D} < 0$, or volume preserving ("isochoric" or "undrained"), characterized by $\mathrm{tr}\mathbf{D} = 0$, or volumes increasing, characterized by $\mathrm{tr}\mathbf{D} > 0$. Clearly, the latter are not feasible with cohesionless sand. Let us denote with $\mathbf{R}$ a tensor that has the direction of a proportional stress path. The question arises, how $\mathbf{R}$ depends on the direction of the corresponding proportional strain path. The latter is characterized by the direction of stretching $\mathbf{D}$, i.e. by the normalized stretching $\mathbf{D}^0$. How can we determine the relation $\mathbf{R}(\mathbf{D}^0)$? This question can be easily answered if we observe that all consolidations are mapped into a specific part of the principal stress space formed by the stress components σ_1, σ_2 and σ_3. This part is the octant, where all principal stresses are compressive, i.e. negative. Hence, the product $\sigma_1\sigma_2\sigma_3$ must also be negative. Now, for a proportional stress path we have $\sigma_i = \mu R(D_i)$, $\mu > 0, i = 1, 2, 3$.[2] Thus, the following condition must hold:

$$R_1(D_1)R_2(D_2)R_3(D_3) < 0 \qquad \text{for} \quad \mathrm{tr}\mathbf{D} = D_1 + D_2 + D_3 < 0\,. \tag{2.1}$$

This implies that $R_1(D_1)R_2(D_2)R_3(D_3)$ must be a function of $D_1 + D_2 + D_3$, a requirement which is fulfilled by the exponential mapping

$$\mathbf{R}(\mathbf{D}) = \exp(c_1\mathbf{D}^0)\,. \tag{2.2}$$

Equation (2.2) maps all volume-reducing ($\mathrm{tr}\mathbf{D} < 0$) proportional strain paths into a cone in the stress space with appex at $\mathbf{T} = \mathbf{0}$, which can be called the $\mathbf{R}$-cone. Its boundary is the critical state surface and corresponds to paths with $\mathrm{tr}\mathbf{D} = 0$. Consider the intersection of the $\mathbf{R}$-cone with a plane $\mathrm{tr}\mathbf{T} = \text{const}$, as shown in Fig. 2.1. This curve expresses the critical limit state in a so-called deviatoric plane in the stress space. The mathematical representation of this curve can be easily derived from Eq. (2.2): For isochoric deformations ($\mathrm{tr}\mathbf{D}^0 = 0$) we can eliminate $\mathbf{D}^0$ from (2.2) and obtain:

$$\mathbf{D}^0 = \frac{1}{c_1}\ln(-\mathbf{R})\,. \tag{2.3}$$

The requirement $\mathrm{tr}\mathbf{D}^0 = 0$ results in $\ln(-R_1R_2R_3) = 0$ or $R_1R_2R_3 = -1$. From the additional requirement $|\mathbf{D}^0| = 1$ we obtain:

$$(\ln R_1)^2 + (\ln R_2)^2 + (\ln R_3)^2 = c_1^2. \tag{2.4}$$

[2]Herein, D_i are the principal values of $\mathbf{D}$, and $R_j(D_i)$ are the principal values of $\mathbf{R}(\mathbf{D})$.

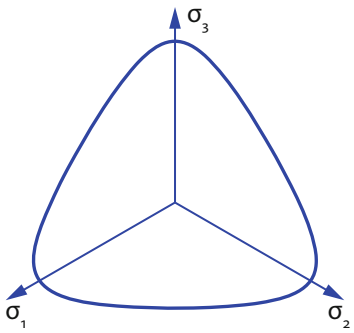

Fig. 2.1 Cross section of the **R**-cone with a deviatoric plane. Numerically obtained with Eq. (2.2) in the following way: the shown curve collects stress states that correspond to isochoric stretchings $\mathbf{D}^0$, $\mathrm{tr}\mathbf{D}^0 = 0$. For any such stretching, Eq. (2.2) yields a stress ray $\mathbf{T} = \lambda\mathbf{R}$, $\lambda > 0$. Its intersection with a π-plane ($\mathrm{tr}\mathbf{T} = \text{const}$) is a point of the shown curve

For the here considered proportional paths holds: $\mathbf{T} = \mu\mathbf{R}$, $0 < \mu < \infty$, hence we can replace in this equation $\mathbf{R}$ by $\mathbf{T}/\mu$ and obtain finally the equation of critical states in the stress space:

$$\left(\ln \frac{T_1}{\sqrt[3]{T_1 T_2 T_3}}\right)^2 + \left(\ln \frac{T_2}{\sqrt[3]{T_1 T_2 T_3}}\right)^2 + \left(\ln \frac{T_3}{\sqrt[3]{T_1 T_2 T_3}}\right)^2 = c_1^2. \tag{2.5}$$

Equation (2.5) is homogeneous of the zero-th degree in $\mathbf{T}$ and describes thus a conical surface in the stress space with apex at $\mathbf{T} = \mathbf{0}$. Its intersection with a plane $\mathrm{tr}\mathbf{T} = \text{const}$ is shown in Fig. 2.1. Note that its shape *practically coincides* [4] with the curve obtained by the expression of Matsuoka and Nakai:

$$\frac{(T_1 + T_2 + T_3)(T_1 T_2 + T_1 T_3 + T_2 T_3)}{T_1 T_2 T_3} = \text{const}. \tag{2.6}$$

Equation (2.2) also relates the critical friction angle with K_0, the so-called coefficient of earth pressure at rest. We consider a critical state and use the abbreviation $K_c := (1 - \sin\varphi_c)/(1 + \sin\varphi_c)$. The equation $R_2/R_1 = K_c$ yields:

$$c_1 = \sqrt{\frac{2}{3}}\, \ln K_c. \tag{2.7}$$

Now we consider an oedometric proportional stress path. The corresponding stretching is

$$\mathbf{D} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{2.8}$$

Herewith we obtain:

$$K_0 = \frac{T_2}{T_1} = \frac{R_2}{R_1} = \frac{\exp(0)}{\exp(-c_1)} = \exp(c_1) = \exp(\ln(K_c^{\sqrt{2/3}})) = K_c^{\sqrt{2/3}}. \quad (2.9)$$

2.7 Proportional Paths B

Now we start from a stress state $\mathbf{T} \neq \mathbf{0}$ and apply the stretching $\mathbf{D}$. In order to asymptotically approach the corresponding proportional stress path $\mathbf{T} = \mu\mathbf{R}(\mathbf{D})$, the stress rate $\dot{\mathbf{T}}$ must point to the point $\mu_1\mathbf{R}(\mathbf{D})$, i.e.

$$\mathbf{T} + \lambda\dot{\mathbf{T}} = \mu_1\mathbf{R}(\mathbf{D}), \quad (2.10)$$

where the positive constants μ, μ_1 and λ need not be further specified here. If we eliminate $\dot{\mathbf{T}}$ we obtain an evolution equation for the stress:

$$\dot{\mathbf{T}} = \nu_1\mathbf{R}(\mathbf{D}) + \nu_2\mathbf{T} \quad (2.11)$$

with appropriately defined scalar quantities ν_1 and ν_2. Equation (2.11) is the final general form of the barodetic constitutive relation. To comply with barotropy, pyknotropy and rate independence of sand, ν_1 and ν_2 are further specified, such that the barodetic constitutive equation for sand obtains the following specific form:

$$\dot{\mathbf{T}} = h(\sigma) \cdot (f\mathbf{R}^0 + g\mathbf{T}^0) \cdot \dot{\varepsilon}, \quad (2.12)$$

where f and g are functions of the stress and the void ratio. In the sequel it will be shown how all known concepts of soil mechanics can be cast in the frame given by Eq. (2.12).

2.8 Limit States

A limit state is obtained when the stiffness vanishes:

$$\dot{\mathbf{T}} = \mathbf{0}. \quad (2.13)$$

Considering stress–strain curves, the limit states are manifested either as peak or residual limit states, where the curve obtains a horizontal slope. In barodesy (see Eq. (2.12)), yield is modeled by the equation

$$f\mathbf{R}^0 + g\mathbf{T}^0 = \mathbf{0}. \quad (2.14)$$

This tensorial equation implies two equations:

1. The flow rule

$$\mathbf{R}^0 = \mathbf{T}^0. \tag{2.15}$$

   Note that $\mathbf{R}$ depends on $\mathbf{D}$. Thus, the flow rule gives (via an implicit equation) the direction of strain that pertains to a limit state $\mathbf{T}$.
2. The scalar equation

$$f + g = 0, \tag{2.16}$$

   which takes into account the actual void ratio e and the stress magnitude σ. This scalar equation somehow corresponds to the yield surface of plasticity theory.

2.9 Incremental Non-Linearity

Incremental non-linearity (or "non-linearity in the small") means different stiffnesses at loading and unloading and, in general, irreversible or hysteretic mechanical behaviour. Both, elastoplastic and hypoplastic relations comprise incremental non-linearity. The elastoplastic approach consists in introducing two different stiffnesses, one for loading and one for unloading. A criterion has to be added to distinguish when we have loading and when unloading. In the frame of hypoplasticity a unique expression for the stress rate (or stiffness) is used, and the distinction between loading and unloading is accomplished by the non-linearity of this equation. In barodesy, the difference of stiffness at loading and unloading is modelled by the fact that the second term $g\mathbf{T}^0$ in Eq. (2.12) is not changed if $\mathbf{D}$ is switched to $-\mathbf{D}$, whereas the first term (i.e. $f\mathbf{R}^0$) undergoes a change.

2.10 Consolidations and Critical States

Noting that $\mathbf{T}^0 = \mathbf{R}^0$ holds true for proportional paths, we obtain from Eq. (2.12)

$$\dot{\mathbf{T}} = h(\sigma)\, \mathbf{T}^0\, (f + g)\, \dot{\varepsilon}. \tag{2.17}$$

For proportional paths holds also $\dot{\mathbf{T}} = \dot{\sigma}\mathbf{T}^0$, hence Eq. (2.12) reduces to

$$\dot{\sigma} = h(\sigma)\, (f + g)\, \dot{\varepsilon}. \tag{2.18}$$

The quantities f, g and, hence, $f + g$ are functions (still to be defined) of the void ratio e, the stress magnitude σ and of ζ, introduced in Sect. 2.5, which is a measure of dilatancy.

Vanishing stiffness for critical states implies $f + g = 0$ for $\zeta = 0$. Hence, we can set

$$f + g = c_2 \zeta. \tag{2.19}$$

We require Eq. (2.19) to be valid not only for critical states but also for all consolidations, i.e. for proportional paths with $\epsilon < 0$. Introducing Eq. (2.19) into (2.18) we obtain:

$$\dot{\sigma} = h(\sigma)\, c_2\, \zeta\, \dot{\varepsilon}, \tag{2.20}$$

Using

$$\zeta\, \dot{\varepsilon} = \frac{\mathrm{tr}\mathbf{D}}{\dot{\varepsilon}}\, \dot{\varepsilon} = \mathrm{tr}\mathbf{D} = \frac{\dot{e}}{1+e}, \tag{2.21}$$

we obtain for consolidations:

$$\dot{\sigma} = h(\sigma)\, c_2 \frac{\dot{e}}{1+e}. \tag{2.22}$$

It follows that the slope of the e vs. σ curves is the same for oedometric, hydrostatic and, in general, for all consolidations. Adapting $h(\sigma)$, and thus Eq. (2.22), to a compression curve from a laboratory test allows to determine the compression curve $e = \kappa(\sigma)$. If we choose

$$h = \sigma^{c_3}, \tag{2.23}$$

we obtain

$$e = \kappa(\sigma) = (1 + e_0) \exp \frac{\sigma^{1-c_3}}{(1-c_3)c_2} - 1 \tag{2.24}$$

with $c_2 < 0$.

The incremental stiffness of compression tests, in particular of oedometric compression tests, denoted by $E_s := d\sigma_1 / d\varepsilon_1$, is known to be stress-dependent according to a relation attributed to Ohde and/or Janbu [8]:

$$E_s = E_{s0} \left(\frac{\sigma}{\sigma_0} \right)^w. \tag{2.25}$$

With $d\sigma_1 = -d\sigma/\sqrt{1+2K_0^2}$ and $d\varepsilon_1 = de/(1+e)$ we obtain from Eq. (2.22):

$$E_s = \frac{-c_2}{\sqrt{1+2K_0^2}}\,\sigma^{c_3} = \frac{-c_2\sigma_0^{c_3}}{\sqrt{1+2K_0^2}}\left(\frac{\sigma}{\sigma_0}\right)^{c_3} \tag{2.26}$$

in accordance with Eq. (2.25). A typical value for c_3 is ca. 0.5.

As said, the equation $f+g=0$ or $f+g=c_2\epsilon$ expresses for $\epsilon=0$ the critical state line (CSL) $e-e_c(\sigma)=0$. Thus, we can set

$$f+g = c_2\zeta + c_4(e_c(\sigma)-e). \tag{2.27}$$

For peak limit states we also have $f+g=0$. This can be fulfilled by Eq. (2.27) if $e<e_c$ for $c_2\epsilon<0$, i.e. for dilatant deformation with tr$\mathbf{D}>0$.

For Eq. (2.27) to be also valid for consolidations (i.e. to obtain $f+g=c_2\epsilon$, cf. Eq. (2.22)) we have to require that $e-e_c(\sigma)$ vanishes for consolidations, see Eq. (2.24). For this to hold, we have to require:

$$e_c(\sigma) = \kappa(\sigma). \tag{2.28}$$

In other words, the dependence of the critical void ratio e_c on stress σ is given by the same function that holds for consolidations (i.e. proportional compressions). Of course, the initial void ratio $\kappa(0)$ must be appropriately chosen in each case. Note that the general opinion in soil mechanics is not unique in that question. Many authors accept that the dependence $e_c(\sigma)$ is the same as in compression tests, other authors contradict this view. Barodesy leads to the acceptance of this view.

However, it has to be admitted, that the CSL is hard to determine by experiments. Wood [20] writes:

> The paths of tests on loose and dense samples head towards a somewhat diffuse, but clearly pressure dependent, zone of critical void ratios.

The final step in determining the barodetic constitutive equation consists in partitioning equation (2.27) into f and g by setting, e.g.

$$f = c_2\zeta - c_4 e, \tag{2.29}$$

$$g = c_4 e_c(\sigma). \tag{2.30}$$

2.11 Cyclic Loading, Limit Cycles and Shake-Down

Proportional paths are not the only attractors in the constitutive relation presented so far. Being ordinary differential equations, constitutive relations "of the rate type" [19] may exhibit also limit cycles or cyclic orbits as further attractors. In

fact, Eq. (2.12) exhibits periodic orbits (limit cycles) at cyclic loading. In terms of mechanics, this effect is related to "shake-down" and means that stress cycles lead asymptotically to cyclic changes of void ratio. In case of, for example, oedometric deformation (but not for conventional triaxial tests), this implies also cyclic strain, i.e. strains due to cyclic stress are bounded, i.e. they do not increase to infinity. Generally, shake-down is one of the possible responses of sand to cyclic loading, the other one being "incremental collapse" (i.e. unlimited growth of strain with increasing number of cycles). It is yet unclear when exactly shake-down and when incremental collapse are to be expected. However, Eq. (2.12) exhibits shake-down (and periodic orbits) e.g. at cyclic oedometric loading: If the axial stress component σ_1 is periodically changed between a lower and an upper limit, then the corresponding radial stress component σ_2, which is bounded, will also become cyclic, i.e. a limit cycle will eventually be obtained in the stress space.[3]

Cyclic stress, asymptotically obtained with strain cycles of infinitesimally small amplitude, is related with the void ratio $\check{e}$, which can be called the *cyclic void ratio*. Little is known from experiments on the dependence of $\check{e}$ on actual stress $\mathbf{T}$ and on the direction $\mathbf{D}^0$ of strain cycles.

Considering strain cycles with infinitesimal amplitude with the constitutive relation (2.12) and denoting with "+" and "−" loading and unloading, respectively, it is observed that at a limit cycle must hold: $\dot{\mathbf{T}}^+ = -\dot{\mathbf{T}}^-$. Hence, the condition for cyclic response reads

$$(f^+\mathbf{R}^{0+} + f^-\mathbf{R}^{0-}) + (g^+ + g^-)\mathbf{T}^0 = \mathbf{0}. \tag{2.31}$$

This equation constitutes a relation between the direction of the strain amplitude, $\mathbf{D}$, the cyclic void ratio $\check{e}$ and the stress $\sigma\mathbf{T}^0$, around which the stress oscillation occurs. Eliminating $\mathbf{T}^0$ from Eq. (2.31) yields:

$$\mathbf{T}^0 = \frac{-1}{g^+ + g^-}(f^+\mathbf{R}^{0+} + f^-\mathbf{R}^{0-}). \tag{2.32}$$

Using this equation and the additional condition $|\mathbf{T}^0| = 1$ makes it possible to determine for a given $\mathbf{D}^0$ the stress direction $\mathbf{T}^0$ of the corresponding cyclic state and also the pertaining cyclic void ratio $\check{e}(\sigma)$. A discussion of Eq. (2.32) is left for a future paper.

It should be added that the here presented model still exhibits ratcheting at cycles of small amplitude, e.g. in conventional triaxial tests.

[3]Integrity of grains (or permanence of the grain size distribution) has not been assumed for the derivation of the constitutive relation so far. In fact, a constitutive relation that does not contain any measure for the strength of grains presupposes that grain crushing does not occur. In reality, however, grain crushing is inevitable, especially at higher stresses. The corresponding changes of the grain size distribution curve are hard to measure.

2.12 Significance of Barodesy

Compared with the "classical" elastoplastic approaches, Eq. (2.12) constitutes a substantial change of paradigm and introduces not only new concepts but also a remarkable simplicity in a field of paramount complexity[4] dominated by a "morass of equations". Equation (2.12) is a convincing implementation of Noll's[5] program to formulate a constitutive equation as a rate equation of the type $\dot{\mathbf{T}} = \mathbf{h}(\mathbf{T}, \mathbf{D})$, and the importance of this achievement is enhanced by the fact that Eq. (2.12) is *derived* from general properties of sand. The implications of Eq. (2.12) are amazing. Despite its simplicity it captures almost every aspect of the behaviour of granular materials: stress dependent stiffness, hysteretic behaviour, dilatancy, contractancy, hardening up to the peak and subsequent softening to critical states, stress–strain curves and stress-paths for all types of tests, including drained and undrained triaxial tests. In a series of papers [12–15] are shown simulation results including drained and undrained triaxial tests with loose and dense sand, cyclic oedometric tests and cyclic simple shear tests with constant normal stress and constant volume. The range of applicability is huge, as it covers all particulate materials such as soils, granulates and powders. Such materials are addressed not only by geotechnical engineering but also by many other technological branches, such as offshore, mining, petroleum engineering, metallurgy, chemical and food industry.

2.13 Open Questions

Despite its simplicity and elegance, the present version of barodesy cannot cover all aspects of sand behaviour. The memory is still contained only in the actual stress $\mathbf{T}$ and the actual porosity e, and this is not sufficient to cover all aspects of re-loading, in particular the so-called aspects of "small strain stiffness". Though, it is interesting to note how many aspects of memory can be covered with $\mathbf{T}$ and e.

The barodetic equations are homogeneous of the first degree in the stretching $\mathbf{D}$ and, hence, rate-independent. To change this, the degree of homogeneity has to be modified.

[4] "Many properties of sand are equally puzzling to science as the big bang is", Neue Zuercher Zeitung, 13.2.2008.

[5] A prominent representative of a school of thought called Rational Mechanics. The main reference is the classical book "The Non-Linear Field Theories of Mechanics" [19].

References

1. Bauer E.: The critical state concept in hypoplasticity. In: Yuan, J.-X. (ed.) Computer Methods and Advances in Geomechanics, pp. 691–696. Balkema, Rotterdam (1997)
2. Chambon, R.: Une classe de lois de comportement incrementalement non-lineaires pour les sols non-visqueux, resolution de quelques problemes de coherence. C. R. Acad. Sci. Paris Ser II **308**(7), 1571–1576 (1989)
3. Dafalias, Y.F.: Bounding surface plasticity. I: mathematical foundation and hypoplasticity. J. Eng. Mech. ASCE **112**, 966–987 (1986)
4. Fellin, W., Ostermann A.: The critical state behaviour of barodesy compared with the Matsuoka-Nakai failure criterion. Int. J. Numer. Anal. Methods Geomech. (2011). doi: 10.1002/nag.1111
5. Goldscheider, M.: Grenzbedingung und Fließregel von Sand. Mech. Res. Commun. **3**, 463–468 (1976)
6. Gudehus, G.: A visco-hypoplastic constitutive relation for soft soils. Soils Found. **44**(4), 11–26 (2004)
7. Herle, I., Kolymbas, D.: Hypoplasticity for soils with low friction angles. Comput. Geotechnics **31**, 365–373 (2004)
8. Janbu, N.: Soil compressibility as determined by oedometer and triaxial tests. In: Proceedings of the European Conference on Soil Mechanics and Foundation Engineering (1963)
9. Kolymbas, D.: A rate-dependent constitutive equation for soils. Mech. Res. Commun. **4**, 367–372 (1977)
10. Kolymbas, D.: A generalised hypoelastic constitutive law. In: Proceedings of XI International Conference on Soil Mechanics and Foundation Engineering, vol. 5, p. 2626. Balkelma, San Francisco (1985)
11. Kolymbas, D.: The misery of constitutive modelling. In: Kolymbas, D. (ed.) Constitutive Modelling of Granular Materials, pp. 11–24. Springer, Berlin (2000)
12. Kolymbas, D.: Barodesy: a new hypoplastic approach. Int. J. Numer. Anal. Methods Geomechanics (2011). doi: 10.1002/nag.1051
13. Kolymbas, D.: Sand as an archetypical natural solid. In: Kolymbas, D., Viggiani, G. (eds.) Mechanics of Natural Solids, pp. 1–26. Springer, Berlin (2011)
14. Kolymbas, D.: Barodesy: a new constitutive frame for soils. Geotechnique Lett. **2**, 17–23 (2012)
15. Kolymbas, D.: Barodesy as a novel hypoplastic constitutive theory based on the asymptotic behaviour of sand. Geotechnik **35**(3), 187–197 (2012)
16. Masin, D.: A hypoplastic constitutive model for clays. Int. J. Numer. Anal. Methods Geomech. **29**, 311–336 (2005)
17. Niemunis, A.: Extended hypoplastic models for soils, Heft 34. Schriftreihe des Inst. f. Grundbau u. Bodenmechanik der Ruhr-Universitaet Bochum, Bochum (2003)
18. Palmer, A.C., Pearce, J.A.: Plasticity theory without yield surfaces. In: Palmer, A.C. (ed.) Symposium on Plasticity and Soil Mechanics, pp. 188–200. Cambridge University Press, Cambridge (1973)
19. Truesdell, C.A., Noll, W.: The Non-Linear Field Theories of Mechanics. In: Encyclopedia of Physics, vol. IIIc. Springer, Berlin (1965)
20. Wood, D.M.: Soil Behaviour and Critical State Soil Mechanics. Cambridge University Press, Cambridge (1990)
21. Wu, W., Bauer, E., Kolymbas, D.: Hypoplastic constitutive model with critical state for granular materials. Mech. Mater. **23**, 45–69 (1996)

Chapter 3
Seismic Performance of Tuned Mass Dampers with Uncertain Parameters

C. Adam, M. Oberguggenberger, and B. Schmelzer

Abstract This chapter addresses the seismic performance of Tuned Mass Dampers (TMDs). In the design of a TMD, two types of uncertainty are relevant: the stochastic excitation modeling the earthquake, and the inherent uncertainty of internal parameters of the damping device and the subsoil. Modeling the excitation by a continuous-time stochastic process the structure-damper system can be described by a linear system of stochastic differential equations. The response is a stochastic process depending on the uncertain parameters of the damping device and the subsoil. These uncertainties are modeled by random sets, i.e., interval-valued random variables. A framework is presented here that admits the combination of these two types of uncertainty leading to a set-valued stochastic process, which is interpreted as containing the true system response. The approach is applied to show how the efficiency of TMDs can be realistically assessed in the presence of uncertainty. The main focus of this paper is on non-stationary models for the excitation based on colored noise multiplied by a prescribed intensity function.

C. Adam (✉)
Unit of Applied Mechanics, University of Innsbruck, Technikerstr. 13, 6020 Innsbruck, Austria
e-mail: christoph.adam@uibk.ac.at

M. Oberguggenberger
Unit of Engineering Mathematics, University of Innsbruck, Technikerstr. 13, 6020 Innsbruck, Austria
e-mail: michael.oberguggenberger@uibk.ac.at

B. Schmelzer
Unit of Engineering Mathematics, University of Innsbruck, Innsbruck, Austria
e-mail: bernhard.schmelzer@uibk.ac.at

G. Hofstetter (ed.), *Computational Engineering*, DOI 10.1007/978-3-319-05933-4_3,

© Springer International Publishing Switzerland 2014

3.1 Introduction

The protection of vibration-prone structures against excessive dynamic response can be accomplished with various passive, active, and semi-active measures, depending on the complexity of the problem, available resources, expected lifespan, available technological standard, environmental conditions, etc. [31, 32]. Since these structures exhibit in general low inherent damping, the installation of a Tuned Mass Damper (TMD) [8] is one effective classical measure to add damping. A TMD is a simple vibratory mechanical device with a single dynamic degree-of-freedom (SDOF) of either mass-spring-dashpot or a pendulum-dashpot type. When appropriately designed, the kinetic energy is transferred from the vibrating structure to the TMD, where it is subsequently dissipated through its viscous element. From the perspective of its weight added to the structure, visual appearance, and space considerations the maximum mass ratio, i.e. the ratio of TMD mass and effective structural mass, is limited in general to 8 %. The efficiency of this device depends on appropriate tuning of its system parameters and on the frequency content of the excitation.

In current engineering practice, TMDs are frequently used to reduce narrow-band structural vibrations induced by wind, traffic, machines, etc. For the tuning of TMD system parameters and prediction of the actual response reduction analytic relations are readily available [8, 14, 40]. However, the efficiency of a TMD to mitigate broad-band earthquake-induced structural vibrations is a topic that is still controversially discussed [5, 13, 17, 26]. Nonetheless, the seismic behavior of a TMD aimed at particularly protecting the building against narrow-band vibrations (excited, e.g., by wind) needs to be assessed reliably if the building is located in an earthquake environment [37]. For example, the stroke, i.e., the peak displacement of the TMD, must not exceed a certain design limit when subjected to severe earthquake excitation with low probability of occurrence [18,39]. Consequently, one objective of this paper is to provide a fundamental study of the seismic performance of a TMD attached to a vibration-prone load-bearing structure that can be modeled as SDOF system.

Seismic assessment of a TMD should consider the quantification of aleatory and epistemic uncertainties. The record-to-record variability of the earthquake excitation is the source of aleatory uncertainty. One option to capture the aleatory uncertainty of the structure-TMD interaction system is to evaluate the responses to several base accelerations with overall characteristic properties recorded during real earthquakes. Based on this approach recent studies [1, 37, 38] have revealed that a TMD reduces the root mean square response effectively, depending on the mass ratio, inherent structural damping, and fundamental structural frequency. However, it was also shown that a TMD might be less avid to decrease the seismic peak response. Additionally, in [37] analytic approximations of the response quantities for design purposes have been derived. Analytic stochastic excitation modeling of earthquake records (see e.g. [25]) is an alternative approach to capture aleatory uncertainty that is more feasible if the earthquake hazard is not well defined.

Epistemic uncertainties result from the lack of knowledge of internal parameters as well as from approximations to reality of the underlying mechanical model. Here the effect of detuned TMD parameters comes into play, which can be traced back to their internal uncertainty. Structural and TMD parameters can only be determined within certain bounds, and they may be subject to change in the course of time. For example, the stiffness of the soil, and as consequence, natural frequencies of the vibration-prone structure depend on environmental conditions such as temperature and moisture.

In this paper, a framework is presented that admits the combination of stochastic processes (i.e., the earthquake excitation) and interval type parameter uncertainty modeled by random sets (i.e., epistemic uncertainty). In particular, modeling the excitation by a continuous-time stochastic process the structure-TMD system is described by a linear system of stochastic differential equations. The system response is a stochastic process depending on the uncertain parameters of the damping device and the subsoil. These uncertainties are modeled by random sets, i.e., finitely many intervals each coming with a probability weight. The approach is applied to show how the efficiency of passive damping mechanisms can be realistically assessed in the presence of uncertainty. In contrast to a previous study [29], where the ground motion was modeled by white noise, in the following colored noise based on the Kanai–Tajimi power spectral density function [15, 34] describes the base acceleration, which is more realistic for earthquake excitation. Preliminary results of the present study have been presented in [28].

3.2 Mechanical Model

This study discusses the vibration mitigation of earthquake excited linear elastic load-bearing structures, whose dynamic response is primarily governed by the fundamental mode. In general, the mechanical model of an SDOF oscillator represents this category of structures with sufficient accuracy. Subsequently, m_s, k_s, and c_s represent lumped mass, stiffness, and viscous damping parameter of this main system. A second SDOF oscillator with lumped mass m_d ($\ll m_\mathrm{s}$), stiffness parameter k_d, and viscous damping parameter c_d serves as TMD. Combined in series this yields the non-classically damped system shown in Fig. 3.1 with two dynamic degrees-of-freedom, expressed by the displacement x_s of the structure and of the TMD x_d, both measured relative to the base displacement x_g. When subjected to base acceleration $\ddot{x}_\mathrm{g}$, the coupled equations of motion of this system read as follows:

$$\mathbf{M}\begin{bmatrix}\ddot{x}_\mathrm{s}\\ \ddot{x}_\mathrm{d}\end{bmatrix}+\mathbf{C}\begin{bmatrix}\dot{x}_\mathrm{s}\\ \dot{x}_\mathrm{d}\end{bmatrix}+\mathbf{K}\begin{bmatrix}x_\mathrm{s}\\ x_\mathrm{d}\end{bmatrix}=\begin{bmatrix}-1\\ -\mu\end{bmatrix}\ddot{x}_\mathrm{g} \tag{3.1}$$

where

$$\mathbf{M}=\begin{bmatrix}1 & 0\\ 0 & \mu\end{bmatrix},\quad \mathbf{C}=\begin{bmatrix}2\zeta_\mathrm{s}\omega_\mathrm{s}+2\zeta_\mathrm{d}\omega_\mathrm{d}\mu & -2\zeta_\mathrm{d}\omega_\mathrm{d}\mu\\ -2\zeta_\mathrm{d}\omega_\mathrm{d}\mu & 2\zeta_\mathrm{d}\omega_\mathrm{d}\mu\end{bmatrix},\quad \mathbf{K}=\begin{bmatrix}\omega_\mathrm{s}^2+\omega_\mathrm{d}^2\mu & -\omega_\mathrm{d}^2\mu\\ -\omega_\mathrm{d}^2\mu & \omega_\mathrm{d}^2\mu\end{bmatrix}$$

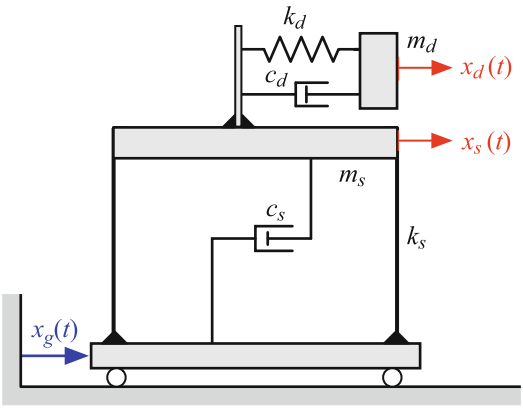

Fig. 3.1 Mechanical model of an SDOF vibration-prone structure equipped with a TMD

The variable μ denotes the mass ratio,

$$\mu = \frac{m_\mathrm{d}}{m_\mathrm{s}}$$

ω_s and ω_d are the natural circular frequencies, and ζ_s and ζ_d denote the non-dimensional damping coefficients of the stand-alone main system and the detuned TMD, respectively,

$$\omega_\mathrm{s} = \sqrt{\frac{k_\mathrm{s}}{m_\mathrm{s}}}, \quad \omega_\mathrm{d} = \sqrt{\frac{k_\mathrm{d}}{m_\mathrm{d}}}, \quad \zeta_\mathrm{s} = \frac{c_\mathrm{s}}{2\omega_\mathrm{s} m_\mathrm{s}}, \quad \zeta_\mathrm{d} = \frac{c_\mathrm{d}}{2\omega_\mathrm{d} m_\mathrm{d}}$$

For an effective reduction of the structural response x_s the parameters of the TMD, i.e., the damping coefficient ζ_d and the frequency ratio δ,

$$\delta = \frac{\omega_\mathrm{d}}{\omega_\mathrm{s}} \tag{3.2}$$

must be tuned "optimally." In general, optimal TMD parameters depend on the type of excitation (harmonic, white noise, etc.) and on the considered response quantity to be optimized (relative or absolute structural displacement or acceleration), see e.g., [3,8,14]. For stationary Gaussian white noise base excitation of an SDOF main system without inherent structural damping (i.e., $\zeta_\mathrm{s} = 0$) the following analytic expressions of optimal TMD parameters have been derived,

$$\delta_\mathrm{opt} = \frac{\sqrt{1-\mu/2}}{1+\mu}, \quad \zeta_\mathrm{d,opt} = \sqrt{\frac{\mu(1-\mu/4)}{4(1+\mu)(1-\mu/2)}} \tag{3.3}$$

assuming that the variance of the stationary relative displacement x_s is minimized, e.g., [3] and [32, p. 234]. Tuning of a TMD according to these expressions minimizes the variance of the relative displacement x_s of the SDOF main system.

3.3 Modeling of the Earthquake Excitation

As outlined in the introduction, base excitation is modeled by a stochastic process. An $\mathbb{R}^d$-valued *stochastic process* $\mathbf{x}$ on a time interval $[0, \bar{t}]$ assigns to each point of time t a random variable $\mathbf{x}(t)$, defined on a probability space Ω with its σ-algebra Σ of measurable sets and the probability measure P. The process is specified if the finite dimensional joint distributions of all random variables $\mathbf{x}(t)$, $t \in [0, \bar{t}]$ are known. A one-dimensional Brownian motion (Wiener process) b is defined for $t \in [0, \infty)$ as follows: each $b(t)$ is a Gaussian variable with mean zero and variance t. Further, the covariance of $b(t_1)$ and $b(t_2)$ equals $\min(t_1, t_2)$ and $b(0) = 0$. The corresponding probability space is denoted by $(\Omega_b, \Sigma_b, P_b)$. Here and in the sequel variable ν_b is reserved for the elements of the space Ω_b with a similar convention for the other probability spaces.

Continuous time white noise $\dot{b}$ is the weak derivative of Brownian motion. It is a generalized process with mean zero, infinite variance, and zero covariance. It is formalized here by means of Itô's integral, for which the reader is referred to the literature, e.g., [2, 22].

Systems of ordinary differential equations with white noise excitation are handled as Itô stochastic differential equations (SDEs):

$$\mathrm{d}\mathbf{x}(t) = \mathbf{f}\big(t, \mathbf{x}(t)\big)\,\mathrm{d}t + \mathbf{g}\big(t, \mathbf{x}(t)\big)\,\mathrm{d}b(t)$$

interpreted as the integral equation

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_0^t \mathbf{f}\big(s, \mathbf{x}(s)\big)\,\mathrm{d}s + \int_0^t \mathbf{g}\big(s, \mathbf{x}(s)\big)\,\mathrm{d}b(s)$$

where time t ranges in some finite time interval $[0, \bar{t}]$, $\mathbf{x}_0$ is a random variable representing the initial value, $\mathbf{f}, \mathbf{g} : [0, \bar{t}] \times \mathbb{R}^d \to \mathbb{R}^d$ are coefficient functions, and b is a one-dimensional Wiener process on $(\Omega_b, \Sigma_b, P_b)$. Their solutions are stochastic processes with continuous trajectories.

In [29] the authors have used white noise to model the base acceleration, for the sake of simplicity. White noise is—due to its covariance structure—a stationary process with a constant power spectral density. Hence, all frequencies appear equally in the base acceleration, which is a contradiction to the properties of most recorded ground motions. Furthermore, the infinite variance can actually not be interpreted physically. Thus, the main goal of the present paper is to use a more realistic model for ground acceleration.

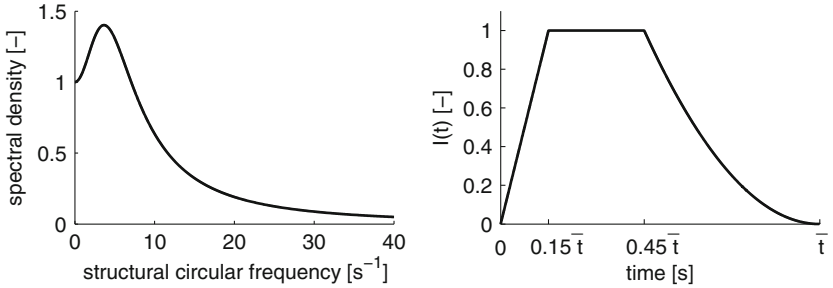

Fig. 3.2 *Left subplot*: power spectral density for Kanai–Tajimi model, $\omega_g = 5\,\mathrm{rad/s}$, $\zeta_g = 0.9$, $S_0 = 1$; *right subplot*: intensity function

In the literature (see, e.g., [4, 24, 33]) one can often find a model proposed by Kanai and Tajimi [15,34]. In their approach the base acceleration of the earth surface layer is approximated by the absolute acceleration of a linear SDOF oscillator excited by white noise. The corresponding equation of motion reads as follows

$$\ddot{z}(t) + 2\zeta_g\omega_g\dot{z}(t) + \omega_g^2 z(t) = -\dot{b}(t)$$

where ω_g and ζ_g are, respectively, the natural circular frequency and non-dimensional damping coefficient of the oscillator corresponding to the properties of the subsoil. The above equation can be written as a two-dimensional linear system of stochastic differential equations,

$$\mathrm{d}\mathbf{z}(t) = \begin{bmatrix} 0 & 1 \\ -\omega_g^2 & -2\zeta_g\omega_g \end{bmatrix} \mathbf{z}(t)\,\mathrm{d}t + \begin{bmatrix} 0 \\ -1 \end{bmatrix} \mathrm{d}b(t)$$

where $\mathbf{z} = [z, \dot{z}]^\mathsf{T}$. The ground acceleration is then modeled by the absolute acceleration of the oscillator, that is $\ddot{x}_g = \ddot{z} + \dot{b}$, which results in the following stochastic process

$$\ddot{x}_g(t) = -2\zeta_g\omega_g\dot{z}(t) - \omega_g^2 z(t) \tag{3.4}$$

Its power spectral density is given by the equation

$$S(\omega) = S_0\,\frac{\omega_g^4 + 4\zeta_g^2\omega_g^2\omega^2}{(\omega_g^2 - \omega^2)^2 + 4\zeta_g^2\omega_g^2\omega^2} \tag{3.5}$$

where S_0 is the (constant) power spectral density of white noise. Obviously, the spectral density S is not constant and thus represents a special type of colored noise. The left picture in Fig. 3.2 shows a plot of the power spectral density for the soil parameters $\omega_g = 5\,\mathrm{rad/s}$, $\zeta_g = 0.9$, and for the uniform spectral density $S_0 = 1$.

Note that the process given by Eq. (3.4) is (asymptotically) stationary. However, it is a well-known fact that the base acceleration of earthquakes is non-stationary.

Typically, at the beginning of the earthquake, amplitudes of the base acceleration are increasing. After a period of quasi-stationary strong motion, amplitudes are decreasing again. Thus, it seems reasonable to multiply the process from Eq. (3.4) with some intensity function I corresponding to the non-stationary behavior of the base acceleration (see, e.g., [6]). This leads to the following model for $\ddot{x}_g$:

$$\ddot{x}_g(t) = I(t)\big(-2\zeta_g\omega_g\dot{z}(t) - \omega_g^2 z(t)\big) \tag{3.6}$$

As suggested in [24] the intensity function plotted in the right-hand side picture of Fig. 3.2 is used: During the first 15 % of the total duration of the earthquake, I increases linearly from 0 to 1. After a constant period over 30 % of the earthquake duration, I decreases in a quadratic manner.

Rewriting the coupled equations of motion from (3.1) as a system of first order, and introducing the stochastic process (3.6) for the base acceleration leads to the following six-dimensional linear system of stochastic differential equations:

$$\mathrm{d}\mathbf{x}(t) = \mathbf{F}(t)\mathbf{x}(t)\,\mathrm{d}t + \mathbf{g}\,\mathrm{d}b(t) \tag{3.7}$$

where $\mathbf{x} = [x_s, x_d, z, \dot{x}_s, \dot{x}_d, \dot{z}]^\mathsf{T}$, $\mathbf{g} = [0,0,0,0,0,-1]^\mathsf{T}$ and

$$\mathbf{F}(t) = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -\omega_s^2 - \omega_d^2\mu & \omega_d^2\mu & -\omega_g^2 I(t) & -2\zeta_s\omega_s - 2\zeta_d\omega_d\mu & 2\zeta_d\omega_d\mu & -2\zeta_g\omega_g I(t) \\ \omega_d^2 & -\omega_d^2 & -\omega_g^2 I(t) & 2\zeta_d\omega_d & -2\zeta_d\omega_d & -2\zeta_g\omega_g I(t) \\ 0 & 0 & -\omega_g^2 & 0 & 0 & -2\zeta_g\omega_g \end{bmatrix}$$

Note that all parameters of the model are contained in the system matrix $\mathbf{F}$. Furthermore, $\mathbf{F}$ is time-dependent because of the intensity function I. This contrasts the situation considered in [29], where the system matrix does not depend on time.

For reasons of comparison, the response of the system without TMD is considered, too. In this case the motion of the corresponding SDOF oscillator is described by the equation

$$\ddot{\tilde{x}}_s + 2\zeta_s\omega_s\dot{\tilde{x}}_s + \omega_s^2\tilde{x}_s = -\ddot{x}_g$$

Substituting the stochastic process (3.6) for $\ddot{x}_g$ leads to the first order SDE system

$$\mathrm{d}\tilde{\mathbf{x}}(t) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ -\omega_s^2 & -\omega_g^2 & -2\zeta_s\omega_s & -2\zeta_g\omega_g \\ 0 & -\omega_g^2 & 0 & -2\zeta_g\omega_g \end{bmatrix} \tilde{\mathbf{x}}(t)\,\mathrm{d}t + \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \end{bmatrix} \mathrm{d}b(t) \tag{3.8}$$

where $\tilde{\mathbf{x}} = [x_s, z, \dot{x}_s, \dot{z}]^T$. In both systems (3.7) and (3.8) the initial values are assumed to be zero.

3.4 Modeling of the Parameter Uncertainty

The epistemic parameter uncertainty is accounted for by means of random sets. In general, a *random set* is a set-valued random variable satisfying certain measurability conditions. The simplest case arises when the underlying probability space is finite. In this case, one speaks of *finite random sets* or *Dempster–Shafer structures*. Such a structure is given by finitely many subsets $A_i, i = 1, \ldots, n$ of a given set $\mathbb{A}$, called the *focal elements*, each of which comes with a *probability weight* p_i, $\sum p_i = 1$.

For example, each set A_i could be the result of an interval-valued measurement and p_i its relative frequency in a sample. Alternatively, the sets A_i could be ranges of a variable obtained from source number i with relative credibility p_i.

As a random set, a Dempster–Shafer structure is viewed as given by an n-point probability space $\Omega_{\mathbb{A}} = \{1, 2, \ldots, n\}$ with probability masses $\{p_1, p_2, \ldots, p_n\}$. The assignment $i \to A_i$ is the defining set-valued random variable.

Following Dempster and Shafer [7, 30], two important set functions are introduced: the *lower probability* and the *upper probability* of an event B are defined by

$$\underline{P}(B) = \sum_{A_i \subset B} p_i, \qquad \overline{P}(B) = \sum_{A_i \cap B \neq \emptyset} p_i \tag{3.9}$$

A good visualization of a random set can be given through its *contour function* on the basic space $\mathbb{A}$, assigning each singleton a its upper probability:

$$a \to \overline{P}(\{a\}) \tag{3.10}$$

It is simply obtained by adding the probability weights p_i of those focal elements A_i to which a belongs. Figure 3.3 shows a random set and the resulting contour function where weights have been chosen as $p_1 = 1/2$, $p_2 = 1/3$, $p_3 = 1/6$.

In the sequel, random sets are needed that are defined on an arbitrary probability space $(\Omega_{\mathbb{A}}, \Sigma_{\mathbb{A}}, P_{\mathbb{A}})$ and whose values are subsets of p-dimensional coordinate space $\mathbb{A} = \mathbb{R}^p$. More precisely, random compact intervals are used, that is, random variables

$$A : \Omega_{\mathbb{A}} \to \mathscr{I}_c(\mathbb{A})$$

where $\mathscr{I}_c(\mathbb{A})$ denotes the set of all non-empty compact intervals in $\mathbb{A}$. Random compact intervals are used since they have and imply advantageous theoretical

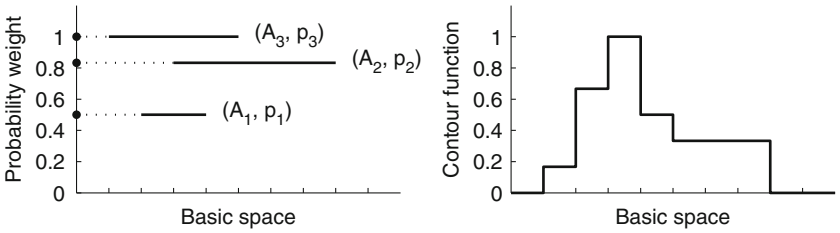

Fig. 3.3 A random set and its contour function; the *three dots* on the vertical axis symbolize the underlying three-point probability space

properties (e.g., concerning measurability, see [19]). In analogy to (3.9) the lower probability and the upper probability of an event B are defined as

$$\underline{P}(B)=P_{\mathbb{A}}(\{\nu_{\mathbb{A}} \in \Omega_{\mathbb{A}} : A(\nu_{\mathbb{A}}) \subseteq B\}), \quad \overline{P}(B)=P_{\mathbb{A}}(\{\nu_{\mathbb{A}} \in \Omega_{\mathbb{A}} : A(\nu_{\mathbb{A}}) \cap B \neq \emptyset\}) \tag{3.11}$$

The event B may be taken as any Borel measurable subset of $\mathbb{A}$. The contour function is given by (3.10).

For further details on interpretations and applications the reader is referred to the articles [10–12, 21, 35, 36] as well as to the monographs [19, 20].

3.4.1 Two Examples of Random Sets for Uncertainty Modeling

In [29] random sets constructed from Tchebycheff's inequality have been used, which require only minimal information about the parameters. More precisely, let a be an uncertain parameter preliminarily viewed as a random variable with expectation (or nominal) value $\overline{a}$ and variance σ^2. Then one can define a random set A on $\Omega_{\mathbb{A}} = (0, 1]$ by setting

$$A(\nu) = \left[\overline{a} - \tfrac{\sigma}{\sqrt{\nu}}, \overline{a} + \tfrac{\sigma}{\sqrt{\nu}}\right], \quad \nu \in \Omega_{\mathbb{A}} \tag{3.12}$$

where $\Omega_{\mathbb{A}} = (0, 1]$ is equipped with the uniform probability distribution. It has been argued in [29] that a focal element $A(\nu)$ may be viewed as an approximate two-sided $(1-\nu)$-fractile range for the parameter a. Furthermore, it has been explained how to compute σ from a probabilistic estimate about the range of the parameter. Figure 3.4 shows the contour function of a generic Tchebycheff random set.

In view of the shape of its contour function, it is obvious that a Tchebycheff random set is an appropriate model for parameter uncertainty when a parameter can take arbitrary (real) values, and negative deviations from the expectation or nominal value seem as likely as positive deviations. In case the parameter range is strictly bounded on (only) one side of the nominal value, it might be better to choose a random set whose contour function reflects this asymmetry.

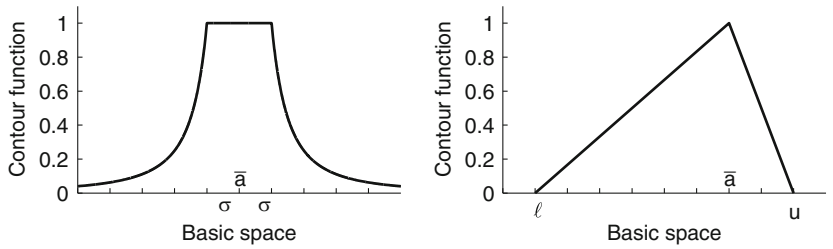

Fig. 3.4 *Left subplot*: a generic Tchebycheff random set; *right subplot*: a triangular random set

One possibility is to use a random set whose contour function has a triangular shape. The latter is then determined by a nominal value $\overline{a}$ and the lower and upper bounds ℓ, u of the parameter range. Such a random set can be defined on $\Omega_{\mathbb{A}} = [0, 1]$ by setting

$$A(\nu) = \big[\ell + (\overline{a} - \ell)\nu, u - (u - \overline{a})\nu\big], \quad \nu \in [0, 1]$$

where $\Omega_{\mathbb{A}} = [0, 1]$ is equipped with the uniform probability distribution. The right picture in Fig. 3.4 shows the contour function of a triangular random set.

3.5 Combination of Stochastic Excitation and Parameter Uncertainty

In the previous two sections it has been demonstrated how to model the base acceleration by stochastic processes and how to use random sets to model (epistemic) parameter uncertainty. The purpose of this section is to demonstrate how the two types of uncertainty can be combined to obtain set-valued assessments of the TMD performance.

As it has been shown at the end of Sect. 3.3, the motion of the combined structure-damper system is described by the linear system of SDEs (3.7), where the parameters μ, ζ_{s}, ω_{s}, ζ_{d}, ω_{d}, ζ_{g}, ω_{g} appear in the system matrix $\mathbf{F}$. In Sect. 3.6, various of these parameters are assumed to be uncertain, and random intervals presented in Sect. 3.4 are used. Note that this is in contrast to [29], where only TMD parameters were assumed to be uncertain. Corresponding to the situation, the tuple of uncertain parameters is denoted by a. The linear system (3.7) then reads

$$\mathrm{d}\mathbf{x}_a(t) = \mathbf{F}(t, a)\mathbf{x}_a(t)\,\mathrm{d}t + \mathbf{g}\,\mathrm{d}b(t)$$

where $\{\mathbf{x}_a(t)\}_{t\in[0,\bar{t}]}$ denotes the solution process corresponding to parameter value a.

As a first indicator for the performance of the TMD the displacement x_{s} of the structure is considered. More precisely, x_{s} is scaled by the largest structural displacement $\tilde{x}_{\mathrm{s}}$ when no TMD is attached. This leads to a map y defined on the

time interval, the set of possible parameter values $\mathbb{A}$, and the probability space Ω_b of Brownian motion:

$$y_a(t, \nu_b) = \frac{x_{s,a}(t, \nu_b)}{\max_{t\in[0,\bar{t}]} |\tilde{x}_{s,a}(t, \nu_b)|}$$

The latter can be seen as a non-dimensional displacement.

The aim is now to combine both kinds of uncertainty. To this end, the set-valued function

$$Y(t, \nu) = \{y_a(t, \nu_b) : a \in A(\nu_{\mathbb{A}})\} \tag{3.13}$$

is introduced, where $(t, \nu) \in [0, \bar{t}] \times \Omega$ and (Ω, Σ, P) denotes the product probability space

$$(\Omega, \Sigma, P) = (\Omega_{\mathbb{A}} \times \Omega_b, \Sigma_{\mathbb{A}} \otimes \Sigma_b, P_{\mathbb{A}} \otimes P_b)$$

This definition means that for each time t and each element $\nu = (\nu_{\mathbb{A}}, \nu_b)$ of the product space Ω the corresponding values of the non-dimensional displacement are merged to one set $Y(t, \nu)$, which is interpreted as containing the true value of the structural non-dimensional displacement. Note that Y is a set-valued stochastic process, that is, at each time t one has a random set $Y(t)$, whose values are compact intervals in $\mathbb{R}$. For further details the reader is referred to [27], where the theory of this approach has been developed.

For reasons of comparison the non-dimensional displacement $\tilde{y}$ of the structure without TMD

$$\tilde{y}_a(t, \nu_b) = \frac{\tilde{x}_{s,a}(t, \nu_b)}{\max_{t\in[0,\bar{t}]} |\tilde{x}_{s,a}(t, \nu_b)|}$$

is also considered. If parameters of the base acceleration are assumed to be uncertain, a set-valued process $\tilde{Y}$ can be defined from the processes $\tilde{y}_a$ in a similar manner as in Eq. (3.13). Furthermore, the absolute values of the non-dimensional displacements y_a are of interest, too, resulting in the set-valued process

$$|Y|(t, \nu) = \{|y_a(t, \nu_b)| : a \in A(\nu_{\mathbb{A}})\}$$

The reader is referred to [29] for equations of the boundary processes of $|\tilde{Y}|$ and their mean value functions.

A central concern is the effectiveness of the TMD, that is, to which extent the dynamic response x_s is reduced compared to the response $\tilde{x}_s$ when no TMD is attached. In the single-valued case the peak response reduction coefficient and the root mean square (RMS) response reduction coefficient are considered

$$r_{\mathrm{m},a}(\nu_b) = \frac{\max_{t\in[0,\bar{t}]} |x_{\mathrm{s},a}(t,\nu_b)|}{\max_{t\in[0,\bar{t}]} |\tilde{x}_{\mathrm{s},a}(t,\nu_b)|}, \quad r_{\mathrm{q},a}(\nu_b) = \sqrt{\frac{\int_0^{\bar{t}} x_{\mathrm{s},a}(t,\nu_b)^2\,\mathrm{d}t}{\int_0^{\bar{t}} \tilde{x}_{\mathrm{s},a}(t,\nu_b)^2\,\mathrm{d}t}}$$

For each parameter value a and for each path of the Brownian motion, the map $r_{\mathrm{m},a}$ represents the reduction of the peak displacement of the structure, whereas $r_{\mathrm{q},a}$ computes the quadratic-mean reduction (over time) of $x_{\mathrm{s},a}$. Similar as in Eq. (3.13) these maps can be extended to the set-valued reduction coefficients R_{m} and R_{q} defined by

$$R_{\mathrm{m}}(\nu) = \{r_{\mathrm{m},a}(\nu_b) : a \in A(\nu_{\mathbb{A}})\}, \quad R_{\mathrm{q}}(\nu) = \{r_{\mathrm{q},a}(\nu_b) : a \in A(\nu_{\mathbb{A}})\}$$

whose values are compact subintervals of the unit interval $[0, 1]$.

The stroke of the TMD is an important design parameter to assure the efficiency of the TMD, and to avoid damage of the TMD and/or of the main structure. It represents the TMD peak displacement with respect to its attachment point at the main structure. As with the reduction coefficients $r_{\mathrm{m},a}$ and $r_{\mathrm{q},a}$ the displacement $\tilde{x}_s$ of the structure without TMD is used for normalization. Thus, the equation for the stroke coefficient reads as follows [37]

$$d_a(\nu_b) = \frac{\max_{t\in[0,\bar{t}]} |x_{\mathrm{s},a}(t,\nu_b) - x_{\mathrm{d},a}(t,\nu_b)|}{\max_{t\in[0,\bar{t}]} |\tilde{x}_{\mathrm{s},a}(t,\nu_b)|}$$

Similar as in Eq. (3.13) one can define the set-valued stroke coefficient D by

$$D(\nu) = \{d_a(\nu_b) : a \in A(\nu_{\mathbb{A}})\}$$

whose values are compact intervals.

3.6 Numerical Simulation and Results

Subsequently, results of numerical simulations are presented. Unless otherwise stated, the results are based on the following nominal values: mass ratio $\mu = 0.05$, structural inherent damping $\zeta_{\mathrm{s}} = 0.005$, soil frequency $\omega_{\mathrm{g}} = 5\,\mathrm{rad/s}$ (soil period $T_{\mathrm{g}} \approx 1.26\,\mathrm{s}$), soil damping $\zeta_{\mathrm{g}} = 0.9$. The latter soil values correspond to soil class C according to Eurocode 8, see [9] and [23]. For the nominal values of the TMD parameters ω_{d} and ζ_{d} the optimal values given by Eqs. (3.2) and (3.3) are used.

For each tuple of parameter values approximations are computed using the Order 2 Implicit Strong Taylor Scheme (see [16]). Each simulation involves 500 sample functions of Brownian motion, and (constant) step size Δt of the time discretization is chosen as $\min(T_{\mathrm{s}}, T_{\mathrm{g}})/12$ if $\min(T_{\mathrm{s}}, T_{\mathrm{g}}) < 1$, or $1/20$ otherwise.

3.6.1 Parametric Studies

In this section the results of parametric studies are presented in an effort to reveal how the expectation values of the reduction coefficients and the stroke coefficient are influenced by structural and soil parameters μ, T_s, ζ_s, T_g, and ζ_g, respectively. In each study the structural period $T_s = 2\pi/\omega_s$ is varied in the range from 0.05 to 5 s. Additionally, one of the remaining parameters is varied while the other parameters are fixed to their nominal values. For each output variable a line plot (with the structural period T_s on the abscissa and one of the response quantities r_m, r_q, d on the ordinate) and a contour plot are presented.[1] All results are compared to the outcomes based on white noise base excitation.

Figure 3.5 shows the expected values of the peak and RMS displacement reduction coefficients r_m and r_q and the stroke coefficient d, respectively, for T_s and T_g varying in the range from 0.05 to 5 s. The bold black lines in the left pictures represent the results for white noise base acceleration (further investigated in [29]), whereas the thin (and partially marked) lines correspond to colored noise excitation for various values of the soil period T_g while fixing the soil damping ζ_g to the value 0.9.

Obviously, reduction coefficients r_m and r_q increase with increasing T_g particularly for short period structures, which means that the TMD is less effective for longer soil periods. However, the stroke coefficient is almost not affected by variation of T_g. The results of this figure suggest that for small structural periods T_s the TMD performance for colored noise excitation is worse than for white noise excitation. This behavior is coherent with computations accomplished with real earthquake records (see [37, 38]) and is due to the fact that the power spectral density (3.5) yields small values for high frequencies (small periods), whereas in the white noise case all frequencies equally likely appear. Another observation is that the smaller the soil period T_g the better the reduction plot approaches the white noise curve. For $T_g = 0.05$ s the expectations of the reduction coefficients actually coincide with those of the white noise case. Again, this can be explained by considering the spectral density of the colored noise process: If $\omega_g \to \infty$ (or equivalently $T_g \to 0$) then $S(\omega) \to S_0$ for all ω, that is, S converges to a constant spectral density, and this corresponds to white noise.

Figure 3.6 depicts the reduction and stroke coefficients for nominal soil frequency $\omega_g = 5$ rad/s ($T_g \approx 1.26$ s) and varying soil damping ζ_g ranging from 0.3 to 0.95. It is remarkable that for a structural period of approximately 1.12 s all values of ζ_g lead to the same reduction. In the structural period range larger than this period the considered response quantities remain almost unaffected by the variation of soil damping. However, in the lower period range RMS and peak reduction coefficients increase considerably with decreasing soil damping, i.e., the TMD becomes less effective.

[1] In the sequel, we shall denote the reduction and stroke coefficients simply by r_m, r_q, d in place of $r_{m,a}$, $r_{q,a}$, d_a, unless explicit reference to a specific tuple of parameters a is required.

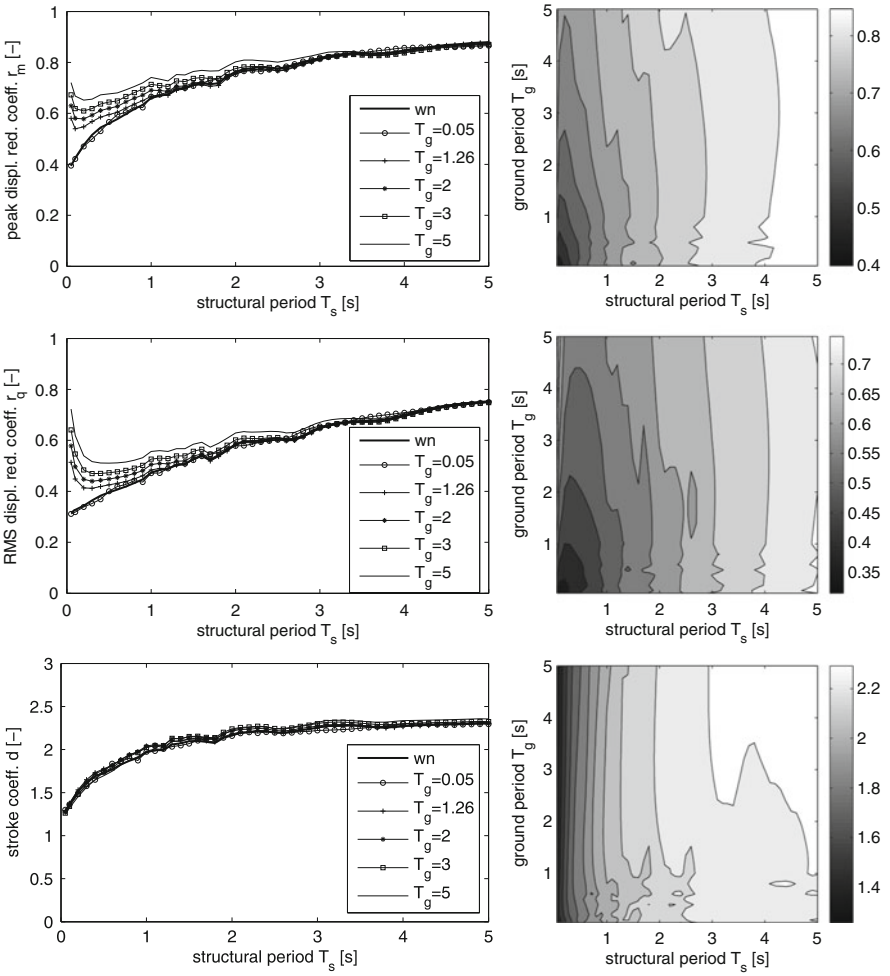

Fig. 3.5 Expectations of reduction coefficients r_{m}, r_{q} and stroke coefficient d, based on white noise excitation (wn) and colored noise excitation for $\zeta_{\mathrm{g}} = 0.9$ and various values of T_{g} [in (s)]

Figure 3.7 shows the behavior of the output variables r_{m}, r_{q}, and d under variation of mass ratio μ in the range of 0.5–8 % based on colored noise excitation with nominal soil parameters ($\omega_{\mathrm{g}} = 5\,\mathrm{rad/s}$, $\zeta_{\mathrm{g}} = 0.9$). One can see that all three output variables decrease when μ increases. This confirms the well-known fact that for larger mass ratios structural displacement is reduced more efficiently and the stroke coefficient is smaller. It is remarkable that for small mass ratios (0.5 % and 1 %) the stroke coefficient depends on the structural period in a non-monotonic manner. From Fig. 3.8 one can conclude that in the case of white noise excitation results are very similar to those based on colored noise excitation, except

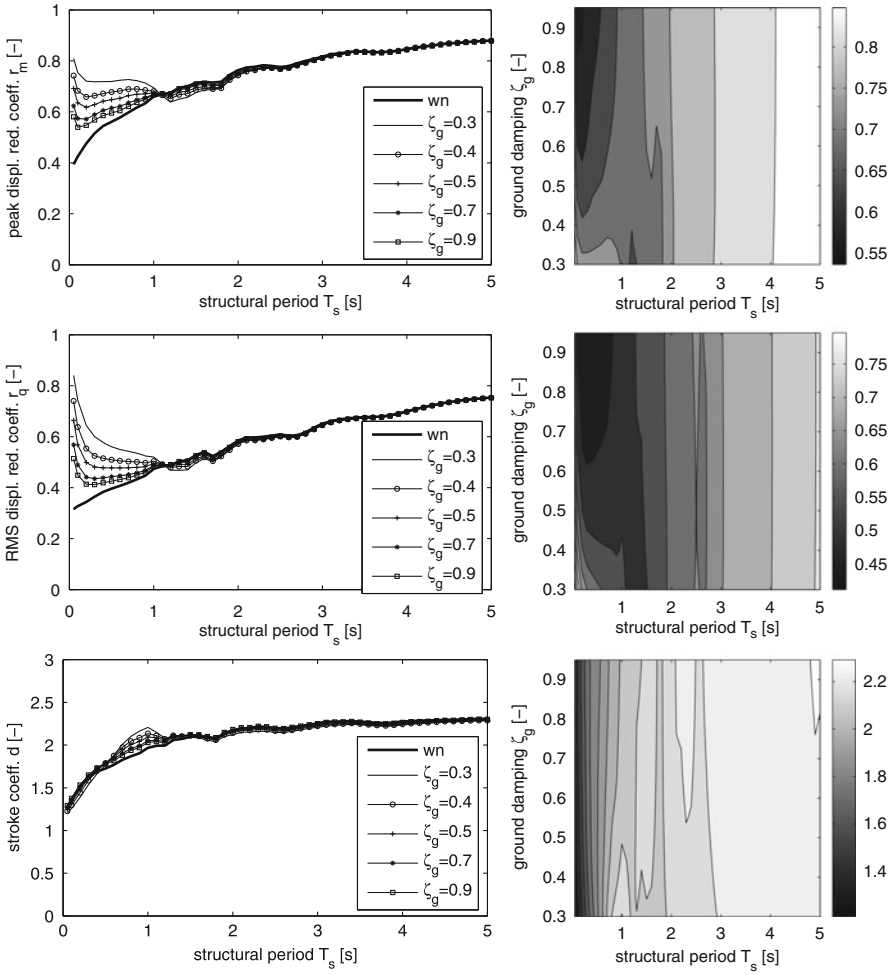

Fig. 3.6 Expectations of reduction coefficients r_m, r_q and stroke coefficient d, respectively, based on white noise excitation and colored noise excitation for $\omega_\mathrm{g} = 5\,\mathrm{rad/s}$ ($T_\mathrm{g} \approx 1.26\,\mathrm{s}$) and various values of ζ_g

for small structural periods, which is coherent with the results displayed in Figs. 3.5 and 3.6. Comparing the results of Figs. 3.7 and 3.8 with outcomes of a study [37] based on a set of recorded ground motions, reveals that not only the dependency of the considered response variables on various structural parameters is the same for the stochastic soil model used here and for real ground motions. These response quantities are even of the same order of magnitude. The approach of this study is thus confirmed.

In Fig. 3.9 results for r_m, r_q, and d are plotted when the structural damping coefficient ζ_s is varied in the range of 0.5–5 %, mass ratio $\mu = 5\,\%$, and the

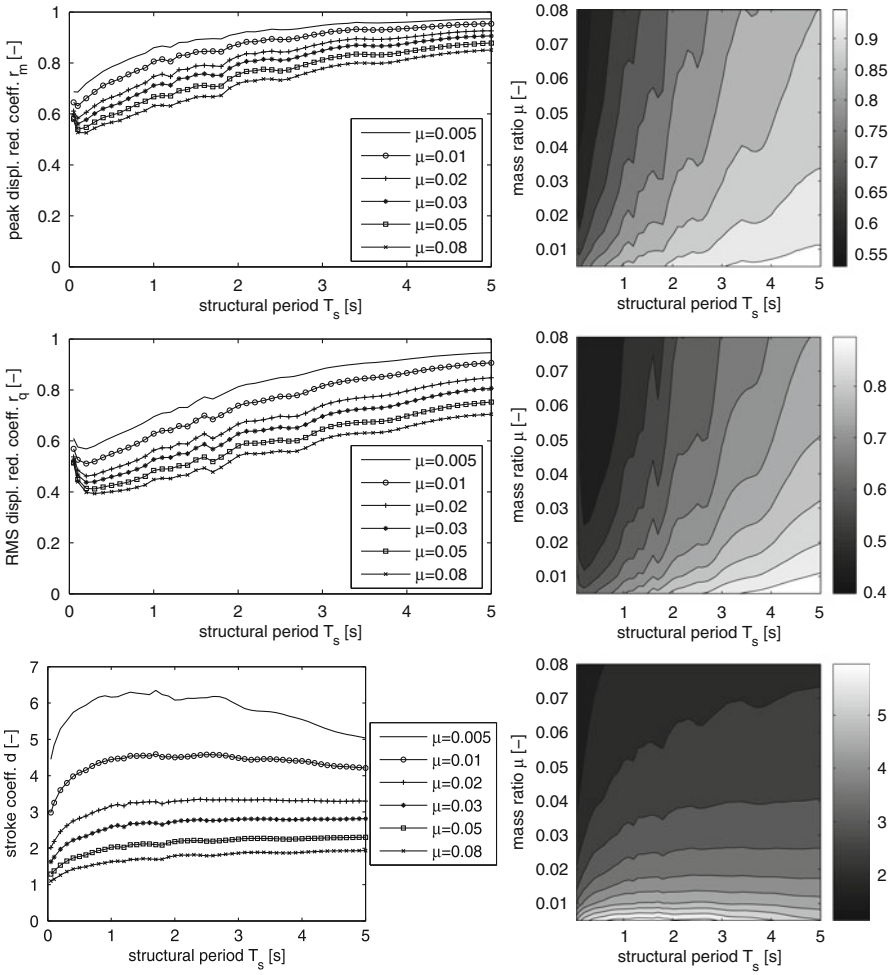

Fig. 3.7 Expectations of reduction coefficients r_{m}, r_{q} and stroke coefficient d based on colored noise excitation ($\omega_{\mathrm{g}} = 5\,\mathrm{rad/s}$, $\zeta_{\mathrm{g}} = 0.9$) for $\zeta_{\mathrm{s}} = 0.005$ and various values of μ, respectively

Kanai–Tajimi model with nominal soil parameters ($\omega_{\mathrm{g}} = 5\,\mathrm{rad/s}$, $\zeta_{\mathrm{g}} = 0.9$) is used. It is readily observed that all three output variables increase when ζ_{s} increases. This means that higher inherent structural damping leads to lower effectiveness of the TMD to reduce the response and to a larger stroke relative to the peak displacement of the main system. This outcome is obvious because the main system becomes less-vibration prone the larger the inherent damping is. From Fig. 3.10 one can see once more that in the case of white noise excitation results are very similar to those based on colored noise excitation, except for small structural periods.

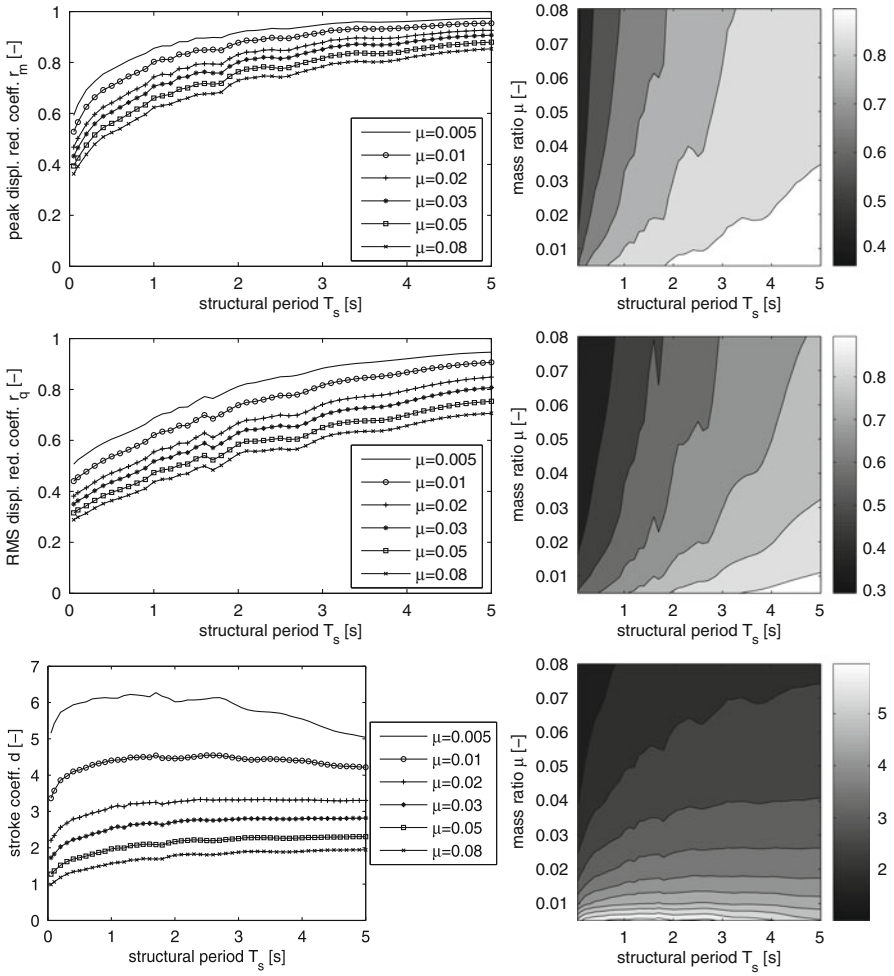

Fig. 3.8 Expectations of reduction coefficients r_{m}, r_{q} and stroke coefficient d, respectively, based on white noise excitation for $\zeta_{\mathrm{s}} = 0.005$ and various values of μ

3.6.2 Set-Valued TMD Parameters

In this subsection the mass ratio μ, the structural inherent damping ζ_{s}, and the soil parameters ω_{g} and ζ_{g} are fixed to their nominal values whereas the TMD parameters ω_{d} and ζ_{d} are assumed to be uncertain. In a first simulation, a Tchebycheff random set A is used only for ω_{d}, and ζ_{d} is assumed to take its nominal value. Concerning the variability it is assumed that the actual value of ω_{d} lies in a range of $\pm 40\,\%$ of its nominal value with 99 % certainty. As explained in [29] this leads to a coefficient of variation of 0.04, that is, $\sigma = 0.04\overline{\omega}_{\mathrm{d}}$. Corresponding to Eq. (3.12) the focal

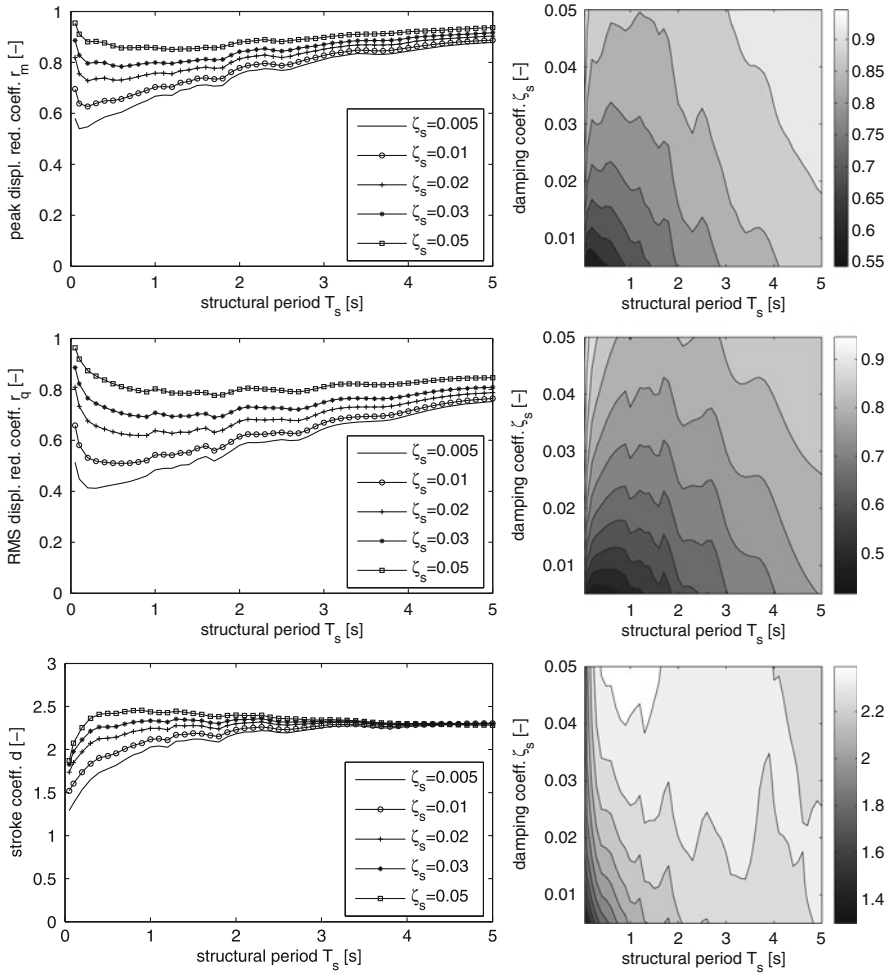

Fig. 3.9 Expectations of reduction coefficients r_m, r_q and stroke coefficient d, respectively, based on colored noise excitation ($\omega_\mathrm{g} = 5\,\mathrm{rad/s}$, $\zeta_\mathrm{g} = 0.9$) for $\mu = 0.05$ and various values of ζ_s

elements are obtained as

$$A(\nu_\mathbb{A}) = \Big[\overline{\omega}_\mathrm{d}\Big(1 - \frac{0.04}{\sqrt{\nu_\mathbb{A}}}\Big), \overline{\omega}_\mathrm{d}\Big(1 + \frac{0.04}{\sqrt{\nu_\mathrm{A}}}\Big)\Big]$$

where $\nu_\mathbb{A} \in (0, 1]$. For the numerical simulation the random set is approximated by a finite random set consisting of the ten focal elements obtained for $\nu_{\mathbb{A},j} = (j/10)^2$, $j = 1, \ldots, 10$. The corresponding weights are then given by $p_1 = 0.01$, $p_j = \nu_{\mathbb{A},j} - \nu_{\mathbb{A},j-1} = (2j-1)/100$, $j = 2, \ldots, 10$. This leads to a better approximation (with respect to the upper and lower probability) than choosing $\nu_{\mathbb{A},j}$ equidistantly from $(0, 1]$.

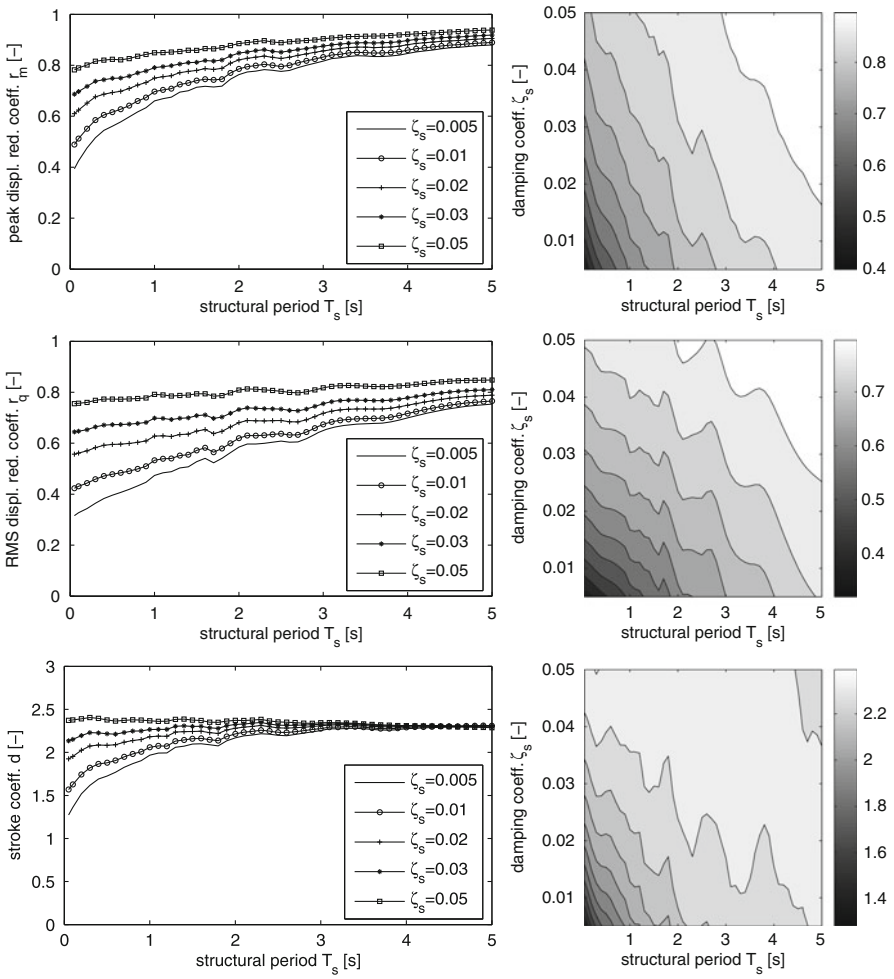

Fig. 3.10 Expectations of reduction coefficients r_{m}, r_{q} and stroke coefficient d, respectively, based on white noise excitation for $\mu = 0.05$ and various values of ζ_{s}

Figure 3.11 shows the expectation of the peak and RMS displacement reduction coefficients and the expectation of the stroke coefficient for 11 different values of the structural period T_{s}, namely, 0.05, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5 (values in seconds). The outer (bold) lines in the pictures are the interval bounds of the expectation of the set-valued reduction coefficients R_{m}, R_{q} and the set-valued stroke coefficient D, respectively. The central (marked) lines represent the output obtained for the optimal parameter value $\overline{\omega}_{\mathrm{d}}$.

Figure 3.12 is obtained by using a Tchebycheff random set for the TMD damping coefficient ζ_{d} with coefficient of variation of 0.04 and fixing ω_{d} to its optimal value. It seems that varying ζ_{d} has almost no influence on the reduction coefficients, which

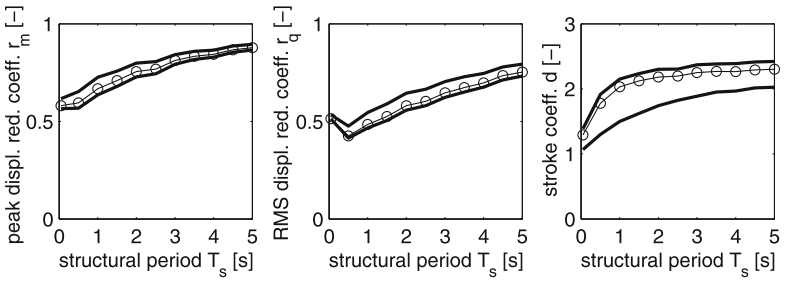

Fig. 3.11 *Left subplot*: bounds of set-valued peak displacement reduction coefficient R_{m} (*outer lines*), peak displacement reduction coefficient $r_{\mathrm{m},\bar{a}}$ (*central line*) for different values of T_{s}, and uncertain TMD frequency ω_{d}; *middle subplot*: bounds of set-valued RMS displacement reduction coefficient R_{q} (*outer lines*), RMS displacement reduction coefficient $r_{\mathrm{q},\bar{a}}$ (*central line*) for different values of T_{s}, and uncertain TMD frequency ω_{d}; *right subplot*: bounds of set-valued stroke coefficient D (*outer lines*), stroke coefficient $d_{\bar{a}}$ (*central line*) for different values of T_{s}, and uncertain TMD frequency ω_{d}

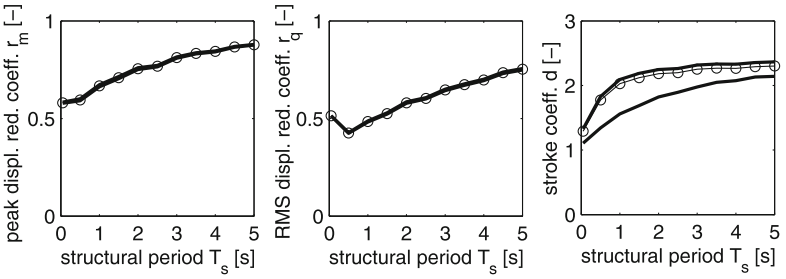

Fig. 3.12 *Left subplot*: bounds of set-valued peak displacement reduction coefficient R_{m} (*outer lines*), peak displacement reduction coefficient $r_{\mathrm{m},\bar{a}}$ (*central line*) for different values of T_{s}, and uncertain TMD damping ζ_{d}; *middle subplot*: bounds of set-valued RMS displacement reduction coefficient R_{q} (*outer lines*), RMS displacement reduction coefficient $r_{\mathrm{q},\bar{a}}$ (*central line*) for different values of T_{s}, and uncertain TMD damping ζ_{d}; *right subplot*: bounds of set-valued stroke coefficient D (*outer lines*), stroke coefficient $d_{\bar{a}}$ (*central line*) for different values of T_{s}, and uncertain TMD damping ζ_{d}

is coherent with the outcomes of [37, 38] based on real recorded ground motions. However, the stroke coefficient is influenced by ζ_{d} in a similar manner as by ω_{d}, i.e., the bounds of the set-valued stroke coefficient are only slightly tighter as in the right picture of Fig. 3.11.

In a further simulation, for both TMD parameters ω_{d} and ζ_{d} Tchebycheff random sets are used. Their approximations are combined to a two-dimensional random set by taking the cartesian product of each of the focal elements of the first with each of the focal elements of the second random set and multiplying the corresponding probability weights. This results in a finite random set consisting of 100 rectangular focal elements. In Fig. 3.13 the expectations of the reduction coefficients and the stroke coefficient are depicted. The plots of the reduction coefficients look very similar to the ones in Fig. 3.11, which emphasizes that the impact of TMD damping ζ_{d} on the reduction coefficient is small.

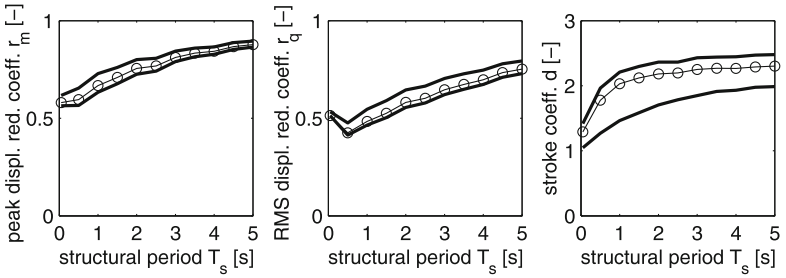

Fig. 3.13 *Left subplot*: bounds of set-valued peak displacement reduction coefficient R_{m} (*outer lines*), peak displacement reduction coefficient $r_{\mathrm{m},\overline{a}}$ (*central line*) for different values of T_{s}, and uncertain TMD parameters ω_{d}, ζ_{d}; *middle subplot*: bounds of set-valued RMS displacement reduction coefficient R_{q} (*outer lines*), RMS displacement reduction coefficient $r_{\mathrm{q},\overline{a}}$ (*central line*) for different values of T_{s}, and uncertain TMD parameters ω_{d}, ζ_{d}; *right subplot*: bounds of set-valued stroke coefficient D (*outer lines*), stroke coefficient $d_{\overline{a}}$ (*central line*) for different values of T_{s}, and uncertain TMD parameters ω_{d}, ζ_{d}

From all three figures one can see that the RMS displacement reduction coefficients are smaller than the peak displacement reduction coefficients. This is due to the fact that in the left pictures only the peak displacements of the trajectories are compared. These maximum displacements usually appear during the period of strong ground motion (where the intensity function equals 1). On the other hand, for the RMS displacement reduction all the displacements observed during the time interval are taken into account. Furthermore, one can observe that the bounds of the set-valued stroke coefficient are much wider than for the reduction coefficients, and that the stroke coefficient $d_{\overline{a}}$ induced by the optimal values of the TMD parameters are close to the upper bound of the set-valued stroke coefficient. These results lead to the well-known conclusion that the optimal TMD parameters from Eq. (3.3) lead to a large stroke, but by variation of the TMD parameters the stroke can be diminished considerably while the efficiency of the TMD is only deteriorating slightly, see, e.g., [37] .

Figure 3.14 shows the bounds of sample functions of the non-dimensional displacement Y of the load-bearing structure (bold lines) obtained by choosing two particular focal elements, a particular path of the ground motion process, and $T_{\mathrm{s}} = 1\,\mathrm{s}$. Thin lines represent the corresponding sample functions of the non-dimensional displacement $y_{\overline{a}}$ obtained for the nominal parameter values $\overline{a} = (\overline{\omega}_{\mathrm{d}}, \overline{\zeta}_{\mathrm{d}})$ and the non-dimensional structural displacement $\tilde{y}$ when no TMD is attached. In the left subplot of Fig. 3.15 the mean value functions of $|Y|$, $|y_{\overline{a}}|$ and $|\tilde{y}|$ for $T_{\mathrm{s}} = 1\,\mathrm{s}$ are plotted. One can see that during the phase of strong ground motion the displacements of the load-bearing system damped by the TMD are fluctuating around a constant value. Due to the increased damping by the TMD, these displacements decay much more quickly than the displacements of the TMD-free system after the end of the strong motion period. The right-hand subplot of Fig. 3.15 depicts for each time the probability that the non-dimensional

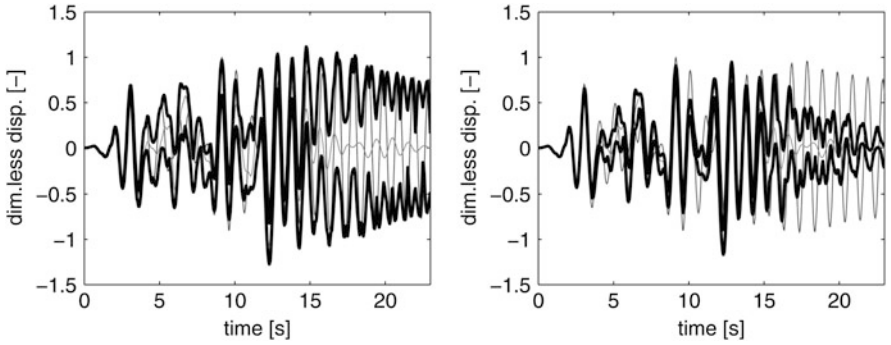

Fig. 3.14 Bounds of a sample function of the non-dimensional structural displacement Y (*bold lines*) and sample functions of $y_{\overline{a}}$ (*central thin line*) and $\tilde{y}$ (*outer thin line*) for $T_s = 1\,\mathrm{s}$ and two different focal elements for uncertain TMD parameters ω_d, ζ_d

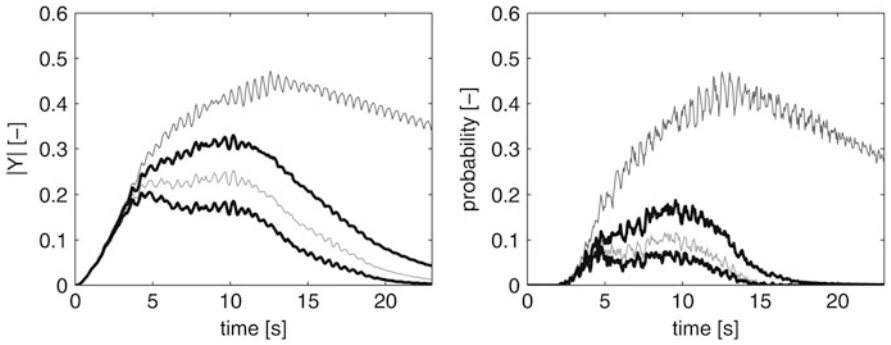

Fig. 3.15 *Left subplot*: mean value functions of the absolute values of the non-dimensional structural displacement $|Y|$ (*bold lines*), $|y_{\overline{a}}|$ (*central thin line*) and $|\tilde{y}|$ (*outer thin line*) for $T_s = 1\,\mathrm{s}$, and uncertain TMD parameters ω_d, ζ_d; *right subplot*: upper and lower probabilities of $[0.5, \infty)$ for $|Y|$ (*bold lines*), probabilities of $|y_{\overline{a}}| > 0.5$ (*central thin line*) and $|\tilde{y}| > 0.5$ (*outer thin line*), TMD parameters ω_d and ζ_d uncertain

displacement exceeds the value 0.5. For the set-valued process $|Y|$ this corresponds to the upper and lower probabilities of the interval $[0.5, \infty)$ (see Eq. (3.11)).

3.6.3 Set-Valued Soil Parameters

In this subsection, results of simulations are discussed when random sets are used for the soil parameters ω_g, ζ_g whereas the mass ratio, the structural damping, and the TMD parameters are fixed to their nominal values.

Figure 3.16 shows the expectations of the reduction coefficients and the stroke coefficient for soil damping $\zeta_g = 0.9$ and a Tchebycheff random set with nominal soil frequency $\overline{\omega}_g = 5\,\mathrm{rad/s}$. The coefficient of variation used for ω_g is 0.04.

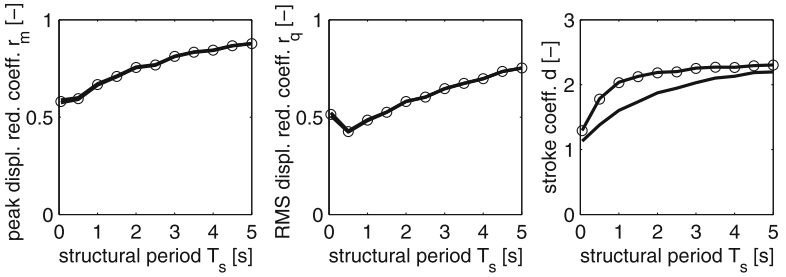

Fig. 3.16 *Left subplot*: bounds of set-valued peak displacement reduction coefficient R_{m} (*outer lines*), peak displacement reduction coefficient $r_{\mathrm{m},\bar{a}}$ (*central line*) for different values of T_{s}, and uncertain soil frequency ω_{g}; *middle subplot*: bounds of set-valued RMS displacement reduction coefficient R_{q} (*outer lines*), RMS displacement reduction coefficient $r_{\mathrm{q},\bar{a}}$ (*central line*) for different values of T_{s}, and uncertain soil frequency ω_{g}; *right subplot*: bounds of set-valued stroke coefficient D (*outer lines*), stroke coefficient $d_{\bar{a}}$ (*central line*) for different values of T_{s}, and uncertain soil frequency ω_{g}

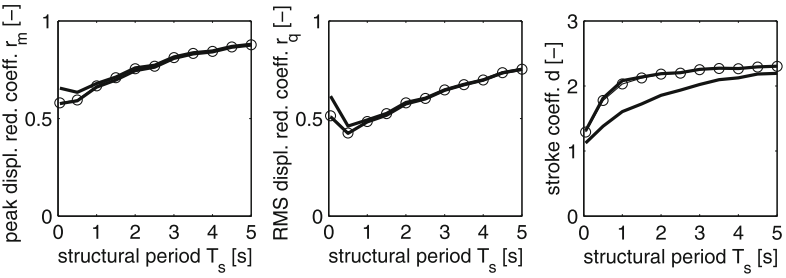

Fig. 3.17 *Left subplot*: bounds of set-valued peak displacement reduction coefficient R_{m} (*outer lines*), peak displacement reduction coefficient $r_{\mathrm{m},\bar{a}}$ (*central line*) for different values of T_{s}, and uncertain soil damping ζ_{g}; *middle subplot*: bounds of set-valued RMS displacement reduction coefficient R_{q} (*outer lines*), RMS displacement reduction coefficient $r_{\mathrm{q},\bar{a}}$ (*central line*) for different values of T_{s}, and uncertain soil damping ζ_{g}; *right subplot*: bounds of set-valued stroke coefficient D (*outer lines*), stroke coefficient $d_{\bar{a}}$ (*central line*) for different values of T_{s}, and uncertain soil damping ζ_{g}

Obviously, varying ω_{g} changes the reduction coefficients only slightly whereas the stroke coefficient is affected considerably. One can further consider the case where $\omega_{\mathrm{g}} = 5\,\mathrm{rad/s}$ and a random set is used for ζ_{g}. As before the nominal value of 0.9 is employed for soil damping ζ_{g}. Concerning the variability it is assumed that ζ_{g} can take values from 0.3 to 0.95; note that ζ_{g} is bounded by 1. This range does not lie symmetrically around the nominal value, and thus it is not appropriate to use a Tchebycheff random set. However, it seems reasonable to utilize a triangular random set instead as shown in Fig. 3.4, right subplot. The latter is approximated by the finite random set obtained by the choices $v_{\mathbb{A},j} = 0.01 + 0.11 \cdot (j-1)$, $j = 1, \ldots, 10$, with probability weights $p_1 = 0.01$, $p_j = 0.11$, $j = 2, \ldots, 10$. Figure 3.17 depicts the expectations of the resulting reduction coefficients and the

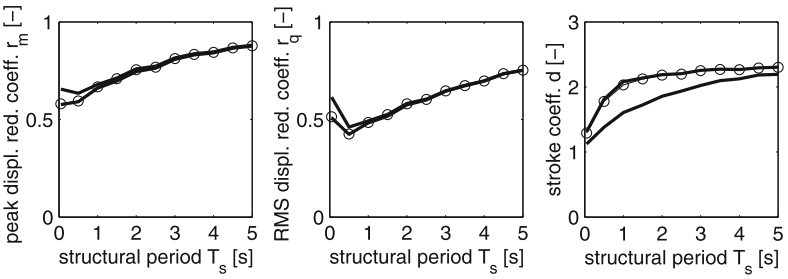

Fig. 3.18 *Left subplot*: bounds of set-valued peak displacement reduction coefficient R_{m} (*outer lines*), peak displacement reduction coefficient $r_{\mathrm{m},\bar{a}}$ (*central line*) for different values of T_{s}, and uncertain soil parameters ω_{g}, ζ_{g}; *middle subplot*: bounds of set-valued RMS displacement reduction coefficient R_{q} (*outer lines*), RMS displacement reduction coefficient $r_{\mathrm{q},\bar{a}}$ (*central line*) for different values of T_{s}, and uncertain soil parameters ω_{g}, ζ_{g}; *right subplot*: bounds of set-valued stroke coefficient D (*outer lines*), stroke coefficient $d_{\bar{a}}$ (*central line*) for different values of T_{s}, and uncertain soil parameters ω_{g}, ζ_{g}

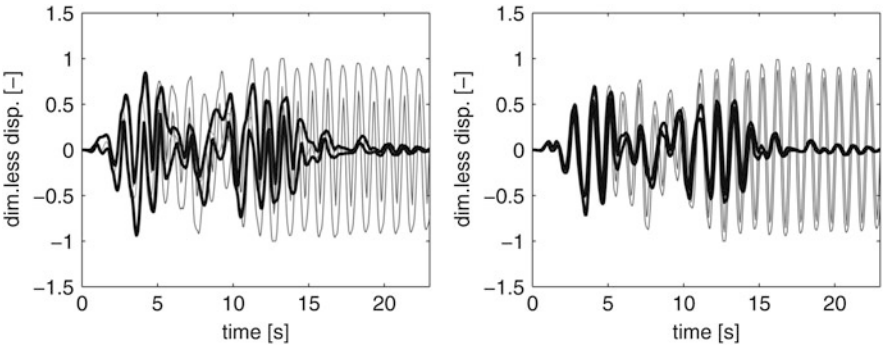

Fig. 3.19 Bounds of sample functions of the non-dimensional structural displacement Y (*bold lines*), sample functions of $y_{\bar{a}}$ (*central thin line*) and bounds of sample functions of $\tilde{Y}$ (*outer thin lines*) for $T_{\mathrm{s}} = 1$ s and two different focal elements for uncertain soil parameters ω_{g}, ζ_{g}

stroke coefficient. Obviously, for the reduction coefficients significant deviations from the nominal values can only be recognized for structural periods T_{s} up to 1 s. For larger values of T_{s} the bounds of the set-valued reduction coefficients more or less coincide with the reductions computed with the nominal values. Similar to Fig. 3.16 the stroke coefficient varies considerably. Very similar outcomes are found when using random sets for both parameters ω_{g} and ζ_{g} (see Fig. 3.18). Figure 3.19 shows sample functions, and in Fig. 3.20 mean value functions and exceedance probabilities are plotted.

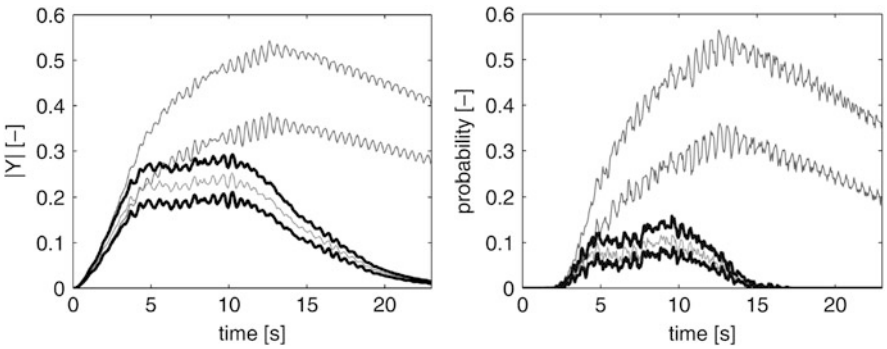

Fig. 3.20 *Left subplot*: mean value functions of the absolute values of non-dimensional structural displacement $|Y|$ (*bold lines*), $|y_{\overline{a}}|$ (*central thin line*) and $|\widetilde{Y}|$ (*outer thin lines*) for $T_s = 1$ s, soil parameters ω_g and ζ_g uncertain; *right subplot*: upper and lower probabilities of $[0.5, \infty)$ for $|Y|$ (*bold lines*) and $|\widetilde{Y}|$ (*outer thin lines*), probabilities of $|y_{\overline{a}}| > 0.5$ (*central thin line*), soil parameters ω_g and ζ_g uncertain

3.7 Conclusion

In this paper a framework to assess the seismic performance of Tuned Mass Dampers (TMDs) in presence of parameter uncertainty has been presented. A stochastic process, based on the Kanai–Tajimi power spectral density function, models earthquake excitation. This constitutes a more realistic excitation model than white noise used in an earlier study [29]. Random sets have been used to describe the uncertainty of the ground parameters and the TMD parameters, which can (in practice) not be tuned optimally. The benefit is an adequate assessment of response reduction coefficients of the main system and the stroke coefficient of the TMD system. The interval-valued description of the behavior of the TMD system is more informative and reliable than a purely stochastic description with single-valued outputs.

Based on this methodology a parametric study has been conducted to quantify the efficiency of a TMD to reduce the seismic response of a vibration-prone structure that can be modeled sufficiently accurately as a single degree-of-freedom oscillator. The results derived are coherent with the outcomes of a similar parametric study [37] that is, however, based on a set of recorded earthquake ground motions. The considered response quantities are both qualitatively and quantitatively comparable, and thus the analytical expression for seismic TMD design presented in [37] is confirmed. Beneficially, the utilized stochastic ground motion model allows one to study the effect of a targeted variation of specific ground motion parameters on the TMD performance, as it has been conducted here.

References

1. Adam, C., Furtmüller, T.: Seismic performance of Tuned Mass Dampers. In: Irschik, H., Krommer, M., Watanabe, K. (eds.) Mechanics and Model Based Control of Smart Materials and Structures, pp. 11–18. Springer, Wien (2010)
2. Arnold, L.: Stochastic Differential Equations: Theory and Applications. Wiley, New York (1974)
3. Ayorinide, E.O., Warburton, G.B.: Minimizing structural vibrations with absorbers. Earthquake Eng. Struct. Dyn. **8**, 219–236 (1980)
4. Bucher, C.: Computational Analysis of Randomness in Structural Mechanics. CRC Press/Balkema, Leiden (2009)
5. Casciati, F., Giuliano, F.: Performance of multi-TMD in the towers of suspension bridges. J. Vibr. Contr. **15**, 821–847 (2009)
6. Clough, R.W., Penzien, J.: Dynamics of Structures. Mc Graw-Hill, Auckland (1975)
7. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat. **38**, 325–339 (1967)
8. Den Hartog, J.P.: Mechanical Vibrations. 4th edn. McGraw-Hill, New York (1956)
9. Building Code EN 1998-1: Eurocode 8: Design of structures for earthquake resistance – Part 1: General rules, seismic actions and rules for buildings, 2005.
10. Fetz, T., Oberguggenberger, M.: Propagation of uncertainty through multivariate functions in the framework of sets of probability measures. Reliab. Eng. Syst. Safety **85**, 73–87 (2004)
11. Goodman, I.R., Nguyen, H.T.: Fuzziness and randomness. In: Bertoluzza, C., Gil, M.Á., Ralescu, D.A. (eds.) Statistical Modeling Analysis and Management of Fuzzy Data. Physica, Heidelberg (2002)
12. Hall, J., Rubio, E., Anderson, M.: Random sets of probability measures in slope hydrology and stability analysis. Zeitschrift für Angewandte Mathematik und Mechanik **84**, 710–720 (2004)
13. Hoang, N., Fujino, Y., Warnitchai, P.: Optimal Tuned Mass Damper for seismic applications and practical design formulas. Eng. Struct. **30**, 707–715 (2008)
14. Jensen, H., Setareh, M., Peek, R.: TMDs for vibration control of systems with uncertain properties. J. Struct. Eng. **118**, 3285–3296 (1992)
15. Kanai, K.: Semi-empirical formula for the seismic characteristics of the ground motion. Bull. Earthquake Res. Inst. Univ. Tokyo **35**, 309–325 (1957)
16. Kloeden, P.E., Platen, E.: Numerical Solution of Stochastic Differential Equations. Springer, Berlin (1992)
17. Leung, A.Y.T., Zhang, H., Chen, C.C., Lee, Y.Y.: Particle swarm optimization of TMD by non-stationary base excitation during earthquake. Earthquake Eng. Struct. Dyn. **37**, 1223–1246 (2008)
18. Meinhardt, C.: Experimental damping assessments of tall buildings to verify the effectivity of damping devices for high rise structures. In: Proceedings of the International Conference on Highrise Towers and Tall Buildings 2010, Munich, Germany, CD-ROM paper, 14–16 April 2010
19. Molchanov, I.: Theory of Random Sets. Springer, Berlin (2005)
20. Nguyen, H.T.: An Introduction to Random Sets. Chapman and Hall/CRC Press, Boca Raton (2006)
21. Oberguggenberger, M.: The mathematics of uncertainty: models, methods and interpretations. In: Fellin, W., Lessmann, H., Oberguggenberger, M., Vieider, R. (eds.) Analyzing Uncertainty in Civil Engineering. Springer, Berlin (2005)
22. Øksendal, B.: Stochastic Differential Equations. An Introduction with Applications. Springer, Berlin (1998)
23. Peil, U., Clobes, M.: Erdbebenbeanspruchung abgespannter Maste. Bauingenieur **87**, 124–129 (2012)
24. Rackwitz, R.: Einwirkungen auf Bauwerke. In: Mehlhorn, G. (ed.) Der Ingenieurbau: Grundwissen, Bd. 8, Tragwerkszuverlässigkeit/Einwirkungen, pp. 73–416. Ernst & Sohn, Berlin (1997)

25. Rofooei, F.R., Mobarake, A., Ahmadi, G.: Generation of artificial earthquake records with a nonstationary Kanai-Tajimi model. Eng. Struct. **23**, 827–837 (2001)
26. Sadek, F., Mohraz, B., Taylor, A.W., Chung, R.M.: A method for estimating the parameters of Tuned Mass Dampers for seismic applications. Earthquake Eng. Struct. Dyn. **26**, 617–635 (1997)
27. Schmelzer, B.: On solutions of stochastic differential equations with parameters modelled by random sets. Int. J. Approx. Reason. **51**, 367–376 (2010)
28. Schmelzer, B., Adam, C., Oberguggenberger, M.: Seismic performance of Tuned Mass Dampers with uncertain parameters. In: Eberhardsteiner, J., Böhm, H.J., Rammerstorfer, F.G. (eds.) Proceedings of the 6th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2012), p. 20, Vienna, Austria, CD-ROM paper, 10–14 September 2012
29. Schmelzer, B., Oberguggenberger, M., Adam, C.: Efficiency of tuned mass dampers with uncertain parameters on the performance of structures under stochastic excitation. Proceedings of the Institution of Mechanical Engineers. Part O J. Risk Reliab. **224**, 297–308 (2010)
30. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
31. Soong, T.T.: Active Structural Control: The Theory and Practice. Longman Scientific and Technical Series, New York (1990)
32. Soong, T.T., Dargush, G.F.: Passive Energy Dissipation Systems in Structural Engineering. Wiley, Chichester (1997)
33. Soong, T.T., Grigoriu, M.: Random Vibration of Mechanical and Structural Systems. Prentice-Hall, Englewood Cliffs (1993)
34. Tajimi, H.: A statistical method of determining the maximum response of a building structure during an earthquake. In: Proceedings of the 2nd Conference on Earthquake Engineering, vol. 2, pp. 781–798. Science Council of Japan, Tokyo (1960)
35. Tonon, F., Bernardini, A., Mammino, A.: Determination of parameters range in rock engineering by means of random set theory. Reliab. Eng. Syst. Saf. **70**, 241–261 (2000)
36. Tonon, F., Bernardini, A., Mammino, A.: Reliability analysis of rock mass response by means of random set theory. Reliab. Eng. Syst. Saf. **70**, 263–282 (2000)
37. Tributsch, A., Adam, C.: Evaluation and analytical approximation of Tuned Mass Damper performance in an earthquake environment. Smart Struct. Syst. **10**, 155–179 (2012)
38. Tributsch, A., Adam, C., Furtmüller, T.: Mitigation of earthquake induced vibrations by Tuned Mass Dampers. In: De Roeck, G., Degrande, G., Lombaert, G., Müller, G. (eds.) Structural Dynamics - EURODYN2011, Proceedings of 8th European Conference on Structural Dynamics, pp. 1742–1749, Leuven, Belgium, CD-ROM paper, 4–6 July 2011
39. Wang, J.-F., Lin, C.-C., Lian, C.-H.: Two-stage optimum design of tuned mass dampers with consideration of stroke. Struct. Control Health Monit. **16**, 55–72 (2009)
40. Warburton, G.B.: Optimum absorber parameters for various combinations of response and excitation parameters. Earthquake Eng. Struct. Dyn. **10**, 381–401 (1982)

Chapter 4
Sensitivity and Reliability Analysis of Engineering Structures: Sampling Based Methods

M. Oberguggenberger

Abstract This chapter intends to present an overview of Monte Carlo-type methods currently in use in the probabilistic analysis of large engineering structures. It starts with an introduction to the generation of multi-dimensional random quantities. Next, spatially distributed random properties, e.g., material or geometrical properties in continuum mechanics, are modeled as random fields. Approximations to random fields by means of Karhunen–Loève expansion and polynomial chaos expansion are introduced. These tools are employed to study the response of continuous structures with loads, material or geometrical properties given by random fields. The main focus is on sensitivity analysis of large engineering structures, where small Monte Carlo sample sizes are mandatory. The transition to reliability is undertaken by means of the concept of tolerance intervals. Further, current sampling methods for accurate reliability estimates are discussed, and practical applications are presented.

4.1 Introduction

Engineering structures are usually modelled as input–output maps: the response Y is a function $Y = g(X_1, \ldots, X_n)$ of input parameters $(X_1, \ldots, X_n)$ like material properties, geometry, boundary conditions, and driving forces (dynamic or distributed loads, noise). It has been acknowledged since a long time that both the structural model (given by the function g) and the input parameters are uncertain. Traditionally, uncertainties have been dealt with by employing safety factors. That is, the traditional codes would require that the load carrying capacity of the structure exceeds the design loads by a certain factor > 1, typically 1.35 for permanent loads (such as dead weight) and 1.5–2.0 for temporary loads.

M. Oberguggenberger (✉)
Unit of Engineering Mathematics, University of Innsbruck, Technikerstr. 13, A6020 Innsbruck, Austria
e-mail: Michael.Oberguggenberger@uibk.ac.at

G. Hofstetter (ed.), *Computational Engineering*, DOI 10.1007/978-3-319-05933-4_4,

© Springer International Publishing Switzerland 2014

This state of affairs is unsatisfactory in as much as no information about the actual distance to failure can be extracted. The desire for a more analytical description of the uncertainties led to the introduction of the probabilistic safety concept in civil engineering, initiated by the pioneering work of Freudenthal [17], Bolotin [6], and others in the 1950s. Starting with the 1980s and 1990s, the European engineering codes have been changed into probability based codes. By now, this is the standard in civil engineering (see, e.g., EN 1990:2002 [15])—interestingly, the civil engineering community has been far ahead of the other engineering fields in adopting the probabilistic point of view.

Under this point of view, every relevant parameter of the engineering model is a random variable. There is no absolute safety, but rather a probability of failure. As a consequence, more information than just the nominal parameter values must be entered in the model, namely a description of the statistical distribution of the input. Further, the response is no longer deterministic, but rather a random variable, whose distribution must be computed in order to describe the behavior of the structure as well as the probability that certain limits are exceeded (described by a limit state function).

In practical applications, the structure is usually represented by a finite element model. These models are generally large, computationally costly, and partially black boxes. Practically, Monte Carlo simulation is the only way to numerically compute the statistics of the system response. Thereby, an artificial sample of $X_1, \ldots, X_n$, a data matrix of size $N \times n$, is generated and N values of $y = g(x_1, \ldots, x_n)$ are calculated, producing a sample of size N of the response Y, which in turn can be evaluated statistically. This approach raises the computational cost dramatically, and so the need for cost-saving algorithms arises.

An adequate understanding of the uncertainties in an engineering task requires a number of actions, among them reflection about the choice of model and the failure mechanisms; assessing the variability of input and output variables and model parameters; sensitivity analysis (i.e., the determination of the relative influence of individual input parameters on the response); assessing the reliability of the structure. This involves a variety of activities to be performed, from laboratory experiments, data collection to model validation.

The reader is alerted that in the present contribution, only the comparatively narrow part of the numerical calculation of sensitivities and of reliability is addressed. As suggested in the title, the focus is on sampling based methods. In view of the need to employ as few model evaluations as possible, the choice of the sample becomes an important issue. This is dealt with under the heading *design of experiment* in Sect. 4.2. In due course, *metamodels* will be encountered there as well. Section 4.3 is devoted to the simulation of *random fields*, that is, spatially distributed random input. Section 4.4 starts off with *sensitivity analysis*, of interest in itself, but also the basis for model reduction. This becomes useful in Sect. 4.5, where *reliability analysis* is addressed. In Sect. 4.6, the concepts will be illustrated using a model from aerospace engineering, supplied by our industrial partner Intales GmbH Engineering Solutions.

The methods presented here have been developed, adapted, and implemented in a number of joint research projects with Intales GmbH Engineering Solutions.[1] Note that approaches to uncertainty analysis going beyond probability theory, such as interval analysis or the combination of both approaches in the form of random sets, are not addressed here. One instance of such a hybrid approach is in Chap. 3 of this volume. For further information the reader is referred to the recent surveys [4, 36].

4.2 Design of Experiment

In this section, the task of simulating the output $Y = g(X_1, \ldots, X_n)$ of an input–output function applied to random input $(X_1, \ldots, X_n)$ will be addressed. Direct Monte Carlo simulation consists in generating a sample $\mathbf{x}_1, \ldots, \mathbf{x}_N$ of the n-dimensional random variable $(X_1, \ldots, X_n)$, collected in an $N \times n$-matrix.[2] The sample has to be generated in such a way that the columns are statistically independent and each of them is distributed according to the distribution of the corresponding random variable. We are not going to detail this step—most scientific software packages come with a pseudorandom number generator that can produce high dimensional independent samples of most familiar statistical distributions [42] of sufficiently large size (the crucial question of accuracy will be addressed below). The term *design of experiment* refers to the choice of the sample so as to achieve certain desirable additional properties.

Subsequently, each sampled row $\mathbf{x}_j = (x_{ji}, \ldots, x_{jn})$ is sent through the input–output map to produce a sample $y_j = g(x_{j1}, \ldots, x_{jn})$, $j = 1, \ldots, N$ of the output Y.

The complete information about the statistical properties of the output Y is contained in its cumulative distribution function

$$F_Y(y) = P(Y \leq y) = P(g(X_1, \ldots, X_n) \leq y)$$

which in turn can be written as an expectation value, namely as

$$F_Y(y) = \mathrm{E}\big(h(Y)\big) = \mathrm{E}\big(h(g(X_1, \ldots, X_n)\big)$$

where h is the indicator function of the interval $(-\infty, y]$, i.e., $h(z) = 1$ for $z \leq y$ and 0 otherwise. Similarly, all statistical properties of the output Y can

[1] ICONA-project 2006–2008, supported by TransIT Innsbruck, ACOSTA-project 2008–2010, supported by The Austrian Research Promotion Agency, MDP-NE 2011–2013, supported by Astrium GmbH; main partners: Intales GmbH Engineering Solutions, Institute of Basic Sciences in Engineering Science and Institute of Mathematics, University of Innsbruck, Czech Technical University in Prague.

[2] We follow the common statistical practice that random variables are denoted by capital letters, while their realizations are denoted by small letters.

be formulated in terms of expectation values of functions of Y. For example, the moments of Y are obtained by choosing $h(z) = z^m$, $m = 1, 2, 3, \ldots$. The core of Monte Carlo simulation is that these expectation values can be approximated by the corresponding sample mean, that is,

$$\mathrm{E}\big(h(Y)\big) \approx \overline{h(Y)} = \frac{1}{N}\sum_{j=1}^{N} h(y_j) = \frac{1}{N}\sum_{j=1}^{N} h\big(g(x_{ji}, \ldots, x_{jn})\big).$$

By construction, $y_1, \ldots, y_N$ is an independent random sample, hence statistical sampling theory tells us that the variance of the estimator $\overline{h(Y)}$ is given by

$$\mathrm{V}\big(\overline{h(Y)}\big) = \frac{1}{N}\mathrm{V}\big(h(Y)\big) = \frac{C^2}{N}$$

where C^2 is the variance of $h(Y) = h\big(g(X_1, \ldots, X_n)\big)$, a fixed number depending only on h, g, and the given distribution of $(X_1, \ldots, X_n)$. Thus the mean error of a Monte Carlo estimate is of order $1/\sqrt{N}$. For methods to generate random samples leading to a numerical error approximately below prescribed bounds see [19].

We note in passing that replacing the pseudorandom numbers by quasirandom numbers, generated from the so-called low-discrepancy sequences, allows one to improve the mean square error to order $(\log N)^n/N$, but demonstrably *not* further [13, 34]. Rather than going into design of experiment based on quasirandom numbers, two sampling plans will be addressed which are of bigger importance in our setting.

Latin Hypercube Sampling The first issue is stratified sampling that is designed to avoid random clustering and produces sampled points with a balanced distribution over the parameter space. A prominent and easy-to-implement method of stratified sampling is *Latin hypercube sampling*. To obtain a sample of size N, the Latin hypercube sampling plan divides the range of each variable X_i into N disjoint subintervals of equal probability. First, N values of each variable X_i, $i = 1, \ldots, n$, belonging to the respective subintervals are randomly selected. Then the N values for X_1 are randomly paired without replacement with the N values for X_2. The resulting pairs are then randomly combined with the N values of X_3 and so on, until a set of N n-tuples is obtained. This set forms the Latin hypercube sample. The advantage of Latin hypercube sampling is that sampled points are evenly distributed through design space, thereby hitting also regions of low probability possibly important for the input–output map which might be missed by direct Monte Carlo simulation. A Latin hypercube estimate is not necessarily more accurate than a standard Monte Carlo estimate at given N, but it can be shown that the variance of a Latin hypercube estimator is asymptotically smaller than the variance of the direct Monte Carlo estimator, and possibly markedly smaller when the input–output map is partially monotonic [53].

Correlation Control The second issue is correlation control, which is an essential ingredient in Monte Carlo simulation with small sample sizes (say, around $N = 100$ or less). As the reader may easily verify, the rows of an independently sampled matrix $\mathbf{X} = (x_{ij}, i = 1, \ldots, N; j = 1, \ldots n)$ of independent random variables $X_1, \ldots, X_n$ may turn out to have correlation coefficients up to 20 % in practice, when N is that small (this undesirable effect disappears for $N \approx 1{,}000$). Thanks to an empirical method due to Iman and Conover [21], it is possible to rearrange the entries of the sampled matrix in such a way that the new columns are nearly uncorrelated. In fact, the method allows one to construct a matrix $\mathbf{X}^*$ of any desired correlation structure $\mathbf{K}$. This is done as follows. The van der Waerden matrix $\mathbf{W}$ is defined by

$$\mathbf{W} = \begin{pmatrix} w_1^{(1)} & \ldots & w_1^{(n)} \\ \vdots & & \vdots \\ w_N^{(1)} & \ldots & w_N^{(n)} \end{pmatrix}$$

where each column consists of a random permutation of the van der Waerden scores

$$\Phi^{-1}\Big(\frac{j}{N+1}\Big), \qquad j = 1, \ldots, N.$$

Here Φ denotes the standard normal cumulative distribution function. Starting with the Cholesky factorizations

$$\mathbf{K} = \mathbf{P}\mathbf{P}^{\mathsf{T}}, \quad \boldsymbol{\rho}_W = \mathbf{Q}\mathbf{Q}^{\mathsf{T}},$$

with the correlation matrix $\boldsymbol{\rho}_W$ of $\mathbf{W}$, one can prove that

$$\mathbf{W}^* = \mathbf{W}\mathbf{Q}^{-\mathsf{T}}\mathbf{P}^{\mathsf{T}}$$

has the target correlation structure $\mathbf{K}$. Empirical investigations [21] showed that the rank correlation matrix of the resulting matrix $\mathbf{W}^*$ is nearly the same, i.e., $\rho_{W^*} \approx \mathbf{R}_{W^*}$. Therefore, rearranging the values in the columns of $\mathbf{X}$ corresponding to the rank order of the columns in $\mathbf{W}^*$ leads to a matrix $\mathbf{X}^*$ which approximately has the desired correlation structure:

$$\rho_{X^*} = \rho_{W^*} \approx \mathbf{R}_{W^*} = \mathbf{K}.$$

Further improvement can usually be achieved by iteration of the procedure. It should be noted that the described method of correlation control does not destroy the Latin hypercube structure of a sample and thus can be directly combined with Latin hypercube sampling. The efficiency of correlation control in dependence on the number of input variables has been studied in [37].

In all simulations presented in this paper, Latin hypercube sampling and correlation control have been routinely implemented.

Bootstrap Resampling The result of a Monte Carlo simulation is a single estimate $\overline{h(Y)}$ of one or more desired quantities $h(Y)$. One would like to be able to assess the accuracy of the estimate, i.e., the variance of the estimator, confidence intervals, etc., without additional calls of the expensive input–output map. A cost-saving method to achieve this is *bootstrap resampling* [14, 50]. The bootstrapping procedure consists in repeatedly drawing from the same sample with replacement to obtain new samples of the same size N. To obtain $B = 1{,}000$, say, bootstrap samples of size N, one proceeds as follows.

From the original data sample, e.g., $h(y_1), \ldots, h(y_N)$, of size N one randomly draws N-times, so that each realization has equal probability of being drawn. The results are combined to produce a bootstrap sample of size N (note that some entries of the bootstrap sample may be repetitions of realizations of the original data sample). This is repeated $B = 1{,}000$ times. The $B = 1{,}000$ bootstrap samples now are used to compute $B = 1{,}000$ realizations of $\overline{h(Y)}$, say, and the distribution of $\overline{h(Y)}$ can be estimated in this way. The reason why this works is that each bootstrap sample has a distribution which approximates the empirical distribution of the original sample.

In this way, the generation of a multitude of samples of the same distribution is mimicked and allows one to assess the variability of the individual sample estimators. For example, confidence intervals for $\overline{h(Y)}$ can be either obtained by computing the percentiles of the $B = 1{,}000$ estimates of $\overline{h(Y)}$, or approximated by a Student's t-distribution based on the empirical standard deviation of those values.

Metamodels Also known as surrogate models or response surfaces, metamodels attempt to save computational cost by approximating the input–output function by a simpler (deterministic) function. Typically, such an approximation is based on evaluating the input–output function at a smaller number of design points and suitable extrapolation. Large size Monte Carlo simulation can then be performed with the metamodel with little computational cost. Metamodels obtained by linear regression with possibly nonlinear shape functions have the advantage that the powerful diagnostic methods of regression analysis can be used. For example, partial coefficients of determination admit to quantify the relative importance of input variables with respect to the variability of the output in nonparametric ways [28, 32, 39]. Other metamodels are based on radial basis functions, smoothing splines, or Kriging (i.e., variance minimizing piecewise linear extrapolation); see, e.g., [25, 45]. The accuracy of a metamodel crucially depends on the degree of smoothness of the input–output function. Metamodels cannot be used, e.g., to accurately describe nonlinear bifurcation as in buckling analysis. On the other hand, given a sufficiently smooth model, the accuracy of a metamodel can be controlled by sequential design of experiment, in which additional design points are added in regions of lower accuracy, optimizing both the global error and the space filling properties of the experimental design, see, e.g., [23].

Stochastic Response Surfaces If the input variables $(X_1, \ldots, X_n)$ are Gaussian or have been transformed into standard Gaussian variables, the input–output function can be seen as a function on standard Gaussian space and approximated by a response surface on that space. As an illustration, consider the univariate case of a single random variable X with distribution function $F(x)$. The transformed random variable $U = \Phi^{-1}(F(X))$ has a standard Gaussian distribution. This transformation reduces the input–output map to a function of the Gaussian variable U as well, by means of $Y(U) = g\big(F^{-1}(\Phi(U))\big)$. For example, if X has a normal distribution with mean μ and variance σ^2, the transformation is simply $X = F^{-1}(\Phi(U)) = \mu + \sigma U$.

Recall that the Hermite polynomials $h_n(u)$ form an orthonormal basis in the space of square integrable functions on the real line with respect to the Gaussian density $\mathrm{e}^{-u^2/2}\,\mathrm{d}u/\sqrt{2\pi}$. The (normalized) Hermite polynomials are given by the recursion

$$h_{n+1}(u) = \frac{u}{\sqrt{n+1}} h_n(u) + \frac{n}{\sqrt{n(n+1)}} h_{n-1}(u)$$

with $h_0(u) = 1, h_1(u) = u$. Every function $Y(U)$ of a Gaussian variable U such that $Y(U)$ has finite second moments has a convergent Hermite expansion of the form

$$Y(U) = \sum_{k=0}^{\infty} c_k h_k(U).$$

The coefficients c_k can be obtained as the inner product $\mathrm{E}(Y(U)h_n(U))$; alternatively, collocation and regression can be used to numerically compute them. More precisely, choose finitely many points $\xi_1, \ldots, \xi_m$ in the domain of the input–output map g. Compute collocation points $u_j = \Phi^{-1}(F(\xi_j))$ on the real line. Record the outputs $y_j = g(\xi_j) = g\big(F^{-1}(\Phi(u_j))\big)$. Evaluate the coefficients $c_1, \ldots, c_M$ of a truncated Hermite expansion by linear regression on the data (u_j, y_j) with the Hermite functions h_k, $k = 1, \ldots, M$, as shape functions. This concludes the construction of a stochastic response surface $Y_M(U)$ for the input–output function, given as a truncated Hermite series. Monte Carlo simulation is now done at no cost by sampling a standard Gaussian variable U and evaluating $Y_M(U)$.

Note that this procedure requires only m evaluations of the costly input–output map on the points $\xi_1, \ldots, \xi_m$. The rest of the burden is put on the transformation $U = \Phi^{-1}(F(X))$, thus parametric studies with differently distributed X, for example, with varying mean and variance, can be easily undertaken. In the one-dimensional case a low value of M, say around ten, and twice the number of collocation points usually suffices. For an application of this method, see, e.g., [35].

As is well-known, the procedure can be generalized to multiple expansions of functions of infinitely many Gaussian variables, known as the polynomial chaos expansion; the reader is referred to e.g., [18, 29].

4.3 Random Fields

Material and geometrical properties (e.g., modulus of elasticity, thickness) of a structure may vary randomly from point to point. Such a behavior can be captured by means of *random fields*, that is, stochastic processes that assign a random variable $q(x)$ to every point x in a region in space. Usually, random fields are chosen so as to have continuous or even differentiable realizations, as opposed to random noise in stochastic mechanics. To define the field, the joint distributions of the values at any finite number of points $q(x_1), \dots, q(x_k)$ should be specified. If the random field is stationary (i.e., the finite dimensional distributions are translation invariant) and Gaussian, it is completely specified by the mean value $\mu_q = \mathrm{E}(q(x))$ and the second moments, i.e., the covariance $\mathrm{COV}(q(x), q(y))$ for any two points x, y. Due to stationarity, the covariance depends only on the distance $\delta = |x - y|$ of the points and is of the form

$$\mathrm{COV}(q(x), q(y)) = C(x, y) = \sigma^2 c(\delta)$$

with the variance σ^2 and the autocorrelation function $c(\delta)$. A frequently used autocorrelation function is of the form

$$c(\delta) = \exp\big(-|\delta|/\ell\big), \tag{4.1}$$

where ℓ is the so-called correlation length. The indicator function of the interval $[-\ell, \ell]$ might be taken as a crude autocorrelation function with correlation equal to 1 for δ in the interval and 0 outside. The area under the curve (4.1) is the same as the area under this indicator function, whence the name correlation length. Other autocorrelation functions in use may be of Gaussian type, in higher dimensions also with anisotropic distance measure.

If measurement data are available, the autocorrelation function can be estimated from the empirical covariance matrix by arranging the values along the distance δ of the measured points and fitting a shape function as in (4.1), thereby estimating σ^2 and ℓ.

In order to simulate a random field, one discretizes the region under consideration with grid points x_i, $i = 1, \dots, M$, measures the distance between the grid points x_i, and sets up a covariance matrix $\mathbf{C} = (C_{ij}, i, j = 1, \dots, M)$ whose values are computed from (4.1) where δ is the distance between x_i and x_j. In case the random field is Gaussian, there are at least three methods to generate realizations of the field.

The first method is the standard simulation method for correlated Gaussian variables. It is based on the Cholesky factorization $\mathbf{C} = \mathbf{A}\mathbf{A}^\mathsf{T}$. If $\mathbf{Y}$ is an M-dimensional Gaussian random variable with mean zero and independent components (i.e., its covariance matrix $\mathrm{E}(\mathbf{Y}\mathbf{Y}^\mathsf{T}) = \mathbf{I}$, the identity matrix), then $\mathbf{X} = \mathbf{A}\mathbf{Y}$ is a mean-zero Gaussian random variable whose covariance matrix is $\mathbf{C}$. This follows from the simple identities

$$\mathrm{E}(\mathbf{X}\mathbf{X}^\mathsf{T}) = \mathrm{E}(\mathbf{A}\mathbf{Y}\mathbf{Y}^\mathsf{T}\mathbf{A}^\mathsf{T}) = \mathbf{A}\mathrm{E}(\mathbf{Y}\mathbf{Y}^\mathsf{T})\mathbf{A}^\mathsf{T} = \mathbf{A}\mathbf{I}\mathbf{A}^\mathsf{T} = \mathbf{C}.$$

Accordingly, starting from a realization of the M-dimensional standard Gaussian random variable $\mathbf{Y}$, the transformation $\mathbf{X} = \mu_q + \mathbf{A}\mathbf{Y}$ yields a realization of the desired random field $q(x)$ in the grid points, i.e., $X_i = q(x_i)$. This is repeated N times to obtain a Monte Carlo sample of the random field. It should be noted that this method works in any space dimension; it just requires enumerating the grid points and keeping track of their distance. The disadvantage of this method is that one cannot easily keep track of the error in terms of the number of grid points, that is, the accuracy of the autocovariance function of the simulated field.

The second method is advantageous in this respect. It is based on the *Karhunen–Loève expansion* of the field. In fact, the eigenvalue problem

$$\int C(x, y)\varphi_k(y)\,\mathrm{d}y = \lambda_k \varphi_k(x)$$

where $C(x, y)$ is the autocovariance function of the random field, has a sequence of positive eigenvalues λ_k and orthonormal eigenfunctions $\varphi_k(x)$ (orthonormality in mean square). Then

$$q(x) = \sum_{k=1}^{\infty} \sqrt{\lambda_k}\xi_k \varphi_k(x) \tag{4.2}$$

where the ξ_k are uncorrelated random variables with unit variance, see, e.g., [30]. If the process is Gaussian, the ξ_k are independent and distributed according to the standard normal distribution $\mathcal{N}(0, 1)$.

For the numerical simulation, the spatial region is again discretized by a grid and the φ_k are taken, e.g., piecewise constant on the grid elements. The eigenvalue problem becomes a matrix eigenvalue problem, and the series (4.2), with approximate eigenvalues and eigenfunctions, truncated after a finite number M of terms, can be used for Monte Carlo simulation of the field trajectories. Here the mean square error due to truncation after M terms is just the sum of the neglected eigenvalues; the discretization error can be estimated through the numerical integration error and its propagation through eigenvalue problems [5]. A further advantage of the method is that it can be directly based on a finite element discretization [46]. However, changing the field parameters, e.g., the correlation length, requires solving the eigenvalue problem with a different matrix anew, which can be costly.

This disadvantage is avoided in the third method, which is applicable in one space dimension. It is based on the observation that the autocorrelation function (4.1) coincides with the autocorrelation function of an Ornstein–Uhlenbeck process, namely the solution process of the Langevin stochastic differential equation

$$\mathrm{d}q(x) = -\tfrac{1}{\ell}q(x) + \sqrt{\tfrac{2}{\ell}}\,\sigma\,\mathrm{d}w(x), \quad q(0) \sim \mathcal{N}(0, \sigma^2), \tag{4.3}$$

where $w(x)$ denotes Wiener process on the real line, see e.g. [2]. Solutions of sufficient accuracy can be easily simulated from discrete white noise input by means of an explicit Euler scheme at little cost [26].

4.4 Sensitivity Analysis

Sensitivity analysis is a core ingredient in understanding the behavior of a structural model. It aims at determining the input parameters that have the largest influence on critical output. In addition, it can be used as a first step in reliability analysis or in optimizing structural properties.

Sensitivity analysis does not necessarily require knowledge of the probabilistic properties of the input and thus is a nonparametric method. If the input–output function is explicitly given and sufficiently smooth, one may use partial derivatives to assess the sensitivity, see, e.g., [40, 45]. In the context we envisage, the input–output function may be non-differentiable and a black box, in addition. For this reason we focus on derivative-free methods of sensitivity analysis, that is, on sampling based methods. The strategy is to produce a sample of the input data $(X_1, \ldots, X_n)$, to compute a sample of the output Y and to analyze the statistical input–output relations or the relations between different output quantities. For all variables, we take a uniform distribution centered around the nominal values μ_j with a spread of a certain equal percentage, say $\pm 15\,\%$. Equally scaled spread and uniform distributions are chosen to avoid distortion of the relative weights of the input variables. If information about the actual statistical distribution of one or the other input variable is known, this knowledge first does not enter in the sensitivity analysis, but may be considered in a second stage.

The computationally least expensive methods are correlation based, which will be described first. An explorative analysis usually starts with inspecting scatterplots of individual variables vs. output. To obtain a refined diagnosis, methods are needed that quantify the correlations, assess their significance, and possibly remove hidden influences of the co-variates on the correlation between a given input variable and the output variable. The simplest indicator is the Pearson correlation coefficient (CC). It detects linear relationships between input and output. To recall the definition, assume given a sample $x_{j1}, \ldots, x_{jn}$, and y_j, $j = 1, \ldots, N$ of the n-dimensional input and the corresponding output. Denote the mean values by $\overline{x}_i$ and $\overline{y}$, respectively. The empirical Pearson correlation coefficient of input number i with the output is defined as

$$r(x_i, y) = \frac{\sum_{j=1}^{N}(x_{ji} - \overline{x}_i)(y_j - \overline{y})}{\sqrt{\sum_{j=1}^{N}(x_{ji} - \overline{x}_i)^2 \sum_{j=1}^{N}(y_j - \overline{y})^2}}.$$

It turns out that the correlation coefficient does not isolate the effect of x_i on the output y, but is influenced by the co-variates (inputs with numbers $k \neq i$),

especially when they have a nonzero correlation with x_i. Regression based indices may be used to mitigate this effect.

A brief recall of linear regression is in order. The goal of linear regression analysis is to fit a linear model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$ to the data, that is, each value y_j is to be approximated by

$$y_j = \beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \ldots + \beta_n x_{jn} + \varepsilon_j, \quad j = 1, \ldots, N$$

with the errors ε_j. The estimated coefficients $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_n$ are obtained as the solution of the minimization problem

$$L(\beta_0, \beta_1, \ldots, \beta_n) = \sum_{j=1}^{N} \varepsilon_j^2 \to \min .$$

The values predicted by the model and the residuals are, respectively,

$$\hat{y}_j = \hat{\beta}_0 + \hat{\beta}_1 x_{j1} + \ldots + \hat{\beta}_n x_{jn}, \qquad e_j = y_j - \hat{y}_j,$$

$j = 1, \ldots, N$. If there is no linear relation between input and output, the best prediction is the mean value $\overline{y}$, in which case the residuals coincide with the measured data y_j, centered at the mean. The other extreme is that the data points y_j already lie on a hyperplane $y_j = \beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \ldots + \beta_n x_{jn}$, in which case the best prediction is simply the data point, $\hat{y}_j = y_j$, and the residuals are identically equal to zero.

It can be shown that the total square variability of y can be partitioned into two summands:

$$\sum_{j=1}^{N} (y_j - \overline{y})^2 = \sum_{j=1}^{N} (\hat{y}_j - \overline{y})^2 + \sum_{j=1}^{N} (y_j - \hat{y}_j)^2.$$

The coefficient of determination is defined as $R^2 = \sum_{j=1}^{N} (\hat{y}_j - \overline{y})^2 \big/ \sum_{j=1}^{N} (y_j - \overline{y})^2$. By what has been said above about the residuals, it equals 1 if the data points already lie on a hyperplane and 0 when no linear relationship between input and output exists, and in general measures the explanatory power of the fitted regression model.

The regression coefficients as such cannot be used as indicators of the influence of the corresponding variables, because they are scale dependent. Rather, the *standardized regression coefficients* (SRC) can be used. These are the regression coefficients of the centered and normalized model, where the data x_{ji} are replaced by $(x_{ji} - \overline{x}_i) \big/ \sqrt{\sum_{j=1}^{N} (x_{ji} - \overline{x}_i)^2}$, and similarly for the y_j. In case the n input columns $x_{j1}, \ldots, x_{jn}$, $j = 1, \ldots, N$ are uncorrelated, the SRCs coincide with the CCs. In general, the expression for the CCs has an additional summand which depends on the correlated co-variates.

A more effective removal of the influence of the co-variates is achieved through the *partial correlation coefficients* (PCCs). The partial correlation between x_i and y, given the set of co-variates $x_{\sim i} = \{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n\}$ is defined as the correlation between the two residuals obtained by regressing x_i on $x_{\sim i}$ and y on $x_{\sim i}$, respectively. That is, one first constructs the two regression models

$$\hat{x}_i = \hat{\alpha}_0 + \sum_{k \neq i} \hat{\alpha}_k x_k, \qquad \hat{y} = \hat{\beta}_0 + \sum_{k \neq i} \hat{\beta}_k x_k,$$

obtaining the residuals $\mathbf{e}_i$ and $\mathbf{e}$ with components

$$e_{ji} = x_{ji} - \hat{x}_{ji}, \qquad e_j = y_j - \hat{y}_j,$$

$j = 1, \ldots, N$. By construction, the residuals e_{ji} and e_j are those parts of x_i and y that remain after subtraction of the predicted linear part depending on $x_{\sim i}$. Thus the PCC $\rho(\mathbf{e}_i, \mathbf{e})$ quantifies the linear relationship between x_i and y after removal of any part of the variation due to the linear influence of the co-variates $x_{\sim i}$.

The advantage of the PCCs is that they are more discriminating. In fact, if the input–output map is a truly linear function and the input parameters are uncorrelated, then the PCC of an input variable that enters with a non-zero coefficient is equal to plus or minus one. In reality, input–output maps are not ideally linear functions and so the effect is somewhat moderated. Still the PCCs are an accentuating measure of influence.

If the input–output function is decidedly nonlinear, but monotonic, sensitivities are better detected when one applies a rank transformation to the data. That is, the data $x_{j1}, \ldots, x_{jn}$, and y_j, $j = 1, \ldots, N$ are ordered and only their rank information is kept. This leads to the Spearman rank correlation coefficients (RCC), the standardized rank regression coefficients (SRRC), and the partial rank correlation coefficients (PRCC). Having the computed Monte Carlo sample $x_{j1}, \ldots, x_{jn}$, and y_j at hand, the calculation of the various coefficients produces no additional cost, thus it is recommended to evaluate all six of them to have a better overview. Finally, bootstrap resampling allows one to compute confidence intervals. If zero is outside the confidence interval, the corresponding coefficient can be considered to be significantly different from zero, and the corresponding input variable is classified as having a non-negligible influence on the output. The degree of influence can then be classified according to the magnitude of the six coefficients.

Alternative methods of sensitivity analysis are variance based. Pinching strategies consist in freezing individual variables at their central value and studying the change of variability in the output. If one produces a sample with X_i fixed to its nominal value μ_i, the reduction in variance says something about the influence of X_i on Y:

$$\frac{\mathrm{V}(g(X_1, \ldots, X_{i-1}, \mu_i, X_{i+1}, \ldots, X_n))}{\mathrm{V}(g(X_1, \ldots, X_n))}$$

In fact, this is the proportion of total variance *not* explained by X_i. As can already be seen from the formulation, this strategy is costly because for each pinched variable a new Monte Carlo simulation is required.

A more sophisticated method is the partition of variance according to the groups of variables, the Sobol' indices introduced in [51,52]. It is based on an expansion of the input–output function into summands of increasing dimensionality.

For a survey of sampling based methods in sensitivity analysis, see [20].

4.5 Reliability Analysis

The central concept of reliability analysis is the *failure probability*. The system is considered in a failed state if a certain combination of the input parameters $(X_1, \ldots, X_n)$ and the output Y exceeds an admissible range. For the present purpose it is not necessary to distinguish into favorable (resistance increasing) and unfavorable (load or stress exerting) influences, as is done in the European civil engineering codes [15] with their partial safety factors, critical values and design values. Since Y is a function of the inputs $(X_1, \ldots, X_n)$, which subsume all random influences on the structure, failure can be described by a *limit state function* $\Phi(X_1, \ldots, X_n)$ of the input alone. Failure is usually defined by $\Phi(X_1, \ldots, X_n) < 0$, while $\Phi(X_1, \ldots, X_n) \geq 0$ signifies a safe state. The *failure region* is the subset of design space (the domain of the input parameters) resulting in violation of the limit state condition, i.e.,

$$\mathscr{F} = \{(x_1, \ldots, x_n) : \Phi(x_1, \ldots, x_n) < 0\}.$$

Then

$$p_f = P\big((X_1, \ldots, X_n) \in \mathscr{F}\big) = P\big(\Phi(X_1, \ldots, X_n) < 0\big)$$

is the failure probability; $R = 1 - p_f$ is the *reliability* of the structure. Occasionally it is useful to describe failure as a ratio of actual and admissible values, leading to a failure region of the form $\Psi(x_1, \ldots, x_n) > 1$. To determine the probability p_f, the types and parameters of the probability distributions of $(X_1, \ldots, X_n)$ are needed. As opposed to sensitivity analysis, this requires detailed information about the statistical properties of the input parameters, obtained from experiments or previous studies.

The acceptable value of the failure probability depends on the circumstances. The civil engineering codes require that the designed structure obtains an instantaneous probability of failure of $p_f = 10^{-6}$ and a long-term failure probability of $p_f = 10^{-5}$. To credibly estimate tail probabilities of such a small magnitude, *a lot* of information is needed. In addition, if time dependent reliability is to be assessed, failure rates and the additional parameters of the time-dependent reliability function are required. The problems arising from this concept of failure probability have

been discussed at many places, including the codes themselves [15, Annex C4(3)], see also [16, 36] and references therein. In technological development phases in aerospace engineering, a failure probability in the range of $p_f = 10^{-3}$ may be acceptable, especially if it is not used as an absolute measure, but as an objective function in optimization (reliability based optimization).

Having performed a Monte Carlo sensitivity analysis of the model output $Y = g(X_1, \ldots, X_n)$, the question comes to mind if one could not use the generated sample for getting reliability estimates of the structure. This is indeed the case, albeit at a possibly low accuracy due to the small sample size of the sensitivity analysis. There are two ways of exploiting the existing sample. One way is by means of tolerance intervals to estimate credible upper and lower bounds for the output Y; another way is by reweighting the generated sample of $(X_1, \ldots, X_n)$ so as to mimic input distributions other than the uniform distributions used in the sensitivity analysis.

Tolerance Intervals While confidence intervals give an estimate for the distribution parameters λ of a random quantity Y, a tolerance interval gives an estimate of the range of possible observations of Y. More precisely, one wants to compute an interval $[a, b]$ that contains a certain proportion p, say $p = 90\,\%$, of the population with a given confidence level $1 - \alpha$, say $1 - \alpha = 95\,\%$. A non-parametric approach based on order statistics is especially attractive, since it is applicable without knowledge of the type of statistical distribution of Y. In fact, given whatever sample of whatever random variable, one may estimate the proportion p of the population that lies within the sample maximum $Y_{\max}$ (the largest value in the sample) and the sample minimum $Y_{\min}$ with a given confidence $1 - \alpha$, depending only on the sample size N. In this situation, the interval $[a, b] = [Y_{\min}, Y_{\max}]$ is given and N is known. Thus depending on the desired confidence level, the proportion p lying within the boundaries $[a, b]$ can be computed.

The derivation of a one-sided non-parametric tolerance interval with upper boundary the sample maximum is particularly easy, using only combinatorics. In fact, a proportion p of the population lies in the interval $(-\infty, Y_{\max}]$ with confidence $1 - \alpha$ if the relation

$$p^N = \alpha$$

holds. This can be seen as follows. Denote by Q_p the p-th quantile of Y. This means that $P(Y \leq Q_p) = p$. On the other hand, the interval $(-\infty, Y_{\max}]$ contains at least the proportion p of the population if $Q_p \leq Y_{\max}$. Thus it is required that

$$P(Q_p \leq Y_{\max}) \geq 1 - \alpha.$$

But $P(Q_p \leq Y_{\max}) = 1 - P(Y_{\max} < Q_p)$. Observe that $Y_{\max} < Q_p$ if and only if each of the N independent realizations of Y in the sample is below Q_p, i.e., $Y_j < Q_p$ for $j = 1, \ldots, N$. By definition, the probability of the event $Y_j < Q_p$ is exactly p. Collecting terms, one arrives at $1 - P(Y_{\max} < Q_p) = 1 - p^N$, whence the assertion.

Table 4.1 Required sample size N for one-sided tolerance intervals at confidence $1-\alpha$

$1-\alpha$	$p=0.90$	$p=0.95$	$p=0.99$
0.90	22	45	230
0.95	29	59	299
0.99	44	90	459

From there, universally valid estimates of the sample size N required so that a proportion p of values lies under the sample maximum at confidence $1-\alpha$ can be established, see e.g. Table 4.1. The same formula applies to one-sided intervals of the form $[Y_{\min},\infty)$. Tolerance intervals with various proportions and confidence levels are tabulated, e.g., in the ISO standard [22]. For the theory, see, e.g., [27].

Monte Carlo Reweighting As outlined in Sect. 4.2, the goal of Monte Carlo simulation is to estimate expectation values $\mathrm{E}\big(h(Y)\big) = \mathrm{E}\big(h(g(X_1,\dots,X_n))\big)$ of functions of the model output. Suppose we have already generated a sample $(x_{ji},\dots,x_{jn})$ and computed the outputs y_j, $j=1,\dots,N$, where the sample has been generated according to a certain probability distribution of the input, say with probability density $f(x_1,\dots,x_n)$. Is it possible to use the same sample to estimate $\mathrm{E}\big(h(\tilde{Y})\big) = \mathrm{E}\big(h(g(\tilde{X}_1,\dots,\tilde{X}_n))\big)$ where the $\tilde{X}_j$ are random variables defined on the same range as the X_j, but with another probability distribution, say with probability density $\varphi(x_1,\dots,x_n)$? To understand the positive answer, it is useful to write the expectation $\mathrm{E}\big(h(\tilde{Y})\big)$ as an integral:

$$\begin{aligned}
\mathrm{E}\big(h(g(\tilde{X}_1,\dots,\tilde{X}_n))\big) &= \int\cdots\int h(g(x_1,\dots,x_n))\varphi(x_1,\dots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_n \\
&= \int\cdots\int h(g(x_1,\dots,x_n))\frac{\varphi(x_1,\dots,x_n)}{f(x_1,\dots,x_n)} \\
&\quad\times f(x_1,\dots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_n \\
&= \mathrm{E}\Big(h(g(X_1,\dots,X_n))\frac{\varphi(X_1,\dots,X_n)}{f(X_1,\dots,X_n)}\Big).
\end{aligned}$$

This shows that the computation of the new expectation value can be accomplished by computing the old expectation value of the input–output function, multiplied by a weight—the quotient of the two densities. In terms of Monte Carlo simulation, one has to compute

$$\mathrm{E}\big(h(\tilde{Y})\big) \approx \frac{1}{N}\sum_{j=1}^{N} h\big(g(x_{ji},\dots,x_{jn})\big)\,\frac{\varphi(x_{ji},\dots,x_{jn})}{f(x_{ji},\dots,x_{jn})}.$$

This causes no additional effort, because one can reuse the expensive computation of $h\big(g(x_{ji},\dots,x_{jn})\big)$. The accuracy of the method depends on the degree of similarity of the old and the new distribution.

A typical application could consist in reusing the sample from the sensitivity analysis, based on uniform distributions, and place truncated Gaussians on the intervals. This concludes the remarks about what can be extracted from the sensitivity analysis towards a reliability assessment.

Importance Sampling When estimating failure probabilities of low value, one cannot expect that the rather small sample sizes of sensitivity analysis suffice. In fact, a standard Monte Carlo estimate of the failure probability is of the form

$$p_f \approx \frac{1}{N} \sum_{j=1}^{N} \chi^{\mathscr{F}}(x_{j1}, \ldots, x_{jn})$$

where $\chi^{\mathscr{F}}$ equals one if $(x_{j1}, \ldots, x_{jn}) \in \mathscr{F}$ and zero otherwise. It thus can be seen as the mean value of an N-fold repetition of a zero-one experiment with success probability p_f. The variance of the Monte Carlo estimator of p_f is hence given by $p_f(1 - p_f)/N$, whence the mean estimation error is approximately equal to $\sqrt{p_f/N}$. This means that an accuracy of $\alpha \cdot 100\,\%$ requires a sample size of $N \approx 1/(\alpha p_f)$. Consequently, cost-saving methods need to be devised, two of which shall be discussed here.

The first one is *importance sampling*. As seen above, the probability of failure is estimated by counting the number of realizations of the input variables $(X_1, \ldots, X_n)$ that fall into the failure region. Since this number is expected to be small compared to the total number of simulated points, the probability density $f(x_1, \ldots, x_n)$ of the input variables will be small on $\mathscr{F}$. The idea of importance sampling is to generate a sample of another distribution, say with probability density $\varphi(x_1, \ldots, x_n)$, which may be concentrated in the region $\mathscr{F}$. The idea is similar to Monte Carlo reweighting, but this time a different sample is produced to begin with. In fact,

$$\begin{aligned} p_f &= \mathrm{E}\big(\chi^{\mathscr{F}}(X_1, \ldots, X_n)\big) = \int \cdots \int \chi^{\mathscr{F}}(x_1, \ldots, x_n) f(x_1, \ldots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n \\ &= \int \cdots \int \chi^{\mathscr{F}}(x_1, \ldots, x_n) \frac{f(x_1, \ldots, x_n)}{\varphi(x_1, \ldots, x_n)} \varphi(x_1, \ldots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n \\ &= \mathrm{E}\Big(\chi^{\mathscr{F}}(\tilde{X}_1, \ldots, \tilde{X}_n) \frac{f(\tilde{X}_1, \ldots, \tilde{X}_n)}{\varphi(\tilde{X}_1, \ldots, \tilde{X}_n)}\Big) \end{aligned}$$

where the random variables $(\tilde{X}_1, \ldots, \tilde{X}_n)$ have the probability density $\varphi(x_1, \ldots, x_n)$. This leads to the following prescription. First, choose a probability density function $\varphi(x_1, \ldots, x_n)$ concentrated in the failure region. Next, generate a random sample $(z_{ji}, \ldots, z_{jn})$ according to the corresponding probability distribution. Finally, estimate the failure probability by

$$p_f \approx \frac{1}{N} \sum_{j=1}^{N} \chi^{\mathscr{F}}(z_{ji}, \ldots, z_{jn}) \frac{f(z_{ji}, \ldots, z_{jn})}{\varphi(z_{ji}, \ldots, z_{jn})}.$$

Of course, this begs the question how to find a probability density $\varphi(x_1, \ldots, x_n)$ concentrated in the failure region. After all, the failure region unfolds itself only *after* Monte Carlo evaluation of the limit state function. Various proposals have been made in this respect, notably Bucher's adaptive sampling [7, 9]. This method starts with a pilot simulation with increased variance of the input variables (to rapidly produce a number of points in the failure region) and then uses certain shifted Gaussians as weight functions.

An interesting proposal has recently been made by Schwarz [49], developed for the situation where the input variables are supported in intervals (as, e.g., the uniform distributions used in sensitivity analysis or transformations thereof). There are two ingredients. First, one may expect that—in most cases—the failure region is concentrated in points near the boundaries of the intervals. Second, input variables with a larger influence on the output should have a larger weight. The starting point of the procedure is a sensitivity analysis with moderate sample size, from which the most important input variables and their correlation coefficients with the output are determined. Then a parametrized family of weight functions is placed on the input intervals, where the weight is just 1 for the unimportant parameters and is shifted more and more towards the boundaries of the input intervals, the larger the correlation.

Subset Simulation The subset simulation method was introduced by Au and Beck in [3]. The idea is to approximate the failure region $\mathscr{F}$ by a sequence of larger regions $\mathscr{F} = \mathscr{F}_m \subset \mathscr{F}_{m-1} \subset \ldots \subset \mathscr{F}_1 \subset \mathscr{F}_0$ and to compute the failure probability by a product of conditional probabilities

$$p_f = P(\mathscr{F}) = P(\mathscr{F}_m|\mathscr{F}_{m-1})P(\mathscr{F}_{m-1}|\mathscr{F}_{m-2}) \ldots P(\mathscr{F}_1|\mathscr{F}_0)P(\mathscr{F}_0)$$

where $\mathscr{F} = \mathscr{F}_m$ and $\mathscr{F}_0$ is the starting region. These conditional probabilities are appreciably larger than the failure probability and hence easier to simulate with smaller samples. In this case it is useful to describe the failure region by means of a ratio based limit state function $\mathscr{F} = \{(x_1, \ldots, x_n) : \Psi(x_1, \ldots, x_n) > 1\}$. The intermediate regions are chosen of the form $\mathscr{F}_i = \{(x_1, \ldots, x_n) : \Psi(x_1, \ldots, x_n) > \alpha_i\}$ with $0 < \alpha_0 < \alpha_1 < \ldots < \alpha_m = 1$. In fact, the choice of α_i is often made during the simulation such that $P(\mathscr{F}_i \,|\, \mathscr{F}_{i-1})$ has a fixed value, say p_0 between 0.1 and 0.3, and the regions $\mathscr{F}_i$ are constructed recursively. The conditional distribution $P(\cdot|\mathscr{F}_i)$ is just the original distribution, restricted to $\mathscr{F}_i$ and scaled by $P(\mathscr{F}_i)$. The latter probability is unknown a priori. This suggests to use the Metropolis–Hastings algorithm, a Markov chain Monte Carlo algorithm, which requires knowledge of the sampling distribution only up to a multiplicative factor (see below).

In the sequel, it is assumed that at each level i, a sample of size N is generated. The algorithm is initiated by generating a sample of the original distribution using standard Monte Carlo simulation. From this sample, the worst $p_0 \cdot 100\,\%$ realizations are declared to belong to $\mathscr{F}_0$. More precisely, the sample is ordered according to the Ψ-values. The threshold level α_0 is chosen as the $(1 - p_0)N$-th largest value among the Ψ-values attained in the sample, and $\mathscr{F}_0$ is defined to be the set

$\{(x_1, \ldots, x_n) : \Psi(x_1, \ldots, x_n) > \alpha_0\}$. Further, $P(\mathscr{F}_0) = p_0$. In the next step, one or more of the points in $\mathscr{F}_0$ is/are chosen as the initial point (root) of a Markov chain whose elements are distributed according to $P(\cdot|\mathscr{F}_0)$, this way generating a second sample of size N. Again, the worst $p_0 \cdot 100\,\%$ realizations are declared to belong to $\mathscr{F}_1$, and α_1 is chosen as the $(1 - p_0)N$-th largest value among the Ψ-values attained in the second sample, and so on. The simulation stops at the first level m at which the Ψ-values of the worst $p_0 \cdot 100\,\%$ bigger than 1. At this stage $P(\mathscr{F}_m|\mathscr{F}_{m-1})$ is estimated by M/N where M is the number of failed realizations in the last sample. Finally, p_f is estimated as $p_0^m \cdot M/N$.

Before discussing some details of the algorithm, a short introduction to Markov chain Monte Carlo simulation is in order. The ideas can be best explicated at the hand of the original Metropolis algorithm for simulating a one-dimensional distribution $\pi(x)$. The goal is to generate a realization $\xi_0, \xi_1, \xi_2, \ldots, \xi_N$ of a Markov chain whose stationary distribution is $\pi(x)$. Since the chain will converge to the stationary distribution as $N \to \infty$, the end-pieces $\xi_M, \ldots, \xi_N$ are approximately distributed according to π (for large M and N).

The algorithm proceeds as follows. Choose a transition kernel $q(x, y)$ *(proposal distribution)* such that $q(x, y) = q(y, x)$ for all x, y. (Often it is taken of the form $q(x, y) = p(x - y)$ where p is a nowhere vanishing probability density.) Choose an initial distribution $p^{(0)}(x)$.

- Sample a value ξ_0 from $p^{(0)}$.
- For $k = 1, \ldots, N$

  Sample a value η from the proposal distribution $q(\xi_{k-1}, \cdot)$.

  Compute the ratio $r = \dfrac{\pi(\eta)}{\pi(\xi_{k-1})}$.

  – If $r \geq 1$, the value η is accepted; set $\xi_k = \xi_{k-1}$.
  – If $r < 1$, the value η is accepted with probability r and rejected with probability $1 - r$.

  Draw a random number ζ from the uniform distribution on $[0, 1]$.

  – If $\zeta \leq r$, set $\xi_k = \eta$.
  – If $\zeta > r$, set $\xi_k = \xi_{k-1}$.

- $\xi_0, \xi_1, \ldots, \xi_N$ has $\pi(x)$ as limiting distribution as $N \to \infty$.

Observe that only the ratio r enters in the computation, so knowledge of $\pi(x)$ is only required up to a multiplicative factor. The Metropolis–Hastings algorithm is similar, but the proposal distribution is not required to be symmetric. More theory can be found in [42].

If the distribution $\pi(x_1, \ldots, x_n)$ is multidimensional, it is advantageous to change only one coordinate in each step in order to keep the number of rejected trials at a moderate level. In its application to subset simulation, the target distribution is $P(\cdot|\mathscr{F}_{i-1})$ in each level. Thus one has to check whether $\xi_k \in \mathscr{F}_{i-1}$ for acceptance, in addition. By construction, each root of the generated Markov chain is already distributed according to $P(\cdot|\mathscr{F}_{i-1})$. It can be shown [3] that the elements

of the whole chain at level i are not only asymptotically but also perfectly distributed according to $P(\cdot|\mathscr{F}_{i-1})$. As can be seen, there are a lot of screws that can be adjusted: choice of the proposal distribution, optimal acceptance/rejection rate, number and length of chains generated in level i—with the possibility of parallelization, choice of p_0, and so on. Based on many recommendations to be found in the literature, subset simulation has developed into an efficient method for estimating failure probabilities. A critical comparison of various simulation methods in reliability analysis can be found in [47, 48].

4.6 Application

This section is devoted to demonstrating the methods at work in a practical application from aerospace engineering, namely in a finite element model of the frontskirt of the ARIANE 5 launcher. The frontskirt is the part of the launcher that connects the tanks section with the payload section and also has to support the booster loads. It consists of a light weight shell structure reinforced by struts. The full finite element model is composed of shell elements and solid elements, altogether with two million degrees of freedom. The models have been supplied by Intales GmbH Engineering Solutions (see Footnote 1). Models of varying complexity and material properties with up to 130 input and a similar number of output parameters have been analyzed.

For the sake of presentation, we shall focus on a smaller finite element model keeping the global structure with about ninety thousand degrees of freedom. Figure 4.1 depicts the model schematically; it is composed of two hemispheres and three cylinders, one of which is made up of composite material. The shadings indicate thickness variations of the tank skin, described by a random field. Booster loads are introduced at two opposite locations in the upper cylinder (not shown in Fig. 4.1). A selection of seventeen input parameters (all loads characterizing various flight scenarios) will be considered; their meaning is described in Table 4.2. As a representative output we start with the load proportionality factor (LPF), a decisive variable indicating buckling failure. It is defined as the limiting value in an incremental procedure in which the mechanical loads during a flight scenario are increased step by step until breakdown of the structure is reached. In the full model, the LPF is computed by means of a path following procedure that follows bifurcations until material failure occurs. In the simplified computations presented here, no distinction of bifurcation or material failure was made, so that the terminal value of the LPF was taken as that value at which the finite element program failed to converge. What concerns the computational effort, a single run of the input–output function computing the LPF in Abaqus takes around 1 h on a personal computer and 10 min on the supercomputer Leo-III of the University of Innsbruck.

Sensitivity Analysis As a basis for the sensitivity analysis, a sample of size $N = 100$ of the $n = 17$ input parameters was generated. Each parameter was taken

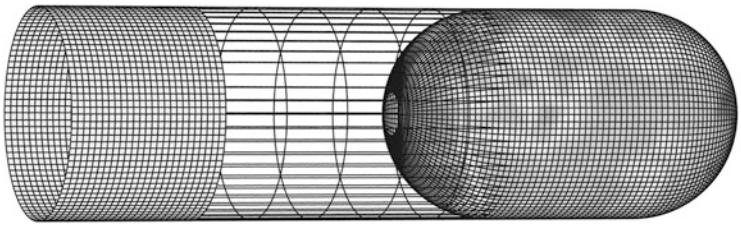

Fig. 4.1 Simplified finite element model of frontskirt; *shadings* show random field (thickness of tank skin) [12]

Table 4.2 Description of input parameters no. 1–17 with their nominal values

i	Parameter X_i	Mean μ_i
1	Initial temperature	293 K
2	Step1 thermal loading cylinder1	450 K
3	Step1 thermal loading cylinder2	350 K
4	Step1 thermal loading cylinder3	150 K
5	Step1 thermal loading sphere1	150 K
6	Step1 thermal loading sphere2	110 K
7	Step2 hydrostatic pressure cylinder3	0.4 MPa
8	Step2 hydrostatic pressure sphere1	0.4 MPa
9	Step2 hydrostatic pressure sphere2	0.4 MPa
10	Step3 aerodynamic pressure	−0.05 MPa
11	Step4 booster loads y-direction node1	40,000 N
12	Step4 booster loads y-direction node2	20,000 N
13	Step4 booster loads z-direction node1	$3.e6$ N
14	Step4 booster loads z-direction node2	$1.e6$ N
15	Step4 mechanical loads x-direction	100 N
16	Step4 mechanical loads y-direction	50 N
17	Step4 mechanical loads z-direction	300 N

Table 4.3 Sample statistics of simulated load proportionality factor

Mean	Minimum	Maximum	Standard deviation	Spread
3.5335	3.4468	3.6457	0.1989	±3 %

uniformly distributed around its nominal value listed in Table 4.2 with a spread of ±15 %. Latin hypercube sampling and correlation control was employed. The statistics of the computed LPF are listed in Table 4.3.

The scatterplot of Fig. 4.2 gives a first impression of the influence of the input variables on the output. It is quite clearly seen that input parameter no. 13 (booster loads) exerts a big influence, whereas the diagrams for the other parameters are less conclusive. To quantify the influences, the six different correlation indices CC, PCC, SRC, RCC, PRCC, SPRC were computed. (A complete list of the values has been published in [38].) As an example, the PRCCs of the 17 input parameters vs. the LPF are visualized in Fig. 4.3 (left).

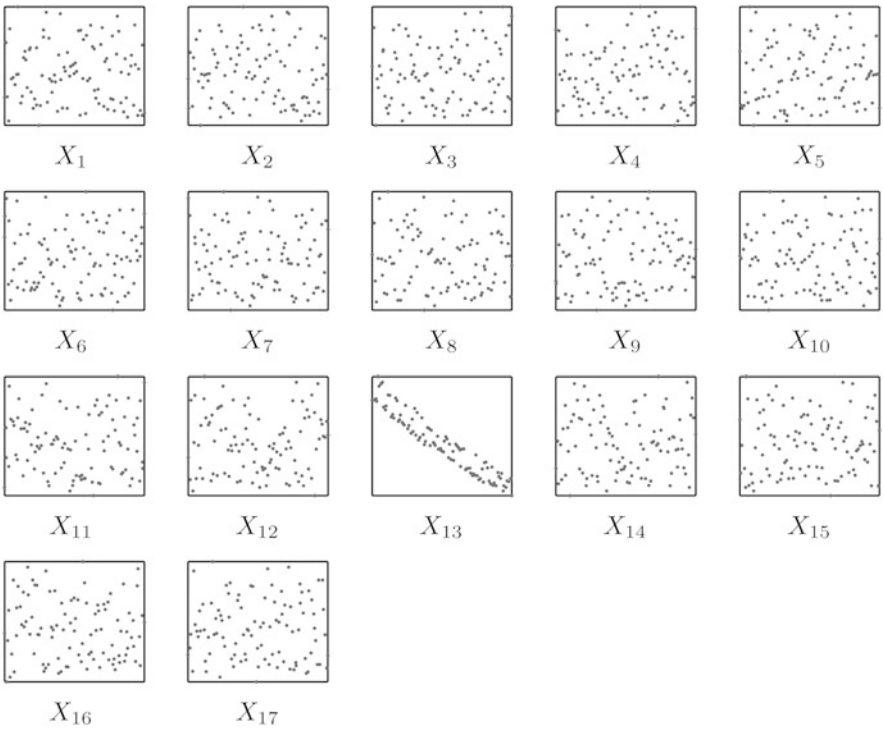

Fig. 4.2 *Scatterplots* of 17 input variables vs. output (LPF) [24, 38]

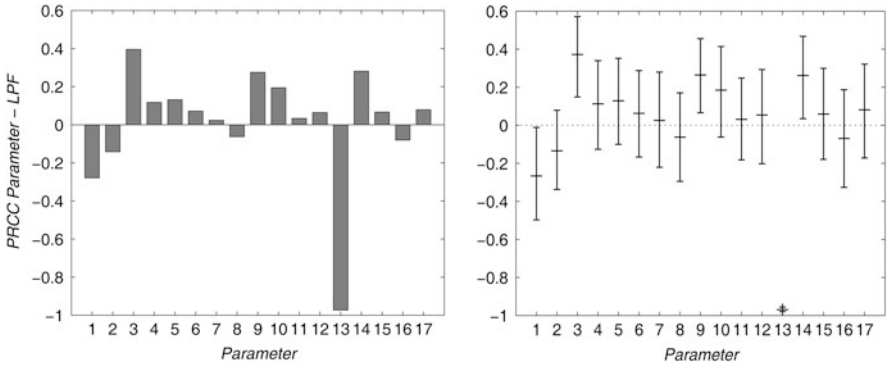

Fig. 4.3 Partial rank correlation coefficients (*left*) and 95 % bootstrap confidence intervals (*right*) [24, 38]

The resulting sensitivity indices induce a ranking of the input parameters according to their influence on the output. Note that the accuracy of the estimate for the correlation coefficients is in the range of $1/\sqrt{N} \cdot 100\,\% = 10\,\%$; this suffices to determine the ranking and confirms the observation that one can get along

Table 4.4 Ranks of the significant input parameters X_i according to the six measures of correlation input–output

i	CC	PCC	SRC	RCC	PRCC	SRCC
1		3	3		4	4
3		2	2		2	2
9					5	5
13	1	1	1	1	1	1
14		4	4		3	3

with small sample sizes for an assertive sensitivity analysis. To check whether the computed correlation indices are significantly different from zero, bootstrap 95 %-confidence intervals were computed (with bootstrap sample size $B = 5{,}000$). As a basis for an overall assessment of the ranking, only those sensitivity estimates with a resultant confidence interval not including 0 have been regarded as significant. As an example, bootstrap confidence intervals for the PRCCs are displayed in Fig. 4.3 (right). Accordingly, only the PRCCs of the parameters X_1, X_3, X_9, X_{13}, and X_{14} test to be nonzero. The ranks of those parameters that tested to be significant according to at least one of the six indices are listed in Table 4.4. The table gives a good impression of the sensitivity assessment—if a single scale is required, one might use the average ranks.

Coefficient of Determination As discussed in Sects. 4.2 and 4.4, metamodels can be used to further quantify the influence of selected parameters. One way to achieve this is to fit a linear regression model

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n$$

and then to compute the partial coefficients of determination. The sequential partial coefficient of determination of variable X_i is computed by first fitting the model $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_{i-1}$, then adjoining the variable X_i and recording the increase in the coefficient of determination R^2. Averaging all partial coefficients of determination which can be obtained by adding the variable X_i to all possible combinations of the already included variables leads to a non-parametric measure of the contribution of variable X_i to the *explanatory power* of the model. The procedure is explained, e.g., in [28, 39]. The result can be conveniently displayed in the form of a pie chart. As an example, a linear model for the *LPF* has been set up with input parameters nos. 1, 3, 4, 7, 9, 13, 14. It resulted in the assessment of the influences depicted in Fig. 4.4, taken from [43]. The eminent influence of parameter X_{13} is once more confirmed. In the figure, the label *Res* refers to the residual proportion which remains unexplained through the linear model.

Random Fields In order to investigate the effect of geometric and material imperfections, the thickness, modulus of elasticity and yield stress were disturbed by two-dimensional random fields with an autocovariance function of the form (4.1). The distance function on the cylinders was taken as the sum of the axial and radial distance, whereas on the spheres, it was taken as the sum of the latitude and altitude, multiplied by the radius. A routine extracting the distances from the finite

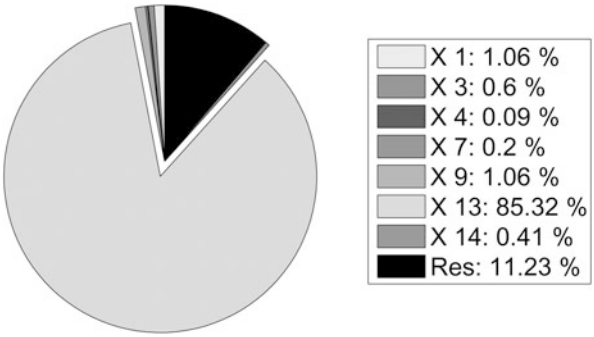

Fig. 4.4 Pie chart of relative explanatory power of LPF through various input variables [43]

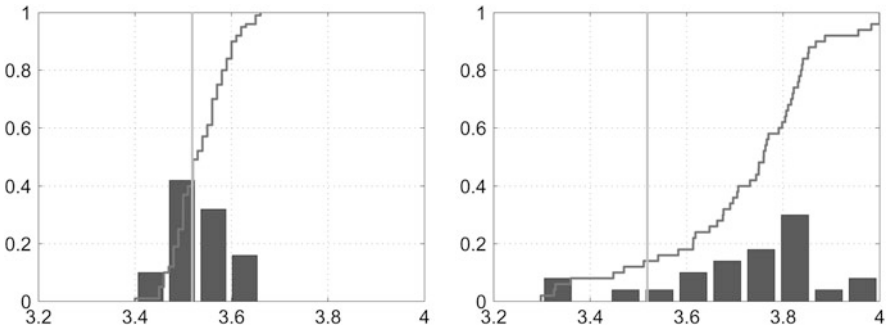

Fig. 4.5 Plots of histogram and cumulative distribution function of LPF produced by random loads, no imperfections (*left*) and with random field turned on (*right*). *Vertical line* indicates LPF corresponding to nominal input values [41]

element grid was implemented by [41]. The nominal values were 1 mm (thickness), 70,000 MPa (modulus of elasticity), 320 MPa (yield stress) for the first sphere, with similar values for the other components of the frontskirt. A coefficient of variation of 10 % was applied throughout. In a parametric study, the correlation lengths were varied between 60 mm (corresponding to the dimension of two elements of the grid) and 1,600 mm, with various combinations in the two respective directions. Random fields were generated by means of the Karhunen–Loève expansion.

In a first investigation, the different effects of the random imperfections and the 17 random loads on the LPF were studied. Figure 4.5 shows one of the results, with a sample size of $N = 100$ both for loads and realizations of the random fields. The correlation lengths were set to 188.5 mm in all angular directions (cylinders and spheres), while the correlation length in axial direction was taken 450 and 900 mm for the cylinders. In the left figure, the distribution of the LPF is shown with the random loads from the sensitivity analysis, but with material and geometric properties kept at their nominal value. In the right figure, the random field is turned on, in addition. One can see that the random field has a stabilizing effect (larger

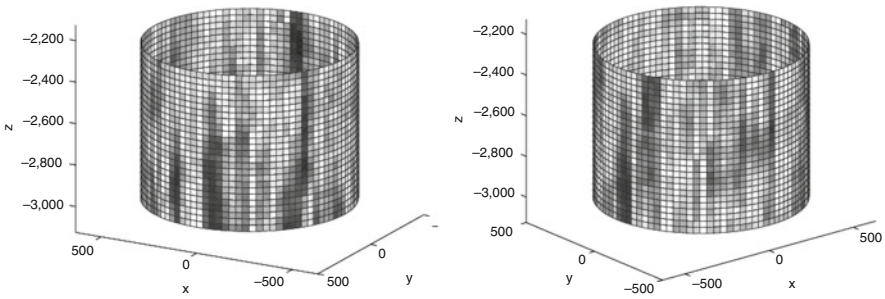

Fig. 4.6 Spatial distribution of influence (measured by CC) of element-wise yield stress (*left*) and thickness (*right*) on the LPF, based on random field simulation [41]

Table 4.5 Proportion p of LPFs lying above $LPF_{\min} = 3.4468$ at confidence level $1 - \alpha$

$1 - \alpha$	90 %	95 %	99 %	99.9 %
p	97.7 %	97.1 %	95.5 %	93.3 %

values of LPF are attained), but at the same time increases the uncertainty of the outcome.

Interestingly, the application of random fields admits a structurally localized study of the correlations. After all, a realization of the random field induces a realization of the modelled quantity (e.g., thickness) in each element of the FE-grid. Thus one can do a standard sensitivity analysis with these variables. For example, cylinder no. 3 has 2,500 elements; the random field simulation produces $N = 100$ realizations of the 2,500 grid values of the thickness, elasticity modulus, yield stress. The spatial distribution of the influence on the LPF can thus be visualized. Figure 4.6 shows such a distribution in terms of the Pearson CC for the yield stress (left) and the thickness (right). In the same way, localized correlations of different output variables can be pictured.

Tolerance Intervals As a first step from sensitivity to reliability, tolerance intervals for the LPF can be established. Recall from Table 4.3 that the minimal sampled LPF was at $LPF_{\min} = 3.4468$. Based on the formula $p^N = \alpha$ with $N = 100$, one can assess the proportions p of the possibly attainable LPF-values with a given confidence $1 - \alpha$. The results are summarized in Table 4.5.

Reliability Analysis In order to test various simulation methods for reliability, a benchmark study of the small launcher model was undertaken by [49], from where all results and tables in this paragraph are taken. For this study, an extended list of 35 input parameters was used. As a limit state function, a combination of allowable limits in the equivalent plastic strain (PEEQ), principal stress in the composite part (SP), and the absolute value of the smallest eigenvalue (EV) was employed:

$$\Psi(\mathbf{x}) = \max\left\{\frac{PEEQ(\mathbf{x})}{0.07}, \frac{SP(\mathbf{x})}{180}, \frac{0.001}{EV(\mathbf{x})}\right\}$$

with failure defined by $\Psi(\mathbf{x}) > 1$, $\mathbf{x} = (x_1, \ldots, x_{35})$. The three criteria correspond to plastification in the metallic part, rupture in the composite part, and buckling of the structure.

A reference brute force Monte Carlo simulation of size $N = 5{,}000$ was undertaken on the supercomputer Leo-III, with three strands in parallel. As in the previous sensitivity analysis, the 35 input variables were taken uniformly distributed on an interval of spread $\pm 15\,\%$ around their nominal values. The resulting failure probability turned out to be $p_f = 0.0116$. A 95 % bootstrap confidence interval was computed (bootstrap sample size $B = 5{,}000$) as $[0.0088, 0.0146]$. A sensitivity analysis with the sample revealed that only 10 of the 35 parameters had a significant influence on the failure criterion Ψ, measured at a 90 % confidence level.

Next, an investigation was undertaken whether an estimate in the same range could be obtained with a smaller sample size by subset simulation or by importance sampling. Subset simulation was undertaken with $p_0 = 0.2$ (see Sect. 4.5). Since $p_0^3 = 0.008$ is already smaller than the intended failure probability, three levels $\mathscr{F}_0, \mathscr{F}_1, \mathscr{F}_2$ suffice for the subset simulation algorithm. Experiments were undertaken with sample size $N = 900$ and $N = 300$ for each level. Recall that 20 % of the generated points 20 % of the generated points of $\mathscr{F}_{i-1}$ are assigned to $\mathscr{F}_i$ in each step. Since these points can be reused when going from level $i - 1$ to level i, the total number of generated points in the three levels is $2,340$ and 780, respectively. Further, it was tested whether including all 35 input influential ones in the simulation changes the value of the failure probability. In addition, bootstrap confidence intervals for the failure probability were computed.

To keep results comparable, the importance sampling procedure was done with a sample size $N = 780$. The method described in Sect. 4.5 was employed, by which the weights are computed in dependence on the magnitude of the correlation coefficient of the respective input parameter with the Ψ-value. This required the actual simulation to be preceded by a sensitivity analysis. The sensitivity analysis was done with a sample of size 99, so that a sample size of 681 remained for the importance sampling part. Correlation control was employed when simulating the input data. It was tested whether weighting of all parameters or weighting only the parameters significant at the 90 % level changes the outcome.

The joint results are recorded in Table 4.6. Here NR refers to the total number of realizations computed in the simulation, NV denotes the number of activated variables (subset simulation), respectively weighted variables (importance sampling); p_f is the failure probability and 95 % BSL/BSU refers to the lower/upper bound of the 95 % bootstrap confidence interval for p_f.

We conclude this section by reporting on a reweighting experiment. As a basis, the sample of size $N = 5{,}000$ of the reference Monte Carlo simulation was taken. The uniformly distributed input was replaced, using reweighting, by truncated Gaussians. The mean values of the Gaussian distributions were taken as the interval midpoints, the variance was computed from assumed coefficients of variation (between 7.5 and 15 %), the truncation was effected at the interval endpoints. The change from uniform distributions to mid-pieces of Gaussians resulted in quite a

Table 4.6 Comparison of results of six different simulation procedures for the failure probability [49]

	Monte Carlo	Subset	Subset	Subset	Importance	Importance
NR	5,000	2,340	780	780	780	780
NV	35	35	35	10	35	10
p_f	0.0116	0.0130	0.0155	0.0120	0.0108	0.0124
95 % BSL	0.009	0.010	0.010	0.008	0.006	0.007
95 % BSU	0.015	0.016	0.022	0.017	0.019	0.019

change of the failure probability, namely to $p_f = 0.0019$ with a 95 % bootstrap confidence interval $[0.001, 0.003]$.

4.7 Conclusion

The purpose of this chapter was twofold. On the one hand, it served to describe current core methods of Monte Carlo simulation, from design of experiment, random fields, metamodelling to concepts of sensitivity and reliability analysis. On the other hand, the chapter demonstrated the implementation of those methods in joint research projects with Intales GmbH Engineering Solutions over the past years.

A number of themes have deliberately not been addressed in order to keep the presentation concise. These include simulation of correlated input using copulas [33], also implemented in the mentioned projects [44], Bayesian methods of reliability analysis [31], Bayesian estimates of the distribution of the failure probability [49, 54], and optimization for finding worst case parameter combinations [12]. Further, the discussion of asymptotic sampling [8], though implemented in our toolbox [49], was omitted because its presentation would have required to go into some details about the safety index and FORM (the first order reliability method).

Acknowledgements The development, adaptation, and implementation in the mentioned research projects is chiefly due to the essential contributions of Christoph Aichinger, Vincent De Groof, Julian King, Katharina Riedinger, Helene Roth, and Martin Schwarz [1, 10, 11, 24, 41, 44, 49]. Many ideas have been developed in discussions with Barbara Goller and Herbert Haller of Intales GmbH, whose continuous support I gratefully acknowledge.

References

1. Aichinger, C.: Monte Carlo methods in iterative solvers. Diploma thesis, University of Innsbruck, Austria (2010)
2. Arnold, L.: Stochastic Differential Equations: Theory and Applications. Wiley, New York (1974)
3. Au, S.-K., Beck, J.L.: Estimation of small failure probabilities in high dimensions by subset simulation. Probab. Eng. Mech. **16**, 263–277 (2001)

4. Beer, M., Ferson, S., Kreinovich, V.: Imprecise probabilities in engineering analysis. Mech. Syst. Signal Process. **37**, 4–29 (2013)
5. Bhatia, R.: Perturbation Bounds for Matrix Eigenvalues. Longman, Harlow (1987)
6. Bolotin, V.V.: Statistical Methods in Structural Mechanics. Holden-Day, San Francisco (1969)
7. Bucher, C.: Adaptive sampling: an iterative fast Monte Carlo procedure. Struct. Saf. **5**, 119–126 (1988)
8. Bucher, C.: Asymptotic sampling for high-dimensional reliability analysis. Probab. Eng. Mech. **24**, 504–510 (2009)
9. Bucher, C.: Computational Analysis of Randomness in Structural Mechanics. CRC Press/Balkema, Leiden (2009)
10. De Groof, V., Oberguggenberger, M., Haller, H., Degenhardt, R., Kling, A.: Quantitative assessment of random field models in finite element buckling analyses of composite cylinders. In: Eberhardsteiner, J., Böhm, H.J., Rammerstorfer, F.G. (eds.) CD-ROM Proceedings of the 6th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2012), Vienna University of Technology, Wien (2012)
11. De Groof, V., Oberguggenberger, M., Haller, H., Degenhardt, R., Kling, A.: A case study of random field models applied to thin-walled composite cylinders in finite element analysis. In: Deodatis, G., Ellingwood, B.R., Frangopol, D.M. (eds.) Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures, p. 379. CRC Press/Balkema, Leiden (2013)
12. De Groof, V., Oberguggenberger, M., Prackwieser, M., Schwarz, M.: Reliability analysis of shell structures. In: Barden, M., Ostermann, A. (eds.) Scientific Computing @ uibk, pp. 39–42. Innsbruck University Press, Innsbruck (2013)
13. Dick, J., Pillichshammer, F.: Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration. Cambridge University Press, Cambridge (2010)
14. Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap. Chapman and Hall, New York (1993)
15. European Committee for Standardization: EN 1990:2002. Eurocode: Basis of Structural Design. CEN, Brussels (2002)
16. Fellin, W., Lessmann, H., Oberguggenberger, M., Vieider, R.: Analyzing Uncertainty in Civil Engineering. Springer, Berlin (2005)
17. Freudenthal, A.N.: Safety and the probability of structural failure. Trans. ASCE **121**, 1337–1397 (1956)
18. Ghanem, R.G., Spanos, P.D.: Stochastic Finite Elements: a Spectral Approach. Springer, New York (1991)
19. Graham, C., Talay, D.: Stochastic Simulation and Monte Carlo Methods: Mathematical Foundations Of Stochastic Simulations. Springer, Berlin (2013)
20. Helton, J.C., Johnson, J.D., Sallaberry, C.J., Storlie, C.B.: Survey of sampling-based methods for uncertainty and sensitivity analysis. Reliab. Eng. Syst. Saf. **91**, 1175–1209 (2006)
21. Iman, R.L., Conover, W.J.: A distribution-free approach to inducing rank correlation among input variables. Commun. Stat. Simul. Comput. **11**, 311–334 (1982)
22. International Standard: ISO 16269-6:2005. Statistical Interpretation of Data: Part 6: Determination of Statistical Tolerance Intervals. ISO, Geneva (2005)
23. Janouchová, E., Kučerová, A.: Competitive comparison of optimal designs of experiments for sampling-based sensitivity analysis. Comput. Struct. **124**, 47–60 (2013)
24. King, J.: Stochastic simulation methods in sensitivity analysis. Diploma thesis, University of Innsbruck, Austria (2007)
25. Kleijnen, J.P.C.: Design and Analysis of Simulation Experiments. Springer Science+Business Media LLC, New York (2008)
26. Kloeden, P.E., Platen, E.: Numerical Solution of Stochastic Differential Equations. Springer, Berlin (1992)
27. Krishnamoorthy, K., Mathew, T.: Statistical Tolerance Regions: Theory, Applications and Computation. Wiley, New Jersey (2009)
28. Kruskal, W.: Relative importance by averaging over orderings. Am. Stat. **41**, 6–10 (1987)

29. Le Maître, O., Knio, O.: Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics. Springer, New York (2010)
30. Loève, M.: Probability Theory, vol. II, 4th edn. Springer, New York (1978)
31. Martz, H.F., Waller, R.A.: Bayesian Reliability Analysis. Wiley, Chichester (1982)
32. Montgomery, D.C., Peck, E.A., Vining, G.G.: Introduction to Linear Regression Analysis, 5th edn. Wiley, New York (2012)
33. Nelsen, R.B.: An Introduction to Copulas, 2nd edn. Springer, New York (2006)
34. Niederreiter, H.: Random number generation and quasi-Monte Carlo methods. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (1992)
35. Oberguggenberger, M.: Analysis and computation with hybrid random set stochastic models. In: Deodatis, G., Ellingwood, B.R., Frangopol, D.M. (eds.) Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures, p. 93. CRC Press/Balkema, Leiden (2013)
36. Oberguggenberger, M.: Combined methods in nondeterministic mechanics. In: Elishakoff, I., Soize, C. (eds.) Nondeterministic Mechanics, pp. 263–356. Springer, Wien (2013)
37. Oberguggenberger, M., Aichinger, C., Caillaud, B., Haller, H., Roth, H.: Simulation tools for assessing the reliability and the design of shell structures. CD-ROM. In: Conference Proceedings, 4th International Conference on "Supply on the Wings", Frankfurt. AIRTEC Frankfurt, Paper No. D13 (2009)
38. Oberguggenberger, M., King, J., Schmelzer, B.: Classical and imprecise probability methods for sensitivity analysis in engineering: a case study. Int. J. Approx. Reason. **50**, 680–693 (2009)
39. Oberguggenberger, M., Ostermann, A.: Analysis for Computer Scientists: Foundations, Methods, and Algorithms. Springer, London (2011)
40. Ostermann, A.: Sensitivity analysis. In: Fellin, W., Lessmann, H., Oberguggenberger, M., Vieider, R. (eds.) Analyzing Uncertainty in Civil Engineering, pp. 101–114. Springer, Berlin (2005)
41. Riedinger, K.: Simulation of random fields for sensitivity analysis. Diploma thesis, University of Innsbruck, Austria (2010)
42. Robert, C.P., Casella, G.: Monte Carlo Statistical Methods, 2nd edn. Springer, New York (2004)
43. Roth, H.: Partial coefficient of determination. Internal Report, University of Innsbruck, Austria (2010)
44. Roth, H.: Sensitivity analysis with correlated variables. Diploma thesis, University of Innsbruck, Austria (2010)
45. Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S.: Global Sensitivity Analysis: The Primer. Wiley, Chichester (2008)
46. Schenk, C.A., Schuëller, G.I.: Uncertainty Assessment of Large Finite Element Systems. Springer, Berlin (2005)
47. Schuëller, G.I. (ed.): A benchmark study on reliability in high dimensions. Special Issue. Struct. Saf. **29**, 165–252 (2007)
48. Schuëller, G.I., Pradlwarter, H.J., Koutsourelakis, P.S.: A critical appraisal of reliability estimation procedures for high dimensions. Probab. Eng. Mech. **19**, 463–474 (2004)
49. Schwarz, M.: Monte Carlo based reliability analysis. Master thesis, University of Innsbruck, Austria (2013)
50. Shao, J., Tu, D.-S.: The Jackknife and Bootstrap. Springer, New York (1995)
51. Sobol, I.M.: Sensitivity analysis for nonlinear mathematical models. Math. Model. Comput. Experiment **1**, 407–414 (1993)
52. Sobol, I.M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math. Comput. Simul. **55**, 271–280 (2001)
53. Stein, M.: Large sample properties of simulations using Latin hypercube sampling. Technometrics **29**, 143–151 (1987)
54. Zuev, K.M., Beck, J.L., Au, S.-K., Katafygiotis, L.S.: Bayesian post-processor and other enhancements of Subset Simulation for estimating failure probabilities in high dimensions. Comput. Struct. **92–93**, 283–296 (2012)

Chapter 5
Multi-Phase Models in Civil Engineering

P. Gamnitzer, M. Aschaber, and G. Hofstetter

Abstract Some problems in civil engineering require the consideration of interactions between solids and fluids and/or between different physical phenomena, like thermal, hygral or chemical processes, for an appropriate description of the material behaviour and of the structural response. This chapter deals with the current developments of multi-phase models focusing on soils and concrete. The latter materials are characterized by a certain degree of permeability allowing liquid or gaseous phases to enter the pore space and to interact with the surrounding solid phase. Since the resulting interactions between the different phases may have a strong impact on the structural behaviour, they have to be accounted for appropriately in numerical models.

5.1 Introduction

Commonly, in civil engineering it is sufficient to model the mechanical material behaviour by appropriate stress–strain relations, frequently derived on the basis of plasticity theory, damage theory or combinations of the former theories, for predicting the structural response. However, some problems require considering additional physical phenomena, like thermal, hygral and/or chemical processes, and interactions between different physical processes for an appropriate description of the material behaviour and of the structural response.

P. Gamnitzer (✉) • G. Hofstetter
Unit of Strength of Materials and Structural Analysis, University of Innsbruck, Technikerstr. 13, A6020 Innsbruck, Austria
e-mail: Peter.Gamnitzer@uibk.ac.at; Guenter.Hofstetter@uibk.ac.at

M. Aschaber
Unit of Strength of Materials and Structural Analysis, University of Innsbruck, Innsbruck, Austria
e-mail: Matthias.Aschaber@uibk.ac.at

G. Hofstetter (ed.), *Computational Engineering*, DOI 10.1007/978-3-319-05933-4_5,
© Springer International Publishing Switzerland 2014

In this context multi-phase models allow to properly take into account coupling effects of different physical processes in a consistent manner and offer the advantage to simultaneously compute all unknowns on the basis of one coupled numerical model.

This chapter focuses on the application of multi-phase models to numerical simulations in civil engineering, in particular, in geotechnical engineering and concrete engineering. In the former case the behaviour of partially saturated soils, including the special case of water saturated soils, is modelled. Presented applications refer to the prediction of ground settlements, induced by lowering the groundwater table, and to the prediction of the instability of an earth dam due to leaking. In the latter case concrete is modelled as a porous material, the pores filled with water, dry air and water vapour. The described application concerns the behaviour of concrete overlays, which are used for the strengthening of existing structures. In this context the interactions between hardening of the overlay, drying due to moisture transfer to the environment and the mechanical behaviour are of interest. All presented applications have in common the formulation of the basic governing equations. Hence, at first the primary unknowns together with the thermodynamic state variables, followed by the governing equations, will be summarized briefly.

5.2 Primary Unknowns and Thermodynamic State Variables

For multi-phase approaches, describing partially saturated porous materials, the current state of a system depends on a set of primary unknowns, which for the present applications consist of

- the displacements $\mathbf{u}^s$,
- the pressure in the gaseous phase p^g,
- the capillary pressure p^c.

If thermal effects are included, then the primary unknowns are supplemented by the temperature T. For the applications discussed in this contribution, it is reasonable to assume the system to be in local thermodynamic equilibrium.

Several derived state variables for liquid phases can be obtained from the primary unknowns. The laws of physics and the empirical relations for determining the derived state variables are summarized subsequently, following [14, 15].

The pressure of the water phase is determined by

$$p^w\left(p^g, p^c\right) = p^g - p^c \ . \tag{5.1}$$

The density of water in the porous medium can be approximated using the state equation for bulk (free) liquid water:

$$\rho^w\left(p^g, p^c, T\right) = \rho_0^w \left(1 - \alpha_T^{\text{water}}\left(T - T_0\right) + C_w\left(p^w\left(p^g, p^c\right) - p_0\right)\right) \ . \tag{5.2}$$

If the pore water is considered to be incompressible, as it is recommended for instance in [14], this equation takes the form

$$\rho^w(T) = \rho_0^w \left(1 - \alpha_T^{\text{water}} (T - T_0)\right) . \tag{5.3}$$

Required constants are the water density $\rho_0^w = 999.84\,\text{kg/m}^3$ at reference temperature $T_0 = 273.15\,\text{K}$ and reference air pressure $p_0 = 101{,}325\,\text{Pa}$, the temperature-dependent thermal expansion coefficient of water α_T^{water}, for which in [14] a value between $0.68 \cdot 10^{-4}\,\text{K}^{-1}$ (at T_0) and $1.01 \cdot 10^{-3}\,\text{K}^{-1}$ (at 420 K) is proposed, and the isothermal compressibility of water $C_w = 4.58 \cdot 10^{-10}\,1/\text{Pa}$. The pressure of the vapour phase is determined from the primary unknowns using the Kelvin–Laplace equation

$$p^{gw}(p^c, p^g, T) = p^{gw,\text{sat}}(T) \cdot \exp\left(-\frac{p^c M_w}{\rho^w(p^g, p^c, T)\, R T}\right) \tag{5.4}$$

with $M_w = 18\,\text{g/mol}$ and $R = 8.314\,\text{J/(mol}\cdot\text{K)}$ denoting the molar mass of water and the universal gas constant, respectively. The Kelvin–Laplace equation (5.4) relates the actual vapour pressure p^{gw} to the corresponding saturated vapour pressure $p^{gw,\text{sat}}$. The latter is a function of temperature and can be computed from the vapour pressure $p_0^{gw,\text{sat}} = 2.3\,\text{kPa}$ at 293.15 K by means of the Clausius–Clapeyron equation

$$p^{gw,\text{sat}}(T) = p_0^{gw,\text{sat}} \cdot \exp\left(-\frac{M_w \cdot \Delta H_{\text{vap}}(T)}{R} \cdot \left(\frac{1}{T} - \frac{1}{293.15\text{K}}\right)\right) . \tag{5.5}$$

In (5.5) the specific enthalpy of evaporation is approximated according to the empirical Watson formula

$$\Delta H_{\text{vap}}(T) = 267.2 \cdot \left(\frac{647.3\text{K} - T}{\text{K}}\right)^{0.38} \quad \left[\frac{\text{kJ}}{\text{kg}}\right] . \tag{5.6}$$

Furthermore, water vapour is assumed to be an ideal gas and, thus, its density

$$\rho^{gw}(p^c, p^g, T) = \frac{M_w p^{gw}(p^c, p^g, T)}{R\,T} \tag{5.7}$$

is obtained from the ideal gas equation.

The pressure of the dry air fraction of the gas phase is defined according to Dalton's law:

$$p^{ga}(p^c, p^g, T) = p^g - p^{gw}(p^c, p^g, T) . \tag{5.8}$$

By analogy to (5.7) the dry air density is given as

$$\rho^{ga}\left(p^c, p^g, T\right) = \frac{M_a\, p^{ga}\left(p^c, p^g, T\right)}{R\, T}\,. \tag{5.9}$$

Since the gas phase is a mixture of vapour and dry air, the resulting combined gas density reads as

$$\rho^g = \rho^{gw} + \rho^{ga}\,. \tag{5.10}$$

If the ideal gas equation is assumed also to hold for the mixture of water vapour and dry air, then its pressure and the molar mass are determined as

$$p^g = \rho^g \frac{R}{M_g} T \tag{5.11}$$

and

$$M_g = \frac{M_a M_w \rho^g}{M_a \rho^{gw} + M_w \rho^{ga}}\,. \tag{5.12}$$

5.3 Governing Equations

The balance equations contain a number of volume fractions. They result from the derivation of the macroscopic balance equations from the microscopic balance equations. Usually, they are expressed in terms of porosity and fluid saturation [24]. The porosity

$$n = \frac{V - V_s}{V} \tag{5.13}$$

describes the ratio of the total volume minus the volume occupied by the solid phase, $V - V_s$, to the total volume V. The degrees of water and gas saturation

$$S_w = \frac{V_w}{V - V_s}, \qquad S_g = \frac{V_g}{V - V_s} = 1 - S_w \tag{5.14}$$

are defined as the ratio of the volume occupied by the respective fluid phase to the volume occupied by all fluid phases. Equations (5.13) and (5.14) are derived from a representative cell of the multi-phase material.

Commonly, the empirical relation according to [27]

$$S_w\left(p^c\right) = S_w^r + \left(S_w^s - S_w^r\right)\left(1 + \left(\frac{p^c}{p_b^c}\right)^{\frac{1}{1-m}}\right)^{-m} \tag{5.15}$$

between capillary pressure and the degree of water saturation is employed. S_w^r and S_w^s denote the residual and maximum degree of water saturation and p_b^c and m represent the air entry value and a fitting parameter.

5.3.1 Balance Laws

The balance laws are summarized along the lines of [15, 19]. They consist of the mass balance equations for each phase and the balance of momentum equation and the enthalpy balance equation for the multi-phase mixture. In a first step the mass balance equations are formulated in terms of the unknown relative velocities between the individual phases. In a second step they will be closed by approximations for the fluid fluxes, which can also be found in the above mentioned publications. Based on the density of the solid phase ρ^s, the velocity $\mathbf{v}^s = d\mathbf{u}^s/dt$ of the solid phase, and a mass exchange term $\dot{m}_{\text{hydr}}$, the balance equation for the solid phase can be derived as

$$\dot{m}_{\text{hydr}} = \left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} \left[(1-n)\,\rho^s\right] + (\nabla \circ \mathbf{v}^s)\left[(1-n)\,\rho^s\right] . \tag{5.16}$$

The operator $\left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} \bullet$ indicates the time derivative to be taken with respect to a fixed position $\mathbf{X}_s$ in the (undeformed) structural reference configuration. Later on, for applications to concrete overlays, the mass exchange term $\dot{m}_{\text{hydr}}$ will be identified with mass exchange due to chemical reaction/hydration processes in concrete. The mass balance for the solid phase is commonly used for eliminating the time derivative of the porosity in the balance equations for the fluid phases [15].

For the water phase, the balance equation is obtained as

$$\begin{aligned} -\dot{m}_{\text{hydr}} - \dot{m}_{\text{vap}} = {} & \left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} (n S_w \rho^w) + (\nabla \circ \mathbf{v}^s)\left[n S_w \rho^w\right] + \\ & + \nabla \circ \left(n S_w \rho^w \left(\mathbf{v}^w - \mathbf{v}^s\right)\right) . \end{aligned} \tag{5.17}$$

Again, although being a balance law for the apparent fluid density in the current configuration, it is stated with respect to the (undeformed) structural configuration. Equation (5.17) is formulated in terms of the water velocity $\mathbf{v}^w$ and a second mass exchange term $\dot{m}_{\text{vap}}$ accounting for the phase transition of water to water vapour.

The mass balance of dry air is derived as

$$\begin{aligned} 0 = {} & \left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} \left(n S_g \rho^{ga}\right) + (\nabla \circ \mathbf{v}^s)\left[n S_g \rho^{ga}\right] \\ & + \nabla \circ \left(n S_g \rho^{ga} \left(\mathbf{v}^{ga} - \mathbf{v}^g\right)\right) + \nabla \circ \left(n S_g \rho^{ga} \left(\mathbf{v}^g - \mathbf{v}^s\right)\right) . \end{aligned} \tag{5.18}$$

It contains the velocity $\mathbf{v}^{ga}$ of dry air, and, furthermore, is augmented by the averaged velocity of the gaseous phase

$$\mathbf{v}^g = \frac{1}{\rho^g}\left(\rho^{ga}\mathbf{v}^{ga} + \rho^{gw}\mathbf{v}^{gw}\right) \tag{5.19}$$

with $\mathbf{v}^{gw}$ denoting the velocity of the vapour phase. Last but not least, the mass balance for the vapour phase is obtained as

$$\begin{aligned}\dot{m}_{\mathrm{vap}} = {} & \left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s}\left(nS_g\rho^{gw}\right) + \left(\nabla\circ\mathbf{v}^s\right)\left(nS_g\rho^{gw}\right) \\ & + \nabla\circ\left(nS_g\rho^{gw}\left(\mathbf{v}^{gw}-\mathbf{v}^g\right)\right) + \nabla\circ\left(nS_g\rho^{gw}\left(\mathbf{v}^g-\mathbf{v}^s\right)\right).\end{aligned} \tag{5.20}$$

By analogy to (5.17), (5.20) is augmented by the averaged velocity of the gas phase $\mathbf{v}^g$. As usual, both mass balances for dry air and vapour are stated with respect to the (undeformed) structural reference configuration. The mass-balance equation of the vapour phase can be used for eliminating $\dot{m}_{\mathrm{vap}}$ from the mass balance of the water phase.

Quasistatic equilibrium is expressed by the balance of momentum equation

$$\nabla\circ\boldsymbol{\sigma} + \rho\mathbf{g} = \mathbf{0}\,. \tag{5.21}$$

In (5.21) $\mathbf{g}$ denotes the gravitational acceleration,

$$\rho = (1-n)\,\rho^s + n\left(S_w\rho^w + S_g\rho^g\right) \tag{5.22}$$

is the averaged density of the three-phase mixture and $\boldsymbol{\sigma}$ represents the Cauchy-stress. Tensile stresses are assumed to be positive. The solid skeleton density is assumed to depend only on temperature:

$$\rho^s = \rho_0^s\cdot\exp\left(-\alpha_T\left(T-T_0\right)\right)\,. \tag{5.23}$$

The system of balance equations is completed by the enthalpy balance of the three-phase mixture

$$\begin{aligned}\dot{m}_{\mathrm{hydr}}\Delta H_{\mathrm{hyd}} - \dot{m}_{\mathrm{vap}}\Delta H_{\mathrm{vap}} = {} & \left[C_p^s\,(1-n)\,\rho^s + C_p^w n S_w\rho^w + C_p^g n S_g\rho^g\right]\left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} T \\ & + \left[C_p^w\left(nS_w\rho^w\left(\mathbf{v}^w-\mathbf{v}^s\right)\right) + C_p^g\left(nS_g\rho^g\left(\mathbf{v}^g-\mathbf{v}^s\right)\right)\right]\circ\nabla T \\ & + \nabla\circ\left[\mathbf{q}\left(T\right)\right]\,.\end{aligned} \tag{5.24}$$

In (5.24) C_p^s, C_p^w, and C_p^g are the heat capacities of the individual phases. The flux $\mathbf{q}(T)$ accounts for the thermal conductivity of the three-phase material, and ΔH_{hyd} and ΔH_{vap} are the specific enthalpies of hydration and vapourization.

5.3.2 Flux Approximations and Stress State Variables

The balance laws, introduced in the previous subsection, contain a number of up to now unspecified flux terms, i.e.,

- mass fluxes of dry air and water vapour with respect to the gas phase caused by diffusive processes,
- mass fluxes of water and gas due to the relative motion of the fluid phases with respect to the solid skeleton,
- thermal fluxes driven by temperature gradients,
- fluxes of momentum, i.e., stresses.

In a porous medium, the individual phases contribute to the total stress present in the balance of momentum equation (5.21). Aiming at the formulation of constitutive equations, the total stress can be decomposed into an effective stress $\boldsymbol{\sigma}_{\text{eff}}$, related to the solid matrix deformations, and a hydrostatic pressure

$$p^s = (p^g - p_{\text{atm}}) - \chi_{\text{Bishop}}(S_w)\, p^c \,, \tag{5.25}$$

which accounts for the pressure exerted by the pore fluids on the solid matrix. Thus,

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}_{\text{eff}} - \alpha_{\text{Biot}} p^s \mathbf{1} \tag{5.26}$$

with the Biot coefficient $\alpha_{\text{Biot}} = 1 - K/K_s$ accounting for the compressibility of the grains, where K and K_s are the bulk moduli of the solid skeleton and the solid grains.

In (5.25) p_{atm} represents the atmospheric air pressure, and $\chi_{\text{Bishop}}(S_w)$ is the generalized, saturation dependent Bishop-parameter.

Heat fluxes are modelled by Fourier's law

$$\mathbf{q} = -\chi_{\text{eff}} \nabla T \tag{5.27}$$

where χ_{eff} denotes the effective thermal conductivity. For wet materials, the latter can be obtained from the thermal conductivity at dry conditions, χ_{dry}, using, for instance, the relation provided in [15] for concrete:

$$\chi_{\text{eff}} = \chi_{\text{dry}}(T) \cdot \left[1 + \frac{4n\rho^w S_w}{(1-n)\rho^s}\right] . \tag{5.28}$$

Darcy type flow of water and gas with respect to the solid skeleton is assumed, i.e.,

$$nS_w\left(\mathbf{v}^w - \mathbf{v}^s\right) = \frac{k_{w,\mathrm{rel}} K}{\mu_w}\left(-\nabla p^w + \rho^w \mathbf{g}\right) \tag{5.29}$$

and

$$nS_g\left(\mathbf{v}^g - \mathbf{v}^s\right) = \frac{k_{g,\mathrm{rel}} K}{\mu_g}\left(-\nabla p^g + \rho^g \mathbf{g}\right) . \tag{5.30}$$

The involved material parameters are the intrinsic permeability of the solid skeleton K, the relative permeabilities $k_{w,\mathrm{rel}}$ and $k_{g,\mathrm{rel}}$ which account for the change in permeability with the degree of water saturation, and the dynamic viscosities of the gas and water phase. The latter are temperature dependent and can be computed according to [14]. The relationship between dynamic water viscosity and temperature reads as

$$\mu_w(T) = 0.6612\left(\frac{T - 229\mathrm{K}}{\mathrm{K}}\right)^{-1.562} \quad [\mathrm{Pa}\cdot\mathrm{s}] . \tag{5.31}$$

For the gas phase, the dynamic viscosity depends on additional thermodynamic state variables. It can be computed from the dynamic viscosities of dry air and water vapour according to

$$\mu_g\left(p^c, p^g, T\right) = \mu_{gw}(T) + \left(\mu_{ga}(T) - \mu_{gw}(T)\right)\left(\frac{p^{ga}\left(p^c, p^g, T\right)}{p^g}\right)^{0.608} . \tag{5.32}$$

In (5.32) the temperature dependence of the dynamic viscosities of water vapour and of dry air is approximated by

$$\mu_{gw}(T) = \mu_{gw}^0 + \alpha_v\left(T - T_0\right) \tag{5.33}$$

and

$$\mu_{ga}(T) = \mu_{ga}^0 + \alpha_a\left(T - T_0\right) + \beta_a\left(T - T_0\right)^2 , \tag{5.34}$$

respectively, with the dynamic viscosities at reference temperature $\mu_{gw}^0 = 8.85 \cdot 10^{-8}\,\mathrm{Pa}\cdot\mathrm{s}$ for water vapour and $\mu_{ga}^0 = 17.17 \cdot 10^{-6}\,\mathrm{Pa}\cdot\mathrm{s}$ for dry air and the constants $\alpha_v = 3.53 \cdot 10^{-9}\,(\mathrm{Pa}\cdot\mathrm{s})/\mathrm{K}$ for water vapour, $\alpha_a = 4.73 \cdot 10^{-8}\,(\mathrm{Pa}\cdot\mathrm{s})/\mathrm{K}$, and $\beta_a = 2.22 \cdot 10^{-11}\,(\mathrm{Pa}\cdot\mathrm{s})/\mathrm{K}^2$ for dry air.

The relative permeabilities are defined based on the effective degree of saturation

$$S_e = \frac{S_w - S_w^r}{S_w^s - S_w^r} , \tag{5.35}$$

as in [17] by

$$k_{w,\mathrm{rel}} = \sqrt{S_e}\left[1 - \left(1 - S_e^{\frac{1}{m}}\right)^m\right]^2 \tag{5.36}$$

and

$$k_{g,\mathrm{rel}} = \sqrt{1 - S_e}\left[1 - S_e^{\frac{1}{m}}\right]^{2m} . \tag{5.37}$$

Equations (5.36) and (5.37) are based on a non-hysteretic capillary pressure-saturation function according to (5.15) and can already be found in a similar form for three-phase flow in [22], for instance. In practice, for the applications described below, a minimum value for the relative air permeability is assumed for numerical reasons, see for instance [23].

Finally, the relative flow of water vapour with respect to the gas phase is assumed to be of Fickian diffusion type [24]:

$$\left[n S_g \rho^{gw}\right]\left(\mathbf{v}^{gw} - \mathbf{v}^{g}\right) = -\rho^g D_g^{gw} \nabla\left(\frac{\rho^{gw}}{\rho^g}\right) \tag{5.38}$$

with the diffusion coefficient according to [14]

$$D_g^{gw}\left(T, p^g p^c\right) = n\left(1 - S_w\left(p^c\right)\right) f_S\, D_0 \frac{p_0}{p^g}\left(\frac{T}{T_0}\right)^{1.667} . \tag{5.39}$$

It is computed from the diffusion coefficient $D_0 = 2.58 \cdot 10^{-5}\,\mathrm{m^2/s}$ for free diffusion of vapour in air at reference temperature, supplemented by the term $n\left(1 - S_w\left(p^c\right)\right) f_S$ including the structure coefficient f_S, which provides a modification for vapour diffusion in porous media.

By definition of the velocity of the gas phase according to (5.19), (5.38) also describes the relative flow of dry air with respect to the gas phase:

$$\left[n S_g \rho^{ga}\right]\left(\mathbf{v}^{ga} - \mathbf{v}^{g}\right) = -\left[n S_g \rho^{gw}\right]\left(\mathbf{v}^{gw} - \mathbf{v}^{g}\right) . \tag{5.40}$$

5.4 Application to Geotechnical Engineering

In soil mechanics the soil grains are commonly assumed as incompressible, i.e., $\alpha_{\mathrm{Biot}} = 1$ in (5.26). In this case volumetric strains of the soil are solely related to changes in porosity. For the Bishop-parameter in (5.25), a common approach is to use the thermodynamically consistent choice $\chi_{\mathrm{Bishop}}\left(S_w\right) = S_w$, see [24]. The

corresponding momentum balance equation then follows from (5.21) and (5.26) together with (5.25) as

$$\nabla \circ \{\boldsymbol{\sigma}_{\text{eff}} - [(p^g - p_{\text{atm}}) - S_w p^c]\,\mathbf{1}\} + \rho \mathbf{g} = \mathbf{0}\,. \tag{5.41}$$

In many geotechnical applications it is furthermore admissible to neglect all exchange terms as well as the influence of temperature variations and the presence of the vapour phase. This gives rise to a simplified three-phase model based on a reduced set of balance equations, which consists of the equilibrium equation (5.41) and the three mass balance equations (5.16), (5.17) and (5.18) for the solid, water, and gas phase, respectively. Inserting the fluxes (5.29) and (5.30), the mass balances are obtained as

$$\left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} n = (\nabla \circ \mathbf{v}^s)\,(1-n)\ , \tag{5.42}$$

$$n \left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} (S_w \rho^w) + (\nabla \circ \mathbf{v}^s)\,[S_w \rho^w] - \nabla \circ \left(\frac{k_{w,\text{rel}} \rho^w K}{\mu_w} (\nabla p^w - \rho^w \mathbf{g}) \right) = 0\,. \tag{5.43}$$

$$n \left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} \left(S_g \rho^g\right) + (\nabla \circ \mathbf{v}^s)\left[S_g \rho^g\right] - \nabla \circ \left(\frac{k_{g,\text{rel}} \rho^g K}{\mu_g} (\nabla p^g - \rho^g \mathbf{g}) \right) = 0\,. \tag{5.44}$$

At this stage, a constitutive law providing a closure for the effective stress is the only missing link to complete the three-phase framework. However, soils exhibit a quite complex constitutive behaviour, depending on the stress state and the capillary pressure. A model for describing the latter will be presented in Sect. 5.4.1.

For simplicity, in what follows small strains and small deformations are assumed. Thus, the strain $\boldsymbol{\varepsilon}$ is given by

$$\boldsymbol{\varepsilon} = \frac{(\nabla \mathbf{u}^s) + (\nabla \mathbf{u}^s)^T}{2}\ , \tag{5.45}$$

using a sign convention that associates compaction with negative strain values. The volumetric strain ε_V follows from (5.45) as

$$\varepsilon_V = \text{tr}\,(\boldsymbol{\varepsilon}) = \nabla \circ \mathbf{u}^s\ . \tag{5.46}$$

By taking the time derivative of (5.46), a relation between volumetric strain rate and solid velocity divergence is obtained,

$$\left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} \varepsilon_V = \left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} \nabla \circ \mathbf{u}^s \approx \nabla \circ \left(\left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} \mathbf{u}^s \right) = \nabla \circ \mathbf{v}^s\ . \tag{5.47}$$

Note that for this approximation the small deformation assumption is crucial, since it allows to interchange the time derivative with respect to the structural frame of reference with the spatial divergence operator. Equation (5.47) can be used to restate the mass balance equation for the solid phase as

$$\frac{1}{1-n} \cdot \left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} n = \left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} \varepsilon_V . \tag{5.48}$$

Furthermore, in geotechnics, it is popular to replace the porosity by the specific volume $v = 1/(1-n)$. In terms of the latter, the mass balance equation for the solid phase (5.48) can be restated as

$$\left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} \varepsilon_V = \left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} \ln v . \tag{5.49}$$

5.4.1 A Modified Cap Model for Unsaturated and Saturated Soil

The modified cap model for partially saturated soils, including the limiting case of water saturated soils, is a further development of the model from [18], which is an extension of the original cap model, proposed in [8], to partially saturated soils. It is characterized by [12, 13]

- nonlinear elastic behaviour,
- a yield surface enclosing the elastic domain in water saturated conditions,
- evolution of the yield surface depending on the capillary pressure,
- a non-associated flow rule for the plastic strain rate,
- a hardening rule for the cap.

The elastic response is described in rate form, separately for the volumetric part $I_1^{\text{eff}} = \operatorname{tr}(\boldsymbol{\sigma}_{\text{eff}})$ and the deviatoric part of the effective stress $\mathbf{s} = \boldsymbol{\sigma}_{\text{eff}} - \frac{1}{3} I_1^{\text{eff}} \cdot \mathbf{1}$:

$$\left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} \ln\left(-I_1^{\text{eff}}\right) = -\frac{v}{\kappa_E} \left(\left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} \varepsilon_V - \left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} \varepsilon_V^p \right), \tag{5.50}$$

$$\left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} \mathbf{s} = 2G \left(\left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} \mathbf{e} - \left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} \mathbf{e}^P \right). \tag{5.51}$$

The (inverse) elastic volumetric stiffness parameter κ_E and the shear modulus G are constants. $\varepsilon_V^p = \operatorname{tr}\left(\boldsymbol{\varepsilon}^P\right)$ is the volumetric part of the plastic strain tensor $\boldsymbol{\varepsilon}^P$. Furthermore, $\mathbf{e} = \boldsymbol{\varepsilon} - \frac{1}{3}\varepsilon_V \cdot \mathbf{1}$ and $\mathbf{e}^P = \boldsymbol{\varepsilon}^P - \frac{1}{3}\varepsilon_V^p \cdot \mathbf{1}$ denote the deviatoric parts of the total strain tensor and the plastic strain tensor, respectively. In the present

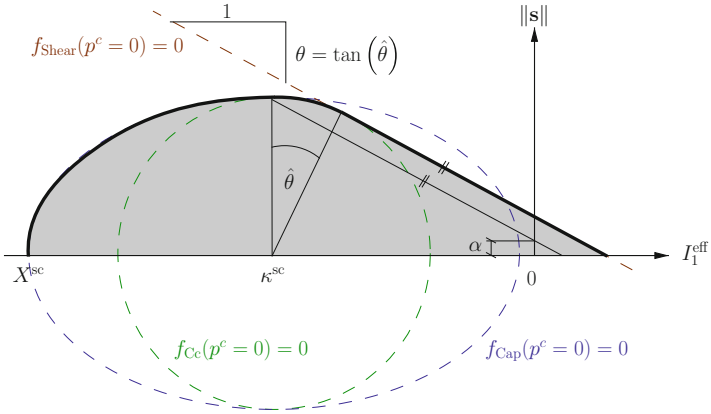

Fig. 5.1 The elastic domain at full saturation (*grey*) is bounded by three segments, defined by the three yield functions f_{Shear}, f_{Cap}, and f_{Cc}

form, the volumetric part of the elastic law does not support tensile stress states, i.e. positive values of I_1^{eff} cannot be represented.

The yield surface at water saturated conditions is formulated in the I_1^{eff}-$\|\mathbf{s}\|$-space in the framework of multi-surface plasticity. The three parts of the yield surface consist of a linear shear failure envelope (index *Shear*), an elliptic hardening cap (index *Cap*) and a circular transition zone between the hardening cap and the shear failure envelope (index *Cc*), the latter ensuring a smooth transition between the former yield functions [9] . The material parameters, defining the shape of the elastic domain at full saturation, are the ellipticity parameter R for the cap surface, the slope parameter θ for the shear envelope and the cohesion parameter α. Please note that in contrast to the classical approach, in the current smooth model the effective cohesion is slightly larger due to the smoothing of the corner between shear failure envelope and cap. The shape and size of the elastic domain also depend on the position of the centre of the ellipsoidal hardening cap κ^{sc}, which serves as a hardening parameter. κ^{sc} can be converted to the volumetric preconsolidation stress at full saturation X^{sc}. The zero-isosurfaces of the yield functions at full saturation together with the described parameters are depicted in Fig. 5.1.

The change of the yield surface with capillary pressure is governed by two relationships. The first one defines an increase in cohesion with increasing values of p^c, and the second one, the load-collapse yield curve, defines the change in preconsolidation stress with p^c. The two relationships used in the present model are postulated using net stresses and are then transferred to effective stresses using a smoothed conversion function F_n defined by

$$F_n\left(p^c\right) = \begin{cases} 3p^c \cdot \left[S_w\left(p^c\right) - 1 + \frac{2p^c}{L} - \frac{\left(p^c\right)^2}{L^2}\right], & 0 \le p^c < L\,, \\ 3p^c \cdot S_w\left(p^c\right) & , \; L \le p^c\,. \end{cases} \tag{5.52}$$

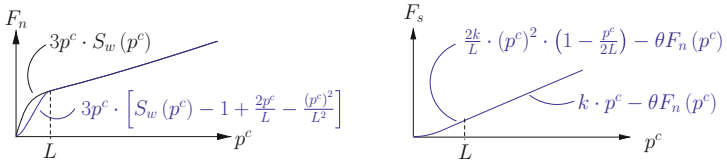

Fig. 5.2 *Left*: The modified conversion function F_n (5.52) (*blue*) from net stress to effective stress space. The sketched example was generated using $L = 0.15\,\text{MPa}$ and the Van Genuchten parameters $p_b^c = 0.09\,\text{MPa}$, $m = 0.55$, $S_w^r = 0.1$, and $S_w^s = 0.9$. *Right*: Schematic plot of the (*smoothed*) function F_s (5.56) defining the increase in cohesion with p^c

For p^c larger than the user-defined parameter L, F_n provides the exact conversion of the first invariant of net stress to the first invariant of effective stress via

$$I_1^{\text{eff}} = I_1^{\text{net}} - F_n(p^c) \ . \tag{5.53}$$

As depicted on the left side of Fig. 5.2, for values of the capillary pressure smaller than L the transformation is modified by a polynomial smoothing such that

$$\frac{\partial F_n}{\partial p^c}(p^c = 0) = 0 \, .$$

Based on this conversion, the load collapse yield curve in the modified cap model, describing the evolution of the effective preconsolidation stress X^{eff} as a function of capillary pressure, reads as

$$X^{\text{eff}}(\kappa^{\text{sc}}, p^c) = X^{\text{net}}(\kappa^{\text{sc}}, p^c) - F_n(p^c) \tag{5.54}$$

with

$$X^{\text{net}}(\kappa^{\text{sc}}, p^c) = \begin{cases} X^{\text{sc}}(\kappa^{\text{sc}}) - X_c \cdot \left(1 - e^{-\frac{(p^c)^2}{\sigma^2 (p^c - p^c_{\max})^2}}\right) & , \ p^c < p^c_{\max} \, , \\ X^{\text{sc}}(\kappa^{\text{sc}}) - X_c & , \ p^c_{\max} \leq p^c . \end{cases} \tag{5.55}$$

The parameter X_c defines the maximum increase in hydrostatic compressive strength with increasing values of p^c. The maximum is reached at $p^c_{\max}$, and σ^2 is a positive parameter controlling how fast the increase in compressive strength takes place, see Fig. 5.3 for a visualization. Up to smoothness corrections, X^{net} is equivalent to the load collapse curve in net stress space. The definition of the load collapse yield curve (5.54) together with (5.55) was motivated by experimental data in [20]. It features a smooth, differentiable transition to full saturation. The assumption of the increase in preconsolidation pressure with p^c independent of the current value of κ^{sc} is rather simplistic and, hence, it will not be valid for arbitrary ranges of p^c and κ^{sc}. However, it allows a reasonable fit to experimental data from

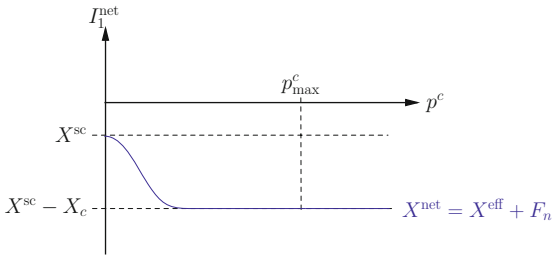

Fig. 5.3 General design of the load collapse yield curve in net stresses

both [5,20]. In addition, it enables an improved representation of wetting paths plus an increased robustness due to the smoothness, see [12,13].

The cohesion is assumed to basically increase proportionally to capillary pressure, augmented by a smoothing term for $p^c < L$. The respective function F_s is given by

$$F_s\left(p^c\right) = \begin{cases} \frac{2k}{L} \cdot \left(p^c\right)^2 \cdot \left(1 - \frac{p^c}{2L}\right) - \theta F_n\left(p^c\right) & , 0 \le p^c < L \,, \\ k \cdot^c - \theta F_n\left(p^c\right) & , L \le p^c. \end{cases} \tag{5.56}$$

k is a proportionality parameter and F_s is defined such that the derivative at $p^c = 0$ vanishes, see the right side of Fig. 5.2. Using these relationships, the three yield functions defining the yield surface read as

$$\begin{aligned} f_{\text{Shear}} = {} & L\left(\vartheta\right) \|\mathbf{s}\| + \theta \cdot I_1^{\text{eff}} \\ & - \left[\left(\alpha + F_s\left(p^c\right)\right) \cdot \sqrt{1+\theta^2} - \theta \cdot \left(\left(\sqrt{1+\theta^2} - 1 \right) \cdot \kappa^{\text{eff}} \right) \right] , \end{aligned} \tag{5.57}$$

$$f_{\text{Cc}} = \sqrt{\left(L\left(\vartheta\right)\|\mathbf{s}\|\right)^2 + \left(I_1^{\text{eff}} - \kappa^{\text{eff}}\right)^2} + \theta\,\kappa^{\text{eff}} - \alpha - F_s\left(p^c\right) \,, \tag{5.58}$$

$$f_{\text{Cap}} = \sqrt{\left(L\left(\vartheta\right)\|\mathbf{s}\|\right)^2 + \left(\frac{I_1^{\text{eff}} - \kappa^{\text{eff}}}{R}\right)^2} + \theta\,\kappa^{\text{eff}} - \alpha - F_s\left(p^c\right) \,. \tag{5.59}$$

The dependence of the shape of the elastic domain on the Lode angle ϑ is described by (see, e.g., [7,10,18])

$$L\left(\vartheta\right) = \left(\frac{1 - \omega\cos\left(3\vartheta\right)}{1-\omega}\right)^{-\eta}, \quad \cos\left(3\vartheta\right) = -\frac{3\sqrt{3}}{2}\frac{I_3^{\text{eff},s}}{\left(I_2^{\text{eff},s}\right)^{\frac{3}{2}}} \tag{5.60}$$

in terms of the parameters ω and η. The Lode angle ϑ can be computed from the second and third invariant of the deviatoric part of the effective stress, $I_2^{\text{eff},s}$ and $I_3^{\text{eff},s}$. Equation (5.60) is visualized in Fig. 5.4. It accounts for the fact that for a

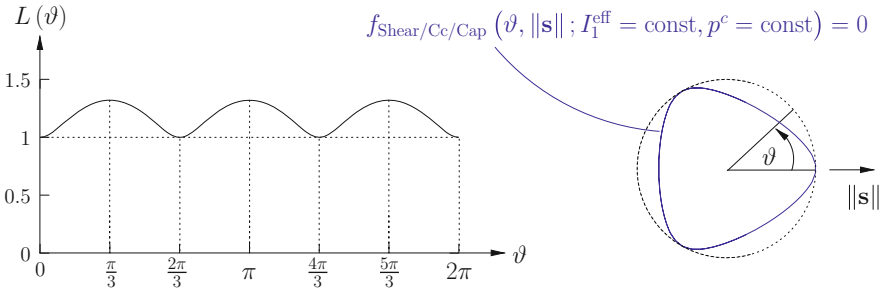

Fig. 5.4 Plots for the function $L(\vartheta)$ and the shape of the yield surface in a deviatoric plane for $\omega = 0.6$ and $\eta = -0.2$ (polar plot of 0-isosurface)

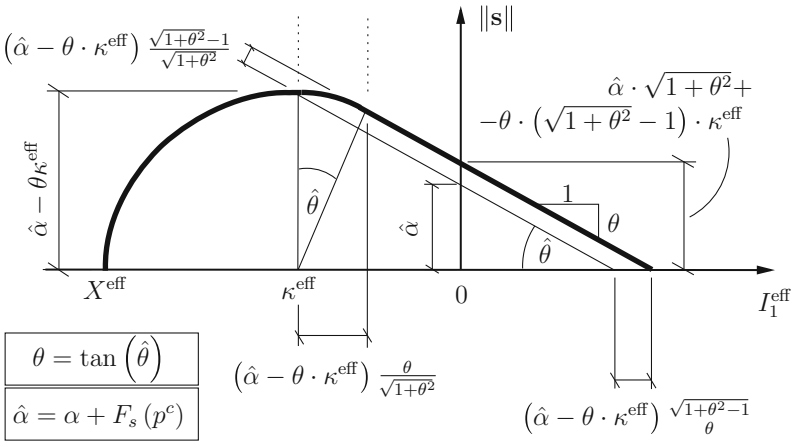

Fig. 5.5 The elastic domain at partial saturation corresponding to a particular capillary pressure p^c

given deviatoric plane the material strength on compressive meridians is higher than on tensile meridians. In (5.59) κ^{eff} determines the position of the centre of the ellipsoidal cap at capillary pressure p^c. It can be obtained from $X^{\text{eff}}(\kappa^{\text{sc}}, p^c)$ according to (5.54) together with (5.55) based on the ellipsoidal shape of the cap using the equation (see Fig. 5.5)

$$\kappa^{\text{eff}}(\kappa^{\text{sc}}, p^c) = \frac{X^{\text{eff}}(\kappa^{\text{sc}}, p^c) + R \cdot [\alpha + F_s(p^c)]}{1 + R\theta} . \tag{5.61}$$

The boundary of the elastic domain is visualized in Fig. 5.6. The figure emphasizes the smooth transition from partial saturation, where capillary pressure influences the material behaviour, to full saturation, where the effective stress is the only stress state variable required for the description of the material response. In all three modes of plastic loading, the plastic strain rate is derived by a non-associated flow rule:

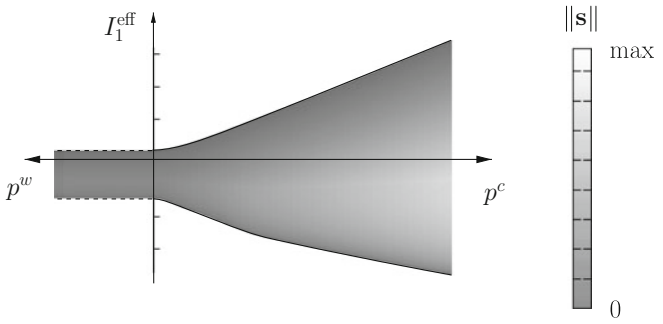

Fig. 5.6 Contour plot of the yield surface in I_1^{eff}-$\|\mathbf{s}\|$-p^c-space

$$\left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} \boldsymbol{\varepsilon}^P = \dot{\gamma}_{\text{Shear}} \cdot \frac{\partial g_{\text{Shear}}}{\partial \boldsymbol{\sigma}_{\text{eff}}} , \tag{5.62}$$

$$\left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} \boldsymbol{\varepsilon}^P = \dot{\gamma}_{\text{Cc}} \cdot \frac{\partial g_{\text{Cc}}}{\partial \boldsymbol{\sigma}_{\text{eff}}} , \tag{5.63}$$

$$\left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} \boldsymbol{\varepsilon}^P = \dot{\gamma}_{\text{Cap}} \cdot \frac{\partial g_{\text{Cap}}}{\partial \boldsymbol{\sigma}_{\text{eff}}} , \tag{5.64}$$

with $\dot{\gamma}_{\text{Cap}}$, $\dot{\gamma}_{\text{Cc}}$, and $\dot{\gamma}_{\text{Shear}}$ denoting the plastic multipliers and the plastic potentials

$$g_{\text{Shear}} = \|\mathbf{s}\| + \theta I_1^{\text{eff}} - \left[\left(\alpha + F_s\left(p^c\right)\right) \cdot \sqrt{1+\theta^2} - \theta \cdot \left(\left(\sqrt{1+\theta^2} - 1 \right) \cdot \kappa^{\text{eff}} \right) \right] , \tag{5.65}$$

$$g_{\text{Cc}} = \sqrt{\|\mathbf{s}\|^2 + \left(I_1^{\text{eff}} - \kappa^{\text{eff}}\right)^2} + \theta \kappa^{\text{eff}} - \alpha - F_s\left(p^c\right) , \tag{5.66}$$

$$g_{\text{Cap}} = \sqrt{\|\mathbf{s}\|^2 + \left(\frac{I_1^{\text{eff}} - \kappa^{\text{eff}}}{R} \right)^2} + \theta \kappa^{\text{eff}} - \alpha - F_s\left(p^c\right) . \tag{5.67}$$

Perfect plasticity is assumed for the shear failure envelope mode and for the circular region between shear failure envelope and cap. Hardening on the cap is governed by the hardening law

$$\left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} \varepsilon_V^p\left(\kappa^{\text{sc}}, p^c\right) = -\lambda\left(p^c\right) \cdot \left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} \ln\left(-X^{\text{net}}\left(\kappa^{\text{sc}}, p^c\right)\right) . \tag{5.68}$$

By analogy to the Barcelona Basic model [1] the volumetric plastic strain rate is assumed to be proportional to the rate of change of preconsolidation stress

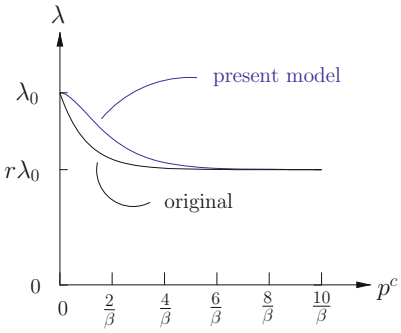

Fig. 5.7 The evolution of the proportionality parameter λ in the hardening law with p^c; comparison of the present approach to the original formulation

(transferred to net stress space). The proportionality factor, depending on the capillary pressure, is given by

$$\lambda\left(p^c\right) = \lambda_0 \cdot \left(\left(1-r\right)\left(1+\beta p^c\right) e^{-\beta p^c} + r\right) . \tag{5.69}$$

It is a smoothed version of the parameter used in [1], based on the material parameters λ_0 at full saturation, r for the residual value of the proportionality parameter at infinite p^c, and β controlling the exponential decrease from the value at full saturation to the residual value (see also Fig. 5.7).

5.4.2 A Model Problem for Ground Settlements

Changes of the groundwater table, as they are caused, for instance, by pumping, typically result in ground settlements. The reliable prediction of the latter is essential for assessing the serviceability and safety of property in affected areas. The model problem, described in the current section, is intended to highlight the capabilities of the three-phase approach for predicting such effects. It refers to a soil column of 10 m height with the groundwater table located 3 m below the surface. Geometric properties and boundary conditions are outlined in Fig. 5.8. Initially both, the air pressure and water pressure distribution, are assumed to be linear, corresponding to hydrostatic equilibrium. Displacements are zero at the bottom of the column. Horizontal displacements are assumed to vanish throughout the domain. At the top surface the atmospheric air pressure defines a constant air pressure boundary condition of Dirichlet type. The vertical boundaries of the column are assumed to be impermeable. Thus, the water content in the column is controlled only by a prescribed mass flux of water across the bottom boundary.

The parameters for the three-phase problem are summarized in Table 5.1, while material parameters for the employed cap model are listed in Table 5.2.

The problem is solved numerically using a finite element approach based on a coarse 2D discretization consisting of ten elements with quadratic serendipity shape

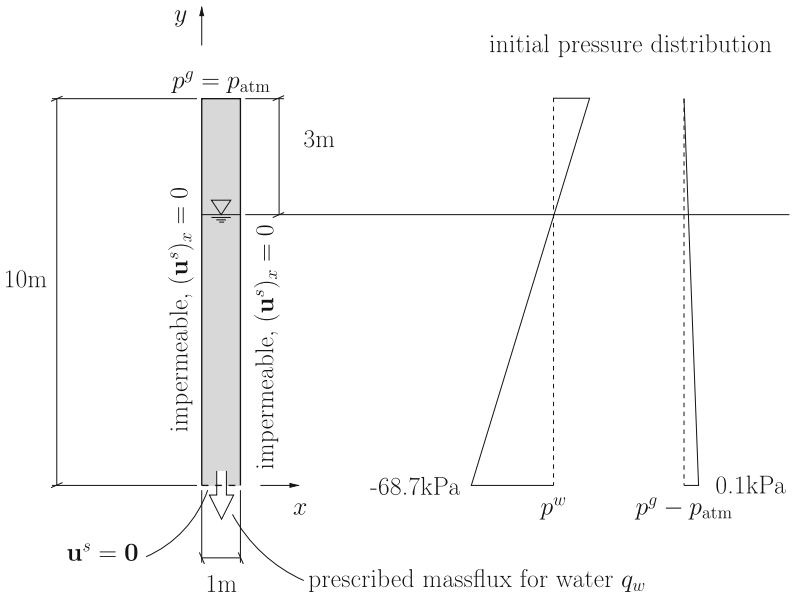

Fig. 5.8 Geometry and boundary conditions for the model problem

Table 5.1 Parameters of the three-phase formulation for the model problem

Gravitational acceleration	g	9.81	m/s^2
Atmospheric air pressure	p_{atm}	101,300	Pa
Dynamic viscosity water	μ_w	$1.0 \cdot 10^{-3}$	Pa·s
Dynamic viscosity gas (air)	μ_g	$1.75 \cdot 10^{-5}$	Pa · s
Compressibility water	C_w	$4.58 \cdot 10^{-10}$	1/Pa
Density water	ρ^w	1,000	kg/m^3
Density air	ρ^g	1.3	kg/m^3
Initial porosity	n	0.5	–
Density solid grains	ρ^s	2,500	kg/m^3
Residual saturation	S_w^r	0.25	–
Maximum saturation	S_w^s	0.95	–
Air entry value	p_b^c	$5 \cdot 10^4$	Pa
Van-Genuchten fitting parameter	m	0.33	–
Intrinsic permeability at max. saturation	K	$1.5 \cdot 10^{-12}$	m^2

functions for displacements and linear shape functions for capillary pressure and air pressure. The employed FE-formulation is based on a weak form of the governing balance equations for momentum and mass together with the constitutive relations for the individual phases. The algorithmic treatment of the cap model is based on a return mapping algorithm, which only requires solving a scalar nonlinear equation at the material point level.

Table 5.2 Cap model parameters for the model problem

(Inverse) volumetric stiffness parameter	κ_E	0.015	–
Shear modulus	G	$1 \cdot 10^7$	Pa
Ellipticity parameter for the cap	R	2.25	–
Friction angle	θ	0.3	–
Cohesion parameter	α	0.0	Pa
First shape factor (shape in dev. planes)	ω	0.8	–
Second shape factor (shape in dev. planes)	η	−0.2	–
First hardening law parameter	λ_0	0.1	–
Second hardening law parameter	β	$3.6 \cdot 10^{-5}$	1/Pa
Third hardening law parameter	r	0.2	–
Cohesion-increase parameter	k	1.0	–
First parameter LC curve	X_c	$1 \cdot 10^5$	Pa
Second parameter LC curve	σ^2	0.02	–
Third parameter LC curve	$p^c_{\max}$	$4.4 \cdot 10^5$	Pa
Parameter defining smoothing interval	L	$3.0 \cdot 10^5$	Pa

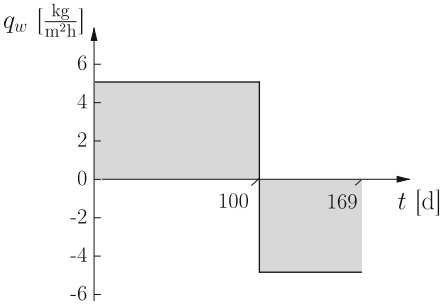

Fig. 5.9 Time-dependent water flux density across the bottom surface of the soil column

In an initial step, preceding the actual simulation, the gravity load is applied for determining the effective stress distribution such that for the given initial capillary (or water) pressure and air pressure fields static equilibrium for the gravity load is obtained. After equilibration, the displacements are reset yielding the final initial state for the subsequent computation. Hence, in addition to the initial stresses, the results at $t = 0$ days include plastic strains and a non-uniform distribution of the hardening parameter, originating from the gravity step. The results are depicted in the leftmost columns of Figs. 5.10, 5.11, 5.12, 5.13, and 5.14.

In a first phase of the computation dewatering of the ground is modelled by a constant water mass outflow across the bottom boundary of $q_w = 0.25\,\mathrm{kg/(m^2h)}$ for 100 days (Fig. 5.9). By this procedure, the ground water table is lowered by 4.6 m. In a second phase, the mass flux is reverted yielding a constant water mass influx of $q_w = 0.25\,\mathrm{kg/(m^2h)}$ for the remaining 69 days, see Fig. 5.9. As a result, the water table in the column is rising again.

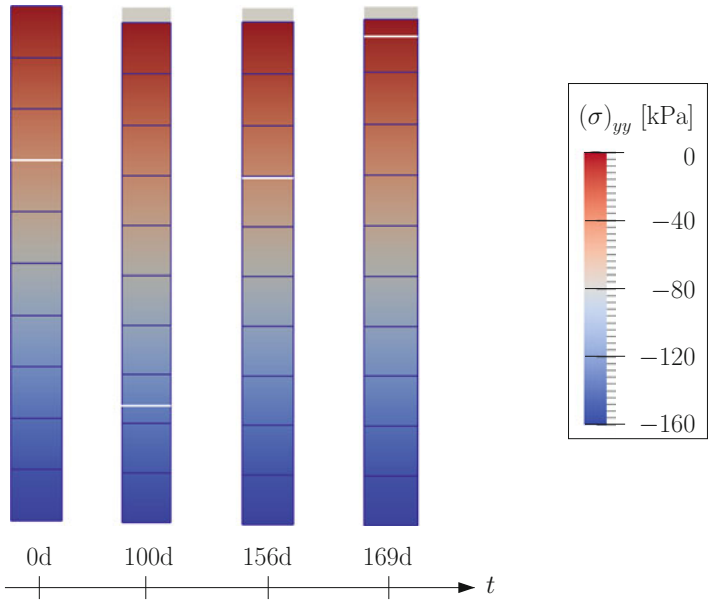

Fig. 5.10 Distribution of the vertical total stress in the soil column at selected time instants

The presented results include distributions of (i) the vertical total stress (Fig. 5.10), (ii) the vertical effective stress (Fig. 5.11), (iii) capillary pressure (Fig. 5.12), (iv) the vertical plastic strain (Fig. 5.13) and (v) the hardening parameter (Fig. 5.14), each (I) at the initial state, (II) at the lowest level of the ground water table, (III) at the re-attained original level of the groundwater table after dewatering and subsequent watering, and (IV) at a significantly higher ground water table than the initial level. In all figures, the results are visualized on the unscaled, deformed configuration with the white line indicating the current position of the groundwater table. It corresponds to the zero isosurface of the capillary pressure, see Fig. 5.12.

The grey shadowed region in the background of the figures indicates the initial shape of the column. By looking at Figs. 5.10, 5.11, 5.12, 5.13, and 5.14 irreversible compaction of the soil column during drawdown of the water table can be observed. At $t = 100$ days larger values of the plastic strains (Fig. 5.13) and of the hardening parameter (Fig. 5.14) are clearly visible compared to the respective initial values.

The computed behaviour is natural by looking at the distribution of the vertical effective stress in Fig. 5.11. If the water table is lowered, then the hydrostatic uplift is reduced, which results in higher (compressive) effective stresses acting on the soil skeleton. The higher compressive effective stresses are associated with plastic material response in the hardening cap mode (Fig. 5.15). By means of the capillary pressure distribution (Fig. 5.12), the effective stress (Fig. 5.11) can be converted to total stress. The vertical total stress is visualized in Fig. 5.10, clearly demonstrating the zero-stress boundary condition at the top of the column.

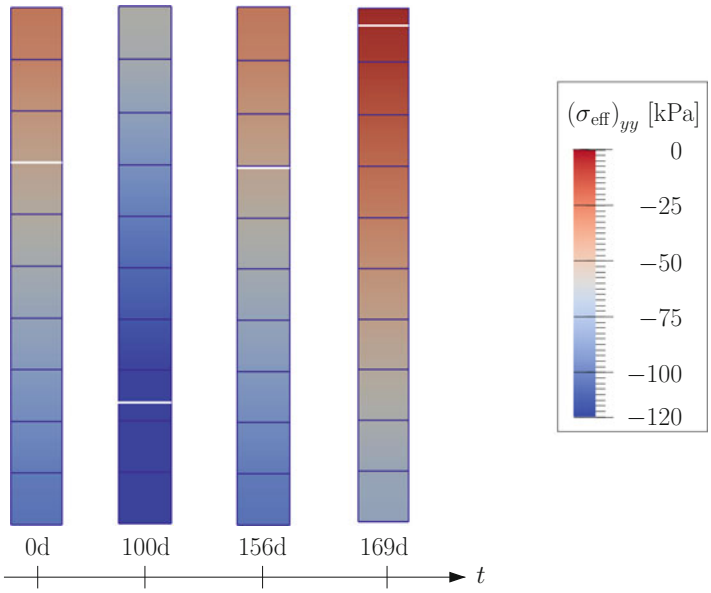

Fig. 5.11 Distribution of the vertical effective stress in the soil column at selected time instants

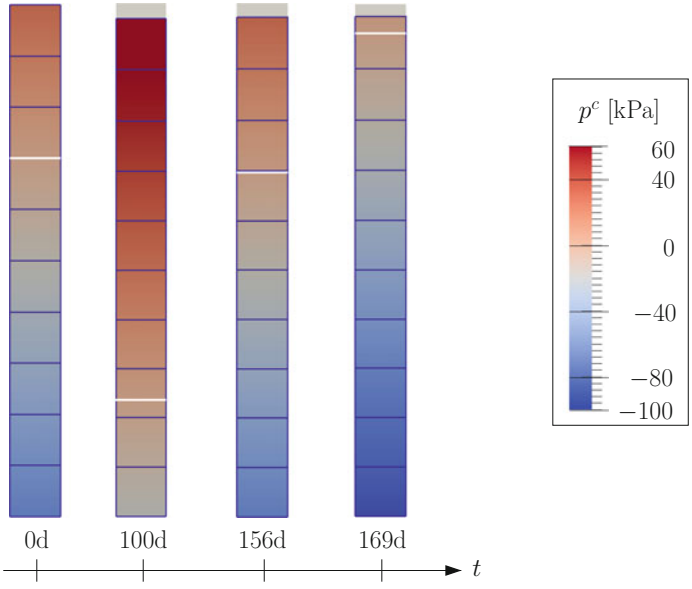

Fig. 5.12 Distribution of the capillary pressure in the soil column at selected time instants (negative values represent water pressure)

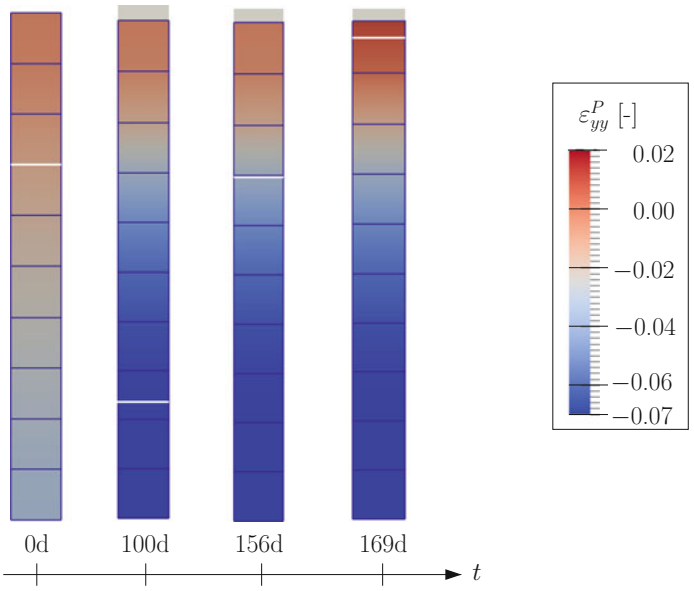

Fig. 5.13 Distribution of the vertical plastic strain in the soil column at selected time instants

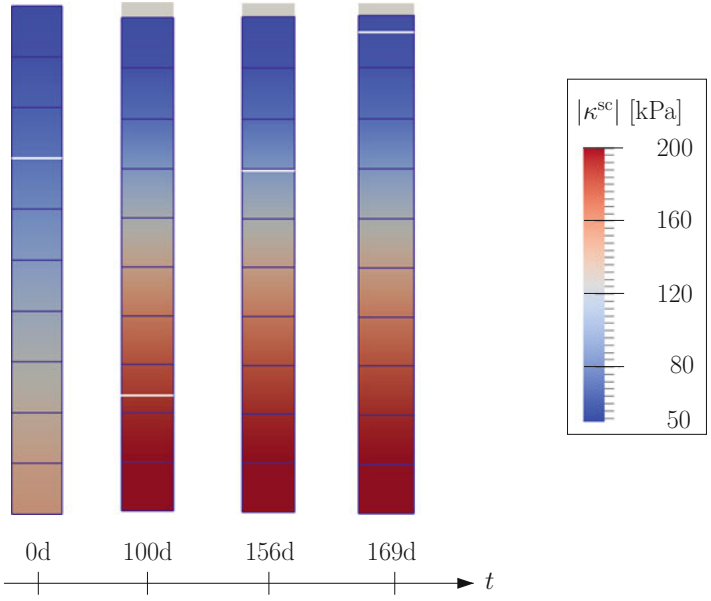

Fig. 5.14 Distribution of the hardening parameter in the soil column at selected time instants

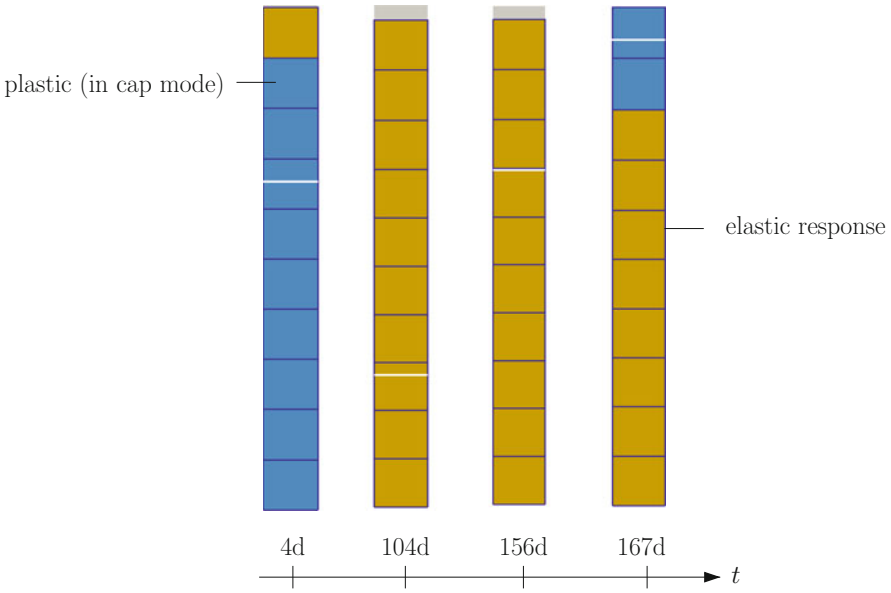

Fig. 5.15 Mode of material response at selected time instants

When rewatering the column up to the initial level, the soil skeleton is unloaded and, thus, the material response during this time period is elastic, as can be seen in the two plots in the middle of Fig. 5.15. Because of irreversible settlements the amount of water required to re-attain the initial water table is smaller than the amount of water that has been extracted before. If the water table is rising beyond the initial level, then wetting-induced plastic deformations can be observed in the region above the initial water table, as can be seen in Fig. 5.15.

5.4.3 Shear Failure of an Embankment Dam

In this section, a failure scenario for an embankment dam will be examined. It concerns slope instability caused by water leakage through the dam. The model problem to be analysed in the following originally was proposed in [11], employing a similar three-phase approach, however, in combination with a single-yield-surface material model, introduced in [10], without considering capillary pressure as a second stress state variable.

The problem is analysed assuming plane strain conditions. The geometry of the dam is depicted in Fig. 5.16. As proposed in [11], the permeability of the dam (material A) is chosen higher than the one of the subsoil (material B). For simplicity, all other three-phase parameters are assumed to be equal for both materials. They are summarized in Tables 5.3 and 5.4.

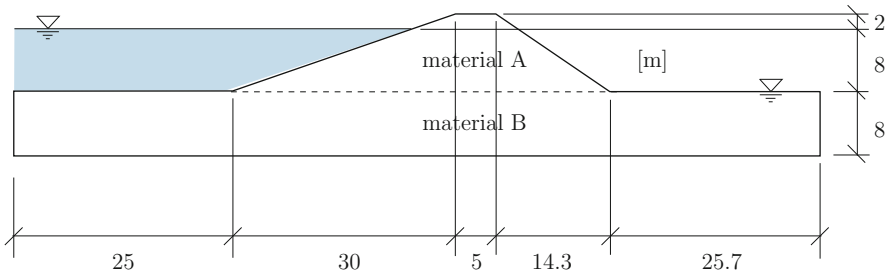

Fig. 5.16 Geometry of the embankment dam

Table 5.3 Parameters of the three-phase formulation for the dam problem

Acceleration of gravity	g	9.81	m/s^2
Temperature	T_0	293.0	K
Atmospheric air pressure	p_{atm}	101,300	Pa
Hydraulic conductivity for water (A)	$(\rho^w g K)/\mu_w$	$1.0 \cdot 10^{-5}$	m/s
Hydraulic conductivity for gas (A)	$(\rho^g g K)/\mu_g$	$1.0 \cdot 10^{-6}$	m/s
Hydraulic conductivity for water (B)	$(\rho^w g K)/\mu_w$	$1.0 \cdot 10^{-8}$	m/s
Hydraulic conductivity for gas (B)	$(\rho^g g K)/\mu_g$	$1.0 \cdot 10^{-9}$	m/s
Compressibility water	C_w	$4.58 \cdot 10^{-10}$	1/Pa
Reference density water	ρ^w	1,000	kg/m^3
Initial porosity	n	0.5	–
Density solid grains	ρ^s	2,700	kg/m^3
Maximum saturation	S_w^s	1.0	–
Residual saturation	S_w^r	0.25	–
Air entry value	p_b^c	$9 \cdot 10^4$	Pa
Van-Genuchten fitting parameter	m	0.55	–

Initially, the reservoir is empty. After 10 days, the water table in the reservoir is assumed to have reached its final position at a level of 8 m. For simplicity, the impoundment is modelled by assuming a linear water pressure distribution between 0 and 8 m above the ground surface and only the amplitude of the respective distribution is increased gradually such that it matches the correct hydrostatic water pressure distribution after 10 days.

Atmospheric air pressure boundary conditions are applied at the upper surface of the structure. Zero capillary pressure boundary conditions are applied at the horizontal ground surface in the downstream region, fixing the water table to the ground surface. Furthermore, water pressure boundary conditions and corresponding surface tractions are prescribed at the upstream face of the dam and at the bottom of the reservoir. Displacements in both directions are fixed at the bottom of the domain. The vertical boundaries of the discretized domain are assumed to be fixed in horizontal direction. The bottom and the left vertical boundary of the domain are assumed to be impermeable. For the right vertical boundary, a hydrostatic water

Table 5.4 Cap model parameters for the dam problem

(Inverse) volumetric stiffness parameter	κ_E	0.011	–
Shear modulus	G	$5.6 \cdot 10^6$	Pa
Ellipticity parameter for the cap	R	3.0	–
Friction angle	θ	0.16	–
Cohesion parameter	α	$2.5 \cdot 10^4$	Pa
First shape factor (shape in dev. planes)	ω	0.6	–
Second shape factor (shape in dev. planes)	η	−0.229	–
First hardening law parameter	λ_0	0.041	–
Second hardening law parameter	β	$1.908 \cdot 10^{-6}$	1/Pa
Third hardening law parameter	r	0.8,390	–
Cohesion-increase parameter	k	0.6	–
First parameter LC curve	X_c	$1.47 \cdot 10^6$	Pa
Second parameter LC curve	σ^2	0.0,679	–
Third parameter LC curve	$p^c_{\max}$	$1.2 \cdot 10^6$	Pa
Parameter defining smoothing interval	L	$3.0 \cdot 10^5$	Pa

pressure distribution is assumed. Since water is assumed leaking through the dam in the course of the computation, for the whole downstream face of the dam a drainage boundary condition is applied. The latter is active only when water is leaking through the dam. In that case it is equivalent to a penalty type flux boundary condition that enforces zero capillary pressure at the respective part of the surface.

The employed computational approach is the same as the one for the ground settlement model problem in the previous subsection. The structured FE-mesh, which is partly shown in Fig. 5.17, consists of 9,156 elements, characterized by quadratic serendipity shape functions for the displacements and linear shape functions for capillary pressure and air pressure with altogether 27,845 nodes and 74,380 unknowns.

Similar to the model problem of the previous section, a gravity initialization step precedes the actual computation. In order to closely match the yield surface used in [11] with the one of the cap model, an initial absolute value of $4.8 \cdot 10^2$ kPa for the hardening parameter is assumed. A comparison of the respective yield surfaces is shown in Fig. 5.18. In Figs. 5.19 and 5.20 the distributions of capillary pressure and excess air pressure (referred to the atmospheric air pressure) are displayed at three selected time instances. Obviously, significant excess air pressure is present only in the initial time period of the numerical simulation. It is caused by the sudden change in boundary conditions. The evolution of the norm of the plastic strain tensor is shown in Fig. 5.21 together with the evolution of the groundwater table in the dam body. Due to the impoundment the water table in the dam body is rising gradually and finally leaking at the toe of the downstream face occurs. The rising water pressure in the dam body causes the effective hydrostatic compressive stress to decrease. Due to the comparably large initial value of the hardening parameter and

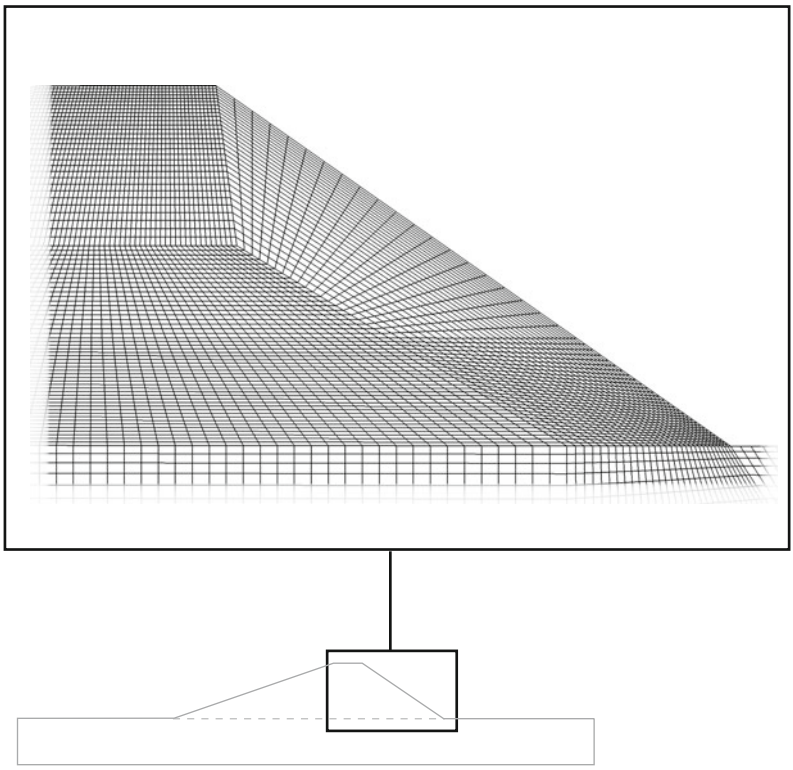

Fig. 5.17 Detail of the structured FE-mesh of the dam body

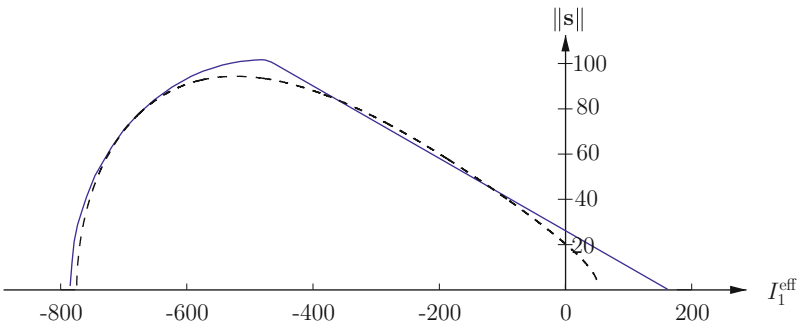

Fig. 5.18 Comparison of the yield surfaces at water saturated conditions for the cap model (*blue, solid line*) and for the soil model used in [11] (*black, dashed line*) [in kPa]

the small value of the friction angle, the shear failure envelope mode is the dominant plastic mode throughout the simulation as shown in Fig. 5.22 for an example step. In Fig. 5.21, the formation of a shear band can be recognized, which is in agreement with results shown in [11]. Similar to what is reported in [11], the formation of

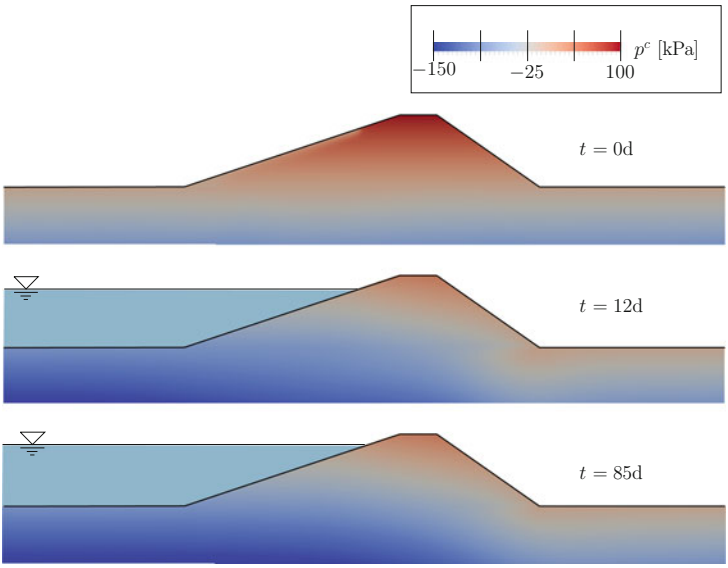

Fig. 5.19 Capillary pressure distributions (negative values represent water pressure)

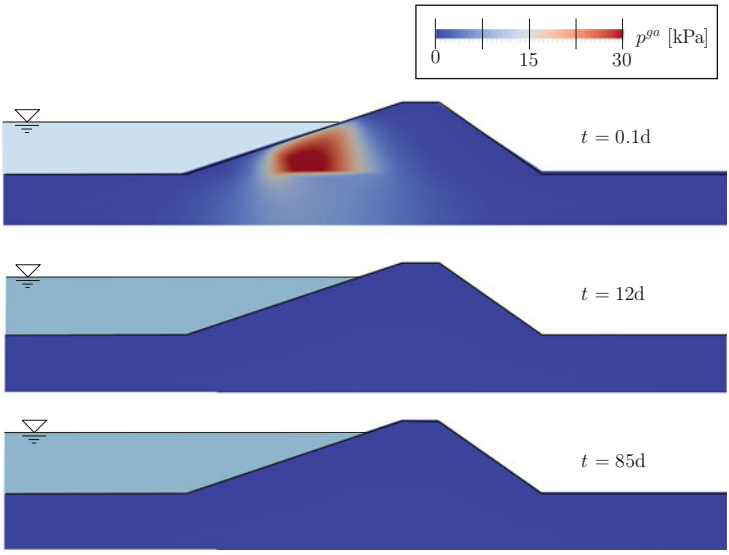

Fig. 5.20 Excess air pressure distributions

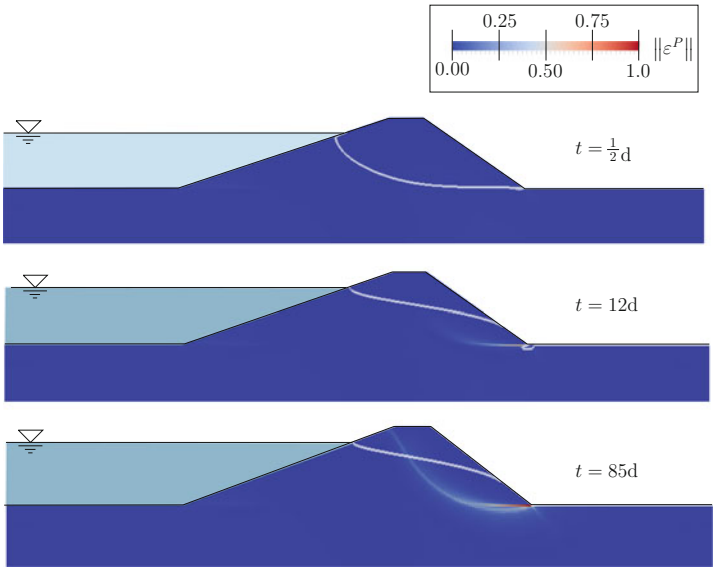

Fig. 5.21 Distribution of the norm of the plastic strain, indicating the formation of a shear band; the *white* curves represent the phreatic surface

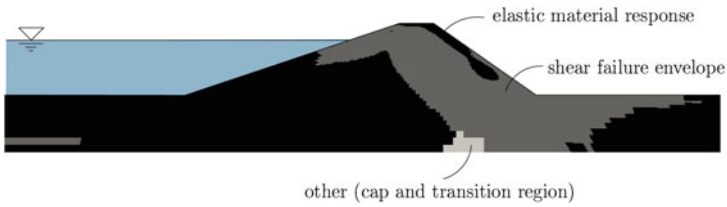

Fig. 5.22 Representative distribution of the active modes of the cap model (the dominant plastic mode is the shear failure envelope mode)

the shear band is initiated at the toe of the dam at the downstream face and is characterized by a curved shape (Fig. 5.23). The plastic deformations are associated with significant displacements, which are visualized in Fig. 5.24.

5.5 Multi-Phase Model for Concrete

For a three-phase model of concrete, a number of additional physical processes, not present in the analysis of the geotechnical problems presented in the previous subsection, have to be taken into account. They include, for instance, changes in material properties due to chemical reactions, i.e. hydration, thermal expansion and

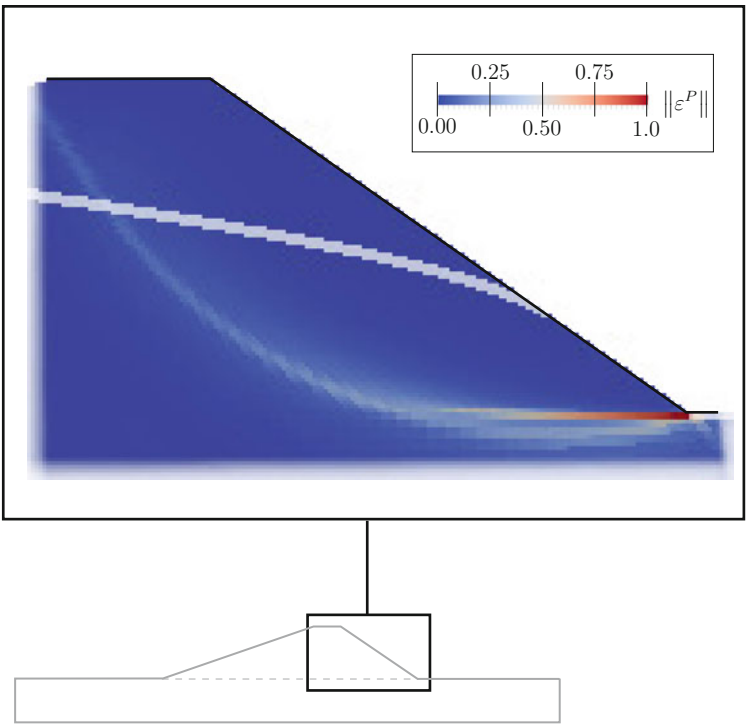

Fig. 5.23 Close-up view of the shear band

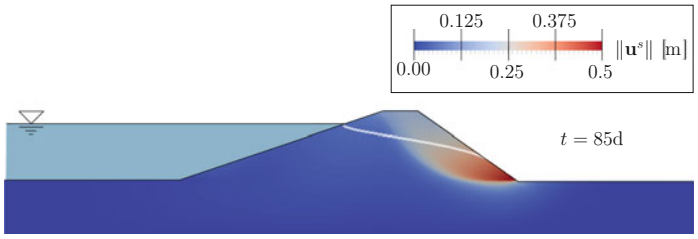

Fig. 5.24 Norm of the final solid displacements indicating slope instability

time dependent behaviour like creep and shrinkage. In order to account for these physical processes, it is important to include temperature as an additional state variable in the analysis, to view the gas phase as a mixture of dry air and water vapour and to include mass exchange terms in the balance laws for the solid phase, the water phase and the vapour phase.

5.5.1 Constitutive Law for Concrete

The constitutive law, relating effective stresses and strains, is stated in rate form as

$$\left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} \boldsymbol{\sigma}_{\text{eff}} = \mathbf{C}\left(\left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} \boldsymbol{\varepsilon} - \left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} \boldsymbol{\varepsilon}^T - \left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} \boldsymbol{\varepsilon}^{CH} - \left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} \boldsymbol{\varepsilon}^C - \left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} \boldsymbol{\varepsilon}^P\right) + \\ + \left(\left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} \mathbf{C}\right)\left(\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^T - \boldsymbol{\varepsilon}^{CH} - \boldsymbol{\varepsilon}^C - \boldsymbol{\varepsilon}^P\right). \tag{5.70}$$

Newly introduced strains are the volumetric strains $\boldsymbol{\varepsilon}^T$ and $\boldsymbol{\varepsilon}^{CH}$ related to temperature variations and chemical reactions during hydration (autogenous shrinkage strains), respectively, and the creep strains $\boldsymbol{\varepsilon}^C$. Although not considered in the numerical example below, plastic strains $\boldsymbol{\varepsilon}^P$ are also included in (5.70) for completeness. Furthermore, changes in the elastic properties due to hydration are accounted for via the term $\left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} \mathbf{C}$.

For the thermal strains, the linear relationship

$$\left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} \boldsymbol{\varepsilon}^T = \alpha_T \left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} T\,\mathbf{1} \tag{5.71}$$

with α_T as the thermal expansion coefficient is employed. Similarly, expansion due to hydration is governed by

$$\left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} \boldsymbol{\varepsilon}^{CH} = \alpha_{CH}\left(\left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} \Gamma_{\text{hydr}}\right)\mathbf{1} \tag{5.72}$$

with Γ_{hydr} and α_{CH} as the degree of hydration and the expansion coefficient due to hydration, respectively.

The reaction law for the degree of hydration is given as [16]

$$\left.\frac{\partial}{\partial t}\right|_{\mathbf{x}_s} \Gamma_{\text{hydr}} = \frac{\tilde{A}_\Gamma\left(\Gamma_{\text{hydr}}\right)}{1 + (a - a\varphi)^4} \exp\left(-\frac{E_a}{RT}\right) \tag{5.73}$$

with the empirical parameter a and the hydration activation energy E_a. As in [15,16] the normalized chemical affinity

$$\tilde{A}_\Gamma\left(\Gamma_{\text{hydr}}\right) = A_1\left(\frac{A_2}{\kappa_\infty} + \kappa_\infty \Gamma_{\text{hydr}}\right)\left(1 - \Gamma_{\text{hydr}}\right)\exp\left(-\bar{\eta}\Gamma_{\text{hydr}}\right) \tag{5.74}$$

is adopted from [6]. It exclusively depends on Γ_{hydr} and the constant parameters κ_∞, A_1, A_2, $\bar{\eta}$. The rate of the degree of hydration (5.73) furthermore depends on the relative humidity $\varphi = p^{gw}/p^{gw,\text{sat}}$ and temperature.

Creep strains $\boldsymbol{\varepsilon}^C$ are modelled according to the microprestress solidification theory by Bazant [2], see also [16].

Assuming a passive gas phase, i.e., $p^g = p_{\text{atm}}$, the relation between total and effective stress (5.26) together with (5.25) is simplified to

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}_{\text{eff}} + \alpha_{\text{Biot}}\, \chi_{\text{Bishop}}\,(S_w)\, p^c \mathbf{1}\,. \tag{5.75}$$

Since drying causes an increase of the capillary pressure, from (5.75) together with the constitutive relations a volumetric compaction of concrete is obtained. Hence, departing from a fully saturated state, characterized by $p^c = 0$, the drying shrinkage strain at a partially saturated state with $p^c \neq 0$ is obtained by means of the bulk modulus K of the solid skeleton as

$$\boldsymbol{\varepsilon}^{SH} = -\frac{\alpha_{\text{Biot}}\, \chi_{\text{Bishop}}\,(S_w)}{3K}\, p^c\, \mathbf{1}\,. \tag{5.76}$$

The combined parameter $\alpha_{\text{Biot}}\, \chi_{\text{Bishop}}\,(S_w)$ can be determined from measurement data for the ultimate isothermal drying shrinkage strain in fully matured concrete samples, which are dried departing from full saturation, i.e. $p^c = 0$, to lower levels of saturation, corresponding to $p^c > 0$. Thus, drying shrinkage strains are not treated explicitly in the material law but are rather a consequence of the (constitutive) effective stress assumption.

Aging elasticity is taken into account based on the effective age or maturity of concrete as it was defined in [3],

$$\left.\frac{\partial}{\partial t}\right|_{\mathbf{X}_s} t_{\text{mat}} = \frac{1}{1 + (a - a\varphi)^4} \exp\left(\frac{E_a}{R}\left(\frac{1}{T_0} - \frac{1}{T}\right)\right). \tag{5.77}$$

According to [21], the current value of the modulus of elasticity is then computed as

$$E\,(t_{\text{mat}}) = E \cdot \exp\left(0.3 \cdot \hat{s}\left[1 - \left(\frac{28 \cdot 24 \cdot 3600\,\text{s}}{t_{\text{mat}}}\right)\right]\right). \tag{5.78}$$

Here, $\hat{s}$ is a material parameter and s denotes time in seconds. Poisson's ratio is assumed to remain constant.

5.5.2 *Numerical Simulation of the Behaviour of Concrete Overlays*

A frequently employed method for strengthening existing RC structures is to add a concrete overlay. The behaviour of the overlay and, hence, of the strengthened structure is affected by drying shrinkage, in particular, by the shrinkage strains of

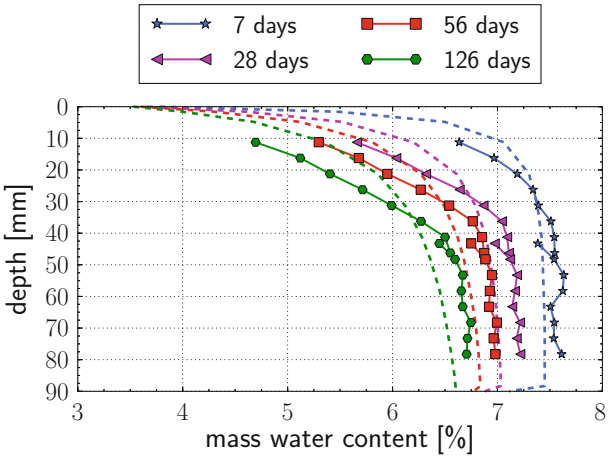

Fig. 5.25 Measured (*continuous lines*) and computed mass water content (*dashed lines*) in the concrete overlay at selected ages of the overlay

the overlay developing during the first days and weeks [4], which are restrained by the substrate concrete.

For investigating the overlay behaviour due to drying shrinkage, a laboratory test program was conducted [25]. A 90 mm thick concrete overlay, made of normal strength concrete, was added to a prismatic concrete specimen with dimensions of $800 \times 300 \times 300$ mm. Before placing of the concrete overlay, the substrate concrete, i.e., the top surface of the prismatic concrete specimen, was prepared by high-pressure water jetting. It resulted in a sharp increase of the mass water content in the superficial zone of the specimen. Subsequently, the overlay was placed, the lateral surfaces were sealed and its top surface was exposed to a relative humidity of 65 %. The mass water content distributions in the overlay were measured during drying. They were determined by a calibrating curve for the respective concrete, relating electrolytic resistances, measured by Multi-Ring-Sensors, to the mass water content.

The measured mass water content distributions in the concrete overlay at selected time instants are displayed by the continuous lines in Fig. 5.25. The measured longitudinal strain (i.e., the strain in the direction of the longest edge of the specimen) at the top surface of the overlay is depicted by the continuous line in Fig. 5.26.

The hygric and also the drying shrinkage properties of the overlay concrete were obtained from tests on thin concrete slices with dimensions of $110 \times 110 \times 6$ mm made of the same concrete mixture. After moist curing for 81 days the concrete slices were exposed to drying at different values of relative humidity between 100 and 50 %, and the mass water content and the ultimate drying shrinkage strains were determined. From the measurement data, the intrinsic permeability K in (5.29) and (5.30), the structure coefficient f_S in (5.39), the air entry value p_b^c and the fitting parameter m in (5.15) are obtained as $K = 3.0 \cdot 10^{-21}\,\mathrm{m}^2$, $f_S = n^2(1 - S_w)^{9/2}$,

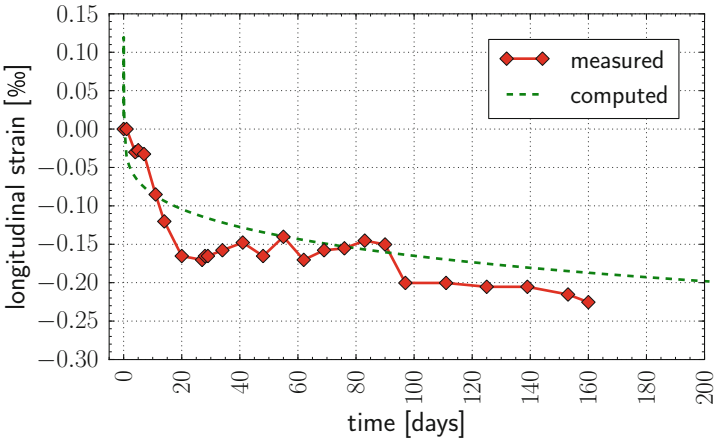

Fig. 5.26 Measured (*continuous line*) and computed evolution (*dashed line*) of the longitudinal strain at the top surface of the concrete overlay

$p_b^c = 13,362\,\text{MPa}$ and $m = 0.313$, respectively. Detailed information is provided in [26]. The combined parameter $\alpha_{\text{Biot}}\,\chi_{\text{Bishop}}\,(S_w)$ in (5.76) can be obtained from a measured relationship between drying shrinkage strain and relative humidity. For instance, in [26], a linear function of the degree of water saturation is determined based on a least-squares fit to data measured at four levels of relative humidity:

$$\alpha_{\text{Biot}}\,\chi_{\text{Bishop}}\,(S_w) = 0.9765 S_w - 0.1086\,. \tag{5.79}$$

According to (5.75) in the multi-phase concrete model, proposed in [16], capillary pressure evolving during drying produces a hydrostatic effective compressive stress acting on the solid matrix. The latter stress, in turn, causes creep strains. However, in the example of this section, these creep strains are not yet considered.

With the material parameters at hand, a numerical simulation of the described lab test is performed. The numerical model of one quarter of the prismatic specimen, representing the substrate concrete and the overlay, consists of 2,856 3D finite elements with quadratic interpolation of the displacements and linear interpolation of temperature and capillary pressure.

In the numerical simulation the following steps are considered: (1) drying of the substrate concrete, (2) high-pressure water pressure jetting of the top surface of the substrate concrete, (3) placement of the concrete overlay and hardening of the latter, exposing its top surface to a relative humidity of 65 % after 12 h of moist curing.

The computed distributions of the mass water content in the overlay are shown at selected time instants during drying by the dashed lines in Fig. 5.25. The computed evolution of the longitudinal strain at the top surface of the concrete overlay is depicted by the dashed line in Fig. 5.26. The computed positive values of the longitudinal strain during the first few days are caused by the temperature increase

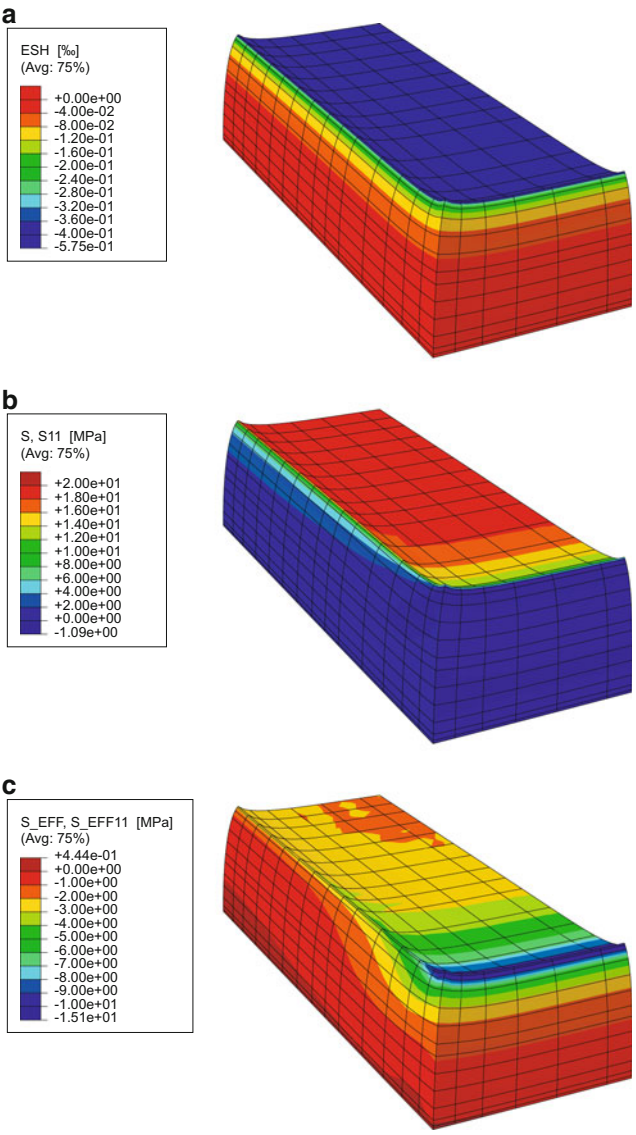

Fig. 5.27 Predicted response of the overlay after 2 weeks of hardening and drying: (**a**) drying shrinkage strains, (**b**) total and (**c**) effective stresses in longitudinal direction

due to hydration. Since the measurements were started about 1 day after casting, the subsequent strain measurements are referred to this time instant and, hence, at least part of the thermal strain is not included in the measured strain.

The deformations of the overlay (magnified by a factor of 1,000) after 2 weeks of hardening and drying are displayed in Fig. 5.27. The shrinkage strains of the overlay,

predicted according to (5.76), are depicted in Fig. 5.27a, and the respective distributions of the total stress and the effective stress, acting in longitudinal direction of the overlay, are shown in Fig. 5.27b,c. From the latter, the difference between total and effective stresses can be interpreted. Whereas at the top surface the predicted total longitudinal stresses are tensile stresses because of the restrained deformations due to drying shrinkage, the respective effective stresses are compressive stresses, which follows from (5.75) by the acting capillary pressure. Not shown in Fig. 5.27 are the predicted effective tensile stresses along the outer boundary of the interface between old and new concrete in the direction normal to the interface.

5.6 Summary and Conclusions

In this contribution, applications of multi-phase models in geotechnical engineering and concrete engineering were presented. In both cases, the ability to account for the presence of water and air inside porous solids and the capability to account for coupling effects between the different phases significantly enhanced the spectrum of mechanical responses that could be represented. They include, for instance, wetting induced ground settlements in geotechnical engineering as well as drying shrinkage in concrete engineering. Despite of the common three-phase nature of the applications presented, problem specific challenges can be observed. For instance, the hydration process of concrete at early ages makes consideration of temperature changes and of the vapour phase indispensable. Furthermore, in the presented framework collapse upon wetting for soils can only be modelled if an appropriate, capillary pressure dependent, material model is used, e.g. the improved elastic-plastic cap model described in Sect. 5.4.1 being a suitable option. Even common parameters for both applications like the intrinsic permeability may vary orders of magnitude between concrete and soils or between different soil types.

Future research will include work on efficient and robust algorithms for the solution of the coupled three-phase problem. In particular, this involves solution strategies at three different levels: (1) the material point level where robust and efficient return-mapping algorithms are required, (2) the nonlinear solution process level for the coupled system of governing equations for which globalization strategies that ensure convergence in case of not ideally chosen initial values are to be explored and (3) the linear solution level where possible benefits of iterative solvers and preconditioners have to be investigated. Furthermore, a more thorough investigation of creep for concrete overlays will be performed, allowing better predictions of drying-induced cracking in structures strengthened by concrete overlays.

References

1. Alonso, E.E, Gens, A., Josa, A.: A constitutive model for partially saturated soils. Géotechnique **40**, 405–430 (1990)
2. Bazant, Z.P., Hauggaard, A., Baweja, S., Ulm, F.: Microprestress-solidification theory for concrete creep. I: aging and drying effects. J. Eng. Mech. (ASCE) **123**, 1188–1194 (1997)
3. Bazant, Z.P., Cusatis, G., Cedolin, L.: Temperature effect on concrete creep modeled by microprestress-solidification theory. J. Eng. Mech. (ASCE) **130**, 691–699 (2004)
4. Beushausen, H.: Failure mechanisms and tensile relaxation of bonded concrete overlays subjected to differential shrinkage. Cement Concrete Res. **36**, 1908–1914 (2006)
5. Bucio, M.B.: Estudio experimental del comportamiento hidro-mecanico de suelos colapsables. Ph.D thesis, Universitat Politecnica de Catalunya (2002)
6. Cervera, M., Olivier, J., Prato, T.: A thermo-chemo-mechanical model for concrete. I: hydration and aging. J. Eng. Mech. (ASCE) **125**, 1018–1027 (1999)
7. De Borst, R., Groen, A.E.: Computational strategies for standard soil plasticity models. In: Zaman, M., Booker, J. (eds.) GiodaModeling in Geomechanics, pp. 23–50. Wiley, Chichester (2000)
8. Di Maggio, F.L., Sandler, I.S.: Material model for granular soils. J. Eng. Mech. Div. (ASCE) **97**, 935–950 (1971)
9. Dolarevic, S., Ibrahimbegovic, A.: A modified three-surface elasto-plastic cap model and its numerical implementation. Comput. Struct. **85**, 419–430 (2007)
10. Ehlers, W.: A single-surface yield function for geomaterials. Arch. Appl. Mech. **65**, 246–259 (1995)
11. Ehlers, W., Graf, T., Ammann, M.: Deformation and localization analysis of partially saturated soil. Comput. Methods Appl. Mech. Eng. **193**, 2885–2910 (2004)
12. Gamnitzer, P., Hofstetter, G.: A cap model for soils featuring a smooth transition from partially to fully saturated state. Proc. Appl. Math. Mech. **13**, 169–170 (2013)
13. Gamnitzer, P., Hofstetter, G.: An improved cap model for partially saturated soils. In: ASCE Conference Proceedings of the Biot Conference on Poromechanics V, pp. 569–578, Vienna (2013)
14. Gawin, D., Majorana, C.E., Schrefler, B.A.: Numerical analysis of hygro-thermal behaviour and damage of concrete at high temperature. Mech. Cohesive Frictional Mater. **4**, 37–74 (1999)
15. Gawin, D., Pesavento, F., Schrefler, B.A.: Hygro-thermo-chemo-mechanical modelling of concrete at early ages and beyond. Part I: hydration and hygro-thermal phenomena. Int. J. Numerical Methods Eng. **67**, 299–331(2006)
16. Gawin, D., Pesavento, F., Schrefler, B.A.: Hygro-thermo-chemo-mechanical modelling of concrete at early ages and beyond. Part II: shrinkage and creep of concrete. Int. J. Numerical Methods Eng. **67**, 332–363 (2006)
17. Hochgürtel, T.: Numerische Untersuchungen zur Beurteilung der Standsicherheit der Ortsbrust beim Einsatz von Druckluft zur Wasserhaltung im schildvorgetriebenen Tunnelbau. Dissertation, RWTH Aachen (1998)
18. Kohler, R., Hofstetter, G.: A cap model for partially saturated soils. Int. J. Numerical Anal. Methods Geomech. **32**, 981–1004 (2008)
19. Lewis, R.D., Schrefler, B.A.: The Finite Element Method in the Static and Dynamic Deformation and Consolidation of Porous Media. Wiley, New York (1998)
20. Macari, E.J., Hoyos, L.R., Arduino, P.: Constitutive modeling of unsaturated soil behavior under axisymmetric stress states using a stress/suction-controlled cubical test cell. Int. J. Plast. **19**, 1481–1515 (2003)
21. ÖNORM EN 1992-1-1: Eurocode 2: Design of Concrete Structures - Part 1-1: General Rules and Rules for Buildings. Austrian Standards Institute, Vienna (2011)
22. Parker, J.C. Multiphase flow and transport in porous media. Rev. Geophys. **27**, 311–328 (1989)
23. Pertl, M.: Grundlagen, Implementierung und Anwendung eines Drei-Phasen Modells für Böden. Dissertation, Universität Innsbruck (2010)

24. Schrefler, B.A.: Mechanics and thermodynamics of saturated/unsaturated porous materials and quantitative solutions. Appl. Mech. Rev. **55**, 351–388 (2002)
25. Theiner, Y., Hofstetter, G.: Evaluation of the effects of drying shrinkage on the behaviour of concrete structures strengthened by overlays. Cement Concrete Res. **42**, 1286–1297 (2012)
26. Valentini, B., Theiner, Y., Aschaber, M., Lehar, H., Hofstetter, G.: Single-phase and multi-phase modeling of concrete structures. Eng. Struct. **47**, 25–34 (2013)
27. 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)

Chapter 6
Concrete Structures Subjected to Fire Loading: From Thermo-Mechanical Modeling of Strain Behavior of Concrete Towards Structural Safety Assessment

T. Ring, M. Zeiml, and R. Lackner

Abstract In this chapter, results obtained within a 4-year research project on the safety of underground structures subjected to fire loading are presented. For this project, a consortium consisting of three scientific partners (Vienna University of Technology, University of Innsbruck, University of Natural Resources and Life Sciences, Vienna) and eight industrial partners (ÖBB-Infrastruktur AG, ASFINAG, Wiener Linien, Arge Bautech, VÖZFI, Büro Dr. Lindlbauer, Schimetta Consult, ZT Reissmann) was established. Whereas the mentioned research project followed a holistic approach, covering simulation of the fire event, experimental investigation of concrete and concrete structures at high temperatures, and modeling and simulation work at both the material and the structural scale (Amouzandeh, Development and application of a computational fluid dynamics code to predict the thermal impact of underground structures in case of fire, Ph.D. thesis, Vienna University of Technology, Vienna, 2012; Ring et al. Brandversuche zum Abplatz- und Strukturverhalten von Tunnel mit Rechtecksquerschnitt [Fire experiments investigating the spalling and structural behavior of rectangular tunnels], Technical Report, Vienna University of Technology and Vereinigung der österreichischen Zementindustrie (VÖZFI), Vienna, 2012; Ring, Experimental characterization and modeling of concrete at high temperatures: Structural safety assessment of different tunnel cross-sections subjected to fire loading, Ph.D. thesis, Vienna University of Technology, Vienna, 2012; Zhang, Simulations for durability assessment of concrete structures: multifield

T. Ring (✉) • M. Zeiml
Institute for Mechanics of Materials and Structures, Vienna University of Technology, Karlsplatz 13, A1040 Vienna, Austria
e-mail: Thomas.Ring@tuwien.ac.at; Matthias.Zeiml@tuwien.ac.at

R. Lackner
Unit of Material Technology, University of Innsbruck, Technikerstr. 13, A6020 Innsbruck, Austria
e-mail: Roman.Lackner@uibk.ac.at

G. Hofstetter (ed.), *Computational Engineering*, DOI 10.1007/978-3-319-05933-4_6,
© Springer International Publishing Switzerland 2014

framework and strong discontinuity embedded approach, Ph.D. thesis, Vienna University of Technology, Vienna, 2013), this chapter focuses on one aspect of the project, namely modeling and simulation of the behavior of concrete and concrete structures under combined thermal and mechanical loading:

1. First, a micromechanical model taking the composite nature of concrete into account is presented. Based on experimental results obtained for cement paste and aggregate subjected to thermal/mechanical loading, a two-scale model formulated within the framework of continuum micromechanics is developed, giving access to the effective elastic and thermal-dilation properties of concrete as a function of temperature.
2. In a second step, these model-based properties are considered within a differential formulation of the underlying stress–strain law, accounting for the influence of mechanical loading on the thermal-strain evolution. The proposed micromechanical approach and its implementation are validated by experimental results obtained from concrete specimens subjected to combined thermo-mechanical loading.
3. Finally, the effect of the underlying model assumptions at the structural scale is illustrated by means of the safety assessment of underground support structures under fire attack.

The obtained results are nowadays considered in the formulation of standards and guidelines for the assessment of the safety of underground structures subjected to fire loading (ÖBV-Richtlinie: Fire protection with concrete for underground traffic infrastructure [Erhöhter baulicher Brandschutz mit Beton für unterirdische Verkehrsbauwerke], Austrian Society for Construction Technology, Vienna, 2013).

6.1 Introduction

Concrete subjected to combined mechanical and thermal loading exhibits a certain path dependence explained by the dependence of physical processes on the actual stress state within the material (see [3, 8, 10, 18, 25, 26, 28]). This path dependence of heated concrete (highlighted in Fig. 6.1) is often related to the introduction of so-called load induced thermal strains (LITS).

The main findings reported in the literature with respect to LITS are:

1. LITS are found only in concrete subjected to first thermal loading [12].
2. The rate of heating (ranging from 0.2 to 5 °C/min) and the water/cement ratio showed only minor influence on LITS [25].
3. The aggregate type has no significant influence in the development of LITS, linking LITS to processes taking place within the cement paste [12].
4. LITS are practically unaffected by the type of cement blend, suggesting that it takes place in a common gel or C-S-H structure [12].
5. LITS seem to increase linearly with the applied stress level (see Fig. 6.2) [3].

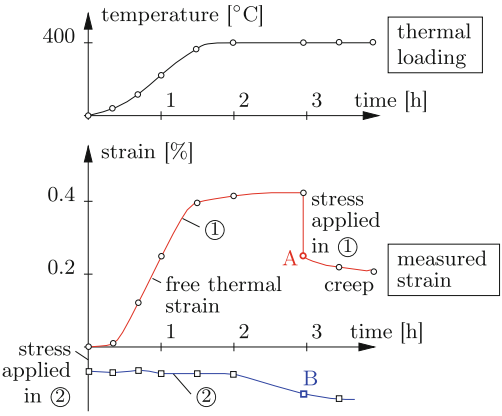

Fig. 6.1 Path dependence of combined mechanical and thermal loading according to [27]; the application of the same thermal and mechanical loading applied in different order leads to the same temperature and stress level ($T_{max} = 400\,°C$ and $\sigma = 0.45 \cdot f_{c,0}$) but to different experimentally observed strains—compare points A and B

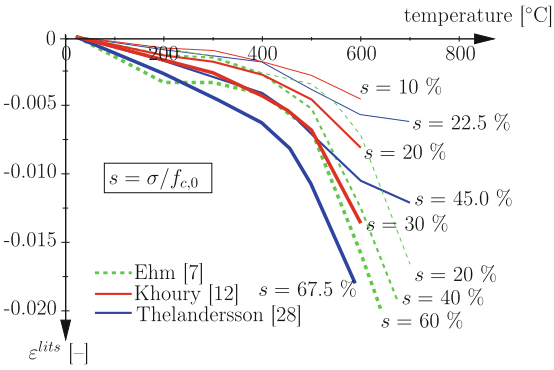

Fig. 6.2 Load dependency of LITS obtained from different experiments [8, 12, 27] ($s = \sigma/f_{c,0}$: level of loading)

Based on these experimental findings, several formulations for LITS can be found in the open literature, ranging from an approach to model LITS within creep of heated concrete [25] over considering LITS via empirical relations [26], to strain-rate formulations for LITS as proposed in [18, 27].

In recent years, micromechanics-based models for concrete have been published in the open literature (e.g., [6, 16]) taking the composite nature of concrete into account. On the one hand, these models were developed in order to identify the behavior of C-S-H at elevated temperatures [6] using nanoindentation. On the other hand, a multiscale model for the determination of the effective stiffness of concrete at high temperatures was proposed in [16].

In the present work, recently published micromechanics-based models [14, 21] are adopted to the description of the change of elastic properties and the thermal dilation of heated concrete. For this purpose, experimental studies on concrete and cement-paste specimens were conducted, and the respective experimental results are presented in Sect. 6.2. In Sect. 6.3, the micromechanic-based model is presented with possible modes of implementation of the underlying stress–strain behavior which is discussed in Sect. 6.4. The so-obtained formulations for the consideration of the combined thermo-mechanical behavior of concrete and their effect within the

analysis of concrete structures subjected to fire loading are highlighted in Sect. 6.5. Concluding remarks are given in Sect. 6.6.

6.2 Experimental Observation

In addition to the existing experimental data available in the open literature, experiments were performed in order to assess the combined thermo-mechanical behavior as well as the elastic properties of concrete under temperature loading. The experiments were conducted in a radiant electric oven which is used to apply the thermal loading (see Fig. 6.3). The cylindrical oven is built around the mechanical testing device, allowing to perform tests under combined thermal and mechanical loading. The cylindrical specimens had a dimension of 100 mm in diameter and a height of 200 mm. In order to monitor the deformations of the heated specimen, steel rings are mounted with steel bars transferring the deformation of the specimen to the outside of the oven (axial direction). In the radial direction, steel bars, directly pointing from the specimen to the outside of the oven, give access to the radial deformation. In the course of the experiments, the specimens are subjected to constant uniaxial loading and heated up to 800 °C with a heating rate of 1 °C/min.

In order to identify the elastic properties of the heated specimens, additional test with modulated mechanical and steadily increasing thermal load was considered within the test program (see [22] for details).

6.2.1 Deformation Under Thermo-Mechanical Loading

In Fig. 6.4, the evolution of strain in axial direction for cement paste[1] subjected to different levels of mechanical loading ($s = 100 \cdot \sigma_a / f_{c,0} = 0$, 5, 10, 20, and 30 %, where $f_{c,0} = 42.6$ MPa) is presented, indicating the load dependency of deformations in case of increasing temperature loading.

With increasing load level, the compaction of cement paste in axial direction and the expansion in radial direction increase, especially at higher temperatures. While the behavior of strains below 500 °C is mainly driven by the degradation of C-S-H- and C-H-phases (see [7, 29]), at temperatures between 500 and 600 °C an abrupt change of the evolution of strain is observed for $s > 5$ %, which is attributed to the development of macro-cracks in longitudinal direction of the cement-paste specimen.

Figure 6.5 shows the evolution of strain in axial and radial direction for concrete specimens under combined mechanical ($s = 100 \cdot \sigma_a / f_{c,0} = 0$, 10, 20, 30, 40, 50, and 60 %, where $f_{c,0} = 39.1$ MPa) and thermal loading. The observed change

[1] For details on the underlying mix-design, the reader is referred to [22]

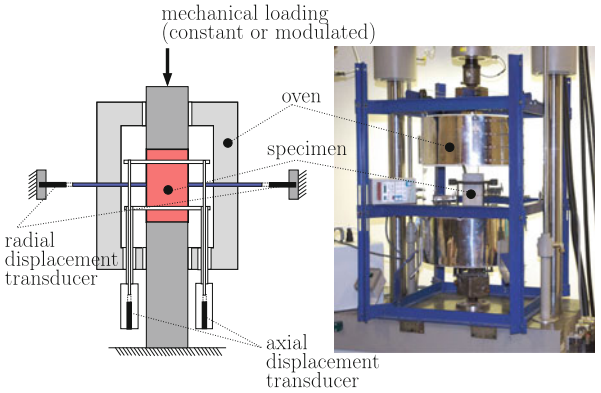

Fig. 6.3 Used device for thermo-mechanical testing (see [22] for details)

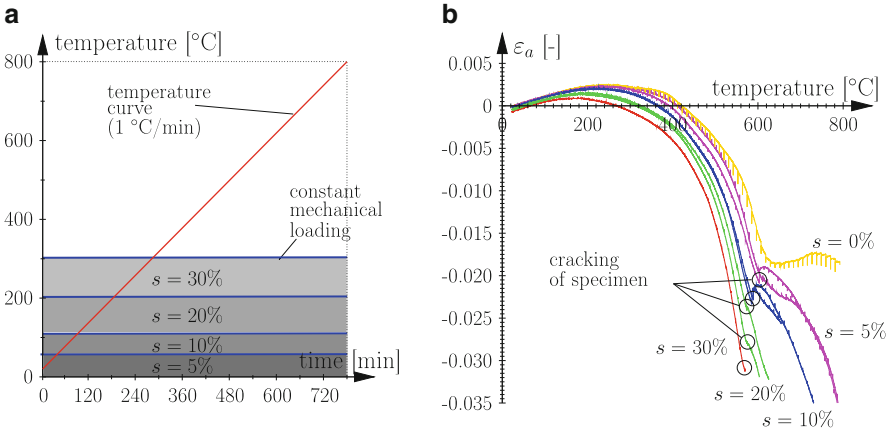

Fig. 6.4 Cement paste: (**a**) temperature curve (1 °C/min) and mechanical loading ($s = 100 \cdot \sigma_a / f_{c,0} = 0$ to 30 %, with initial compressive strength $f_{c,0} = 42.6$ MPa); (**b**) evolution of axial strain as a function of temperature [22]

in the strain evolution between 550 and 620 °C results from the quartz transition at 573 °C. The evolution of axial strain presented in Fig. 6.5b decreases with increasing mechanical loading. At higher load levels ($s \geq 40$ %), the concrete specimens fail before the final temperature of 800 °C is reached.

6.2.2 *Behavior of Siliceous Material*

Since concrete with a high content of siliceous aggregates (89 % in total, consisting of 68 % quartz and 21 % feldspar) was investigated in the previous subsection, the thermal strain behavior as well as the evolution of elastic properties is included in

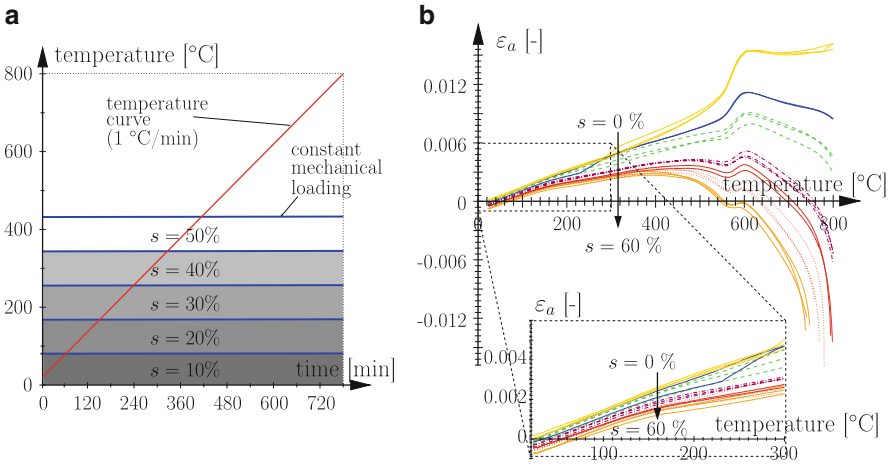

Fig. 6.5 Concrete: (**a**) temperature curve (1 °C/min) and mechanical loading ($s = 100 \cdot \sigma_a / f_{c,0} =$ 0 to 60 %, with the initial compressive strength $f_{c,0} = 39.1$ MPa); (**b**) evolution of axial strain as a function of temperature [22]

Sect. 6.3, with the respective experimental results taken from [11, 15]. The thermal-strain evolution of quartz reported in [11] is shown in Fig. 6.9, indicating quartz transition at 573 °C. For $T > 573$ °C, the evolution of the thermal strain exhibits a plateau at 1.72 %. The elastic properties of quartz (Young's modulus and Poisson's ratio) as a function of temperature were determined in [15] using ultrasonic tests. Both free-thermal strain and elastic properties of quartz will be essential in the following section dealing with the micromechanical modeling of concrete behavior under combined thermo-mechanical loading.

6.3 Micromechanical Model

In order to capture the influence of the constituents of concrete on the overall behavior, a micromechanical model is proposed, consisting of aggregates, cement paste, and pore space (see Fig. 6.6). Hereby, one portion of the air voids is already contained within the cement paste, while an additional portion of air voids is introduced by the mixing process of aggregates and cement paste (see Fig. 6.7a).

Accordingly, the proposed micromechanical model comprises two scales in addition to the macroscale:

- At Scale I, cement-paste composite (pore space, hydration products) and additional pores introduced during the mixing process build up the material microstructure. At this scale, the experimentally determined behavior for cement paste (see [22], for details) is considered.
- At Scale II, the aggregate phase is employed into the homogenized material of Scale I.

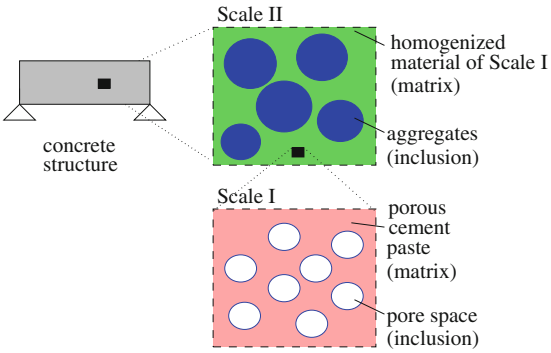

Fig. 6.6 Micromechanical model of concrete

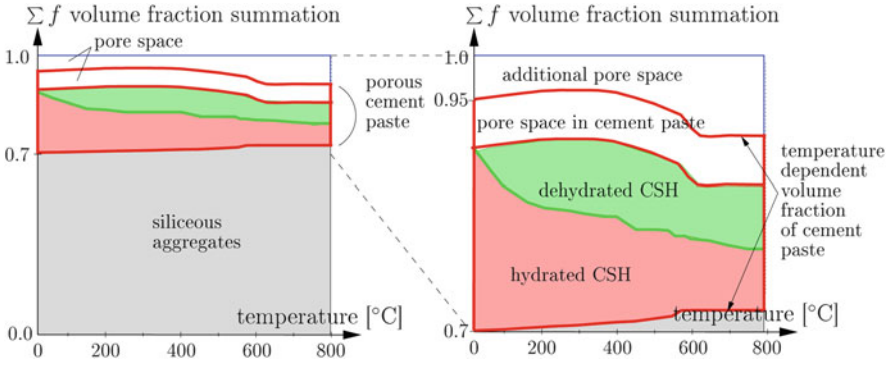

Fig. 6.7 Evolution of volume fractions for heated concrete

Within this micromechanical framework, both the effective elastic and thermal-dilation properties of heated concrete are determined using continuum micromechanics (based on Mori Tanaka [17], applied in [20]).

6.3.1 *Effective Elastic Properties*

The effective shear and bulk modulus, G_{eff} and K_{eff}, are given as:

$$G_{\text{eff}} = \frac{\sum_r f_r G_r \left[1 + \beta \left(\frac{G_r}{G_m} - 1\right)\right]^{-1}}{\sum_r f_r \left[1 + \beta \left(\frac{G_r}{G_m} - 1\right)\right]^{-1}} \tag{6.1}$$

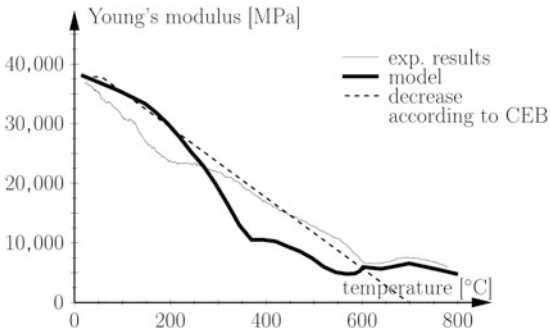

Fig. 6.8 Effective Young's modulus obtained from micromechanical model for unloaded concrete ($s = 0\,\%$) compared to experimental results (initial volume fractions: $f_p = 0.05$, $f_c = 0.25$, $f_a = 0.70$) and to decrease of stiffness according to CEB [4]

and

$$K_{\text{eff}} = \frac{\sum_r f_r K_r \left[1 + \alpha \left(\frac{K_r}{K_m} - 1\right)\right]^{-1}}{\sum_r f_r \left[1 + \alpha \left(\frac{K_r}{K_m} - 1\right)\right]^{-1}}, \tag{6.2}$$

with $r \in$ {porous cement paste (matrix), additional pore space (inclusion)} at Scale I and $r \in$ {homogenized material of Scale I (matrix), aggregates (inclusion)} at Scale II. The coefficients α and β are defined as

$$\alpha = \frac{3K_m}{3K_m + 4G_m} \quad \text{and} \quad \beta = \frac{6(K_m + 2G_m)}{5(3K_m + 4G_m)}. \tag{6.3}$$

In Eqs. (6.1) to (6.3), the index m refers to the matrix phase, while α and β represent the volumetric and deviatoric part of the Eshelby tensor S, specialized for the case of spherical inclusions. Furthermore, f_r [–] refers to the volume fraction of the r-th material phase, which is determined from the concrete mix-design, with 1,860 kg/m^3 siliceous material, 330 kg cement/fly ash, and 185 kg water. Under the assumption of complete hydration, the initial volume fractions (prior to fire loading) for the investigated concrete mixture are set to $f_p/f_c/f_a = 0.05$ / 0.25 / 0.7 (additional pore space (p)/porous cement paste (c)/aggregate (a), see Fig. 6.7).

As the material behavior (Young's modulus, Poisson's ratio) of porous cement paste is taken from the conducted experiments (see [22]), changes associated with dehydration are already considered in the cement-paste phase (see, e.g., [2, 13]).

The model response from Scale II, giving the effective Young's modulus (E_{eff}) for $s = 0\,\%$, is presented in Fig. 6.8, showing good agreement with experimental observations, especially for temperatures up to 200 °C.

Table 6.1 summarizes the evolution of the effective elastic properties obtained from the micromechanical model.

Table 6.1 Effective elastic properties obtained from micromechanical model for unloaded concrete ($s = 0\,\%$)

Temperature (°C)	K_{eff} (GPa)	G_{eff} (GPa)	ν_{eff} (–)	E_{eff} (GPa)
20	19.4	16.4	0.17	38.3
100	17.9	15.2	0.17	35.5
200	15.1	12.9	0.17	30.1
300	9.8	8.6	0.16	19.9
400	4.9	4.6	0.15	10.5
450	4.4	4.1	0.14	9.4
500	3.5	3.3	0.15	7.5
550	2.5	2.2	0.15	5.1
573	2.3	2.1	0.14	4.9
600	2.9	2.2	0.19	5.3
650	2.8	2.1	0.21	5.6
700	3.4	2.3	0.22	5.8
800	2.5	1.8	0.22	4.4

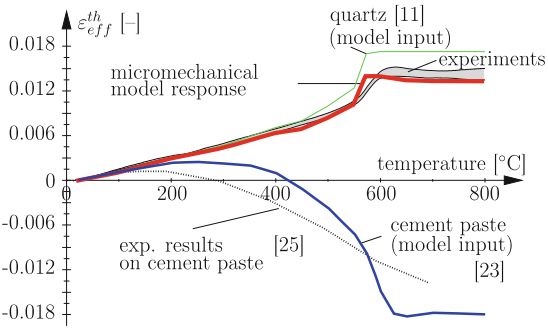

Fig. 6.9 Comparison of experimentally obtained free thermal strain ($s = 0\,\%$) with prediction by micromechanical model together with experimentally obtained thermal strain of the constituents (aggregates, cement paste), serving as model input

6.3.2 *Effective (Free) Thermal Strain*

When aggregates and cement paste are heated, they show a significant discrepancy in their thermal-dilation behavior (see Fig. 6.9). While siliceous aggregates are expanding during heating, cement paste is turning from expansion into shrinkage at 250 °C, which is explained by the continuous dehydration of cement paste [5,25]. Using the morphology of the proposed micromechanical framework (Fig. 6.6), the effective thermal strain is given by (see Appendix A)

$$\varepsilon^{\text{th}}_{\text{eff}} = \varepsilon^{\text{th}}_{m} + (1 - f_m \langle A \rangle_{Vm}) \frac{K_i}{K_{\text{eff}}} (\varepsilon^{\text{th}}_{i} - \varepsilon^{\text{th}}_{m}) \,, \tag{6.4}$$

where $\varepsilon^{\text{th}}_{i}$ and K_i are the thermal strain and bulk modulus of the inclusion phase i (additional pore space at Scale I, aggregates at Scale II), respectively. Furthermore, the index m refers to the matrix phase at the respective scale. Figure 6.9 contains the evolution of the effective thermal strain obtained from the proposed micromechanical model, showing excellent agreement with the experimental data. As indicated in Fig. 6.9, the free thermal strain is mainly driven by the behavior of the aggregates.

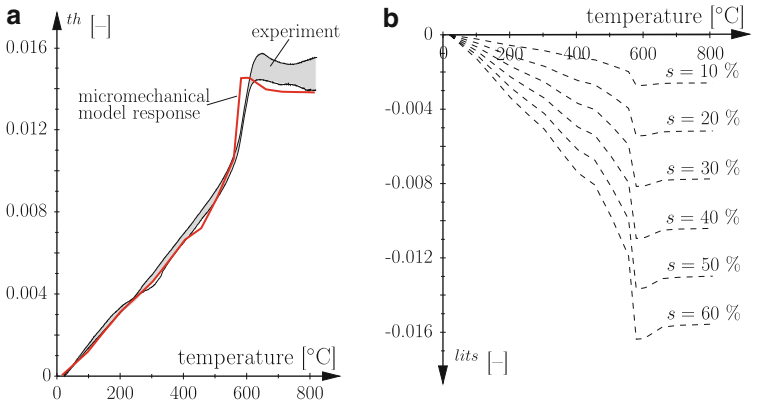

Fig. 6.10 (**a**) Evolution of free thermal strain of concrete, ε^{th}; (**b**) evolution of $\varepsilon^{\text{lits}}$ for different load levels according to Eq. (6.6) ($s = \sigma / f_{c,0} = 10$, 20, 30, 40, 50, and 60 %); $k = 2.35$; ε^{th} taken from Fig. 6.10a

6.4 Implementation

In the open literature, the load-dependent part of thermal strains of concrete is often referred to as "Load Induced Thermal Strains" (LITS). LITS may be considered by an additional strain (see, e.g., [27]), reading

$$\varepsilon = \varepsilon^{\text{el}}(T, \sigma) + \varepsilon^{\text{th}}(T) + \varepsilon^{\text{lits}}(T, \sigma), \tag{6.5}$$

where ε^{el} and ε^{th} represent the elastic strain and the free thermal strain (see ε^{th} in Fig. 6.10a), respectively. Commonly, the stress-dependence of LITS introduced in Eq. (6.5) is considered by empirical relations, such as the Thelandersson-approach [27]:

$$\varepsilon^{\text{lits}} = k \frac{\sigma}{f_{c,0}} \varepsilon^{\text{th}}(T)\,, \tag{6.6}$$

where k is a parameter depending on the type of loading ($k = 2.35$ for uniaxial loading, $k = 1.7$ for biaxial loading [27]) and $\sigma / f_{c,0}$ accounts for the influence of the load level, giving a linear dependence of LITS on the applied stress (see Fig. 6.18b). For determination of LITS, the micromechanical model response for ε^{th} given in Fig. 6.10a is used.

Introducing the LITS-compliance tensor $^{\text{lits}}$, $\varepsilon^{\text{lits}}$ given in Eq. (6.6) may be formulated in a more general form, reading

$$\varepsilon^{\text{lits}} = {}^{\text{lits}}(T) : \boldsymbol{\sigma}\,. \tag{6.7}$$

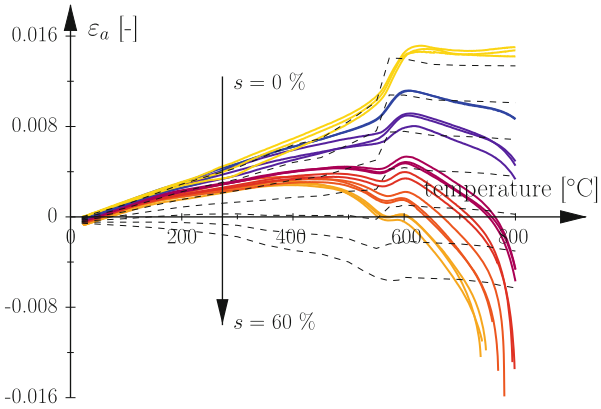

Fig. 6.11 Evolution of axial strain of concrete: Thelandersson-approach (Eq. (6.6) with $k = 2.35$) compared with experimental results

Combining Eqs. (6.5) and (6.7), the stress–strain law for heated concrete becomes

$$\sigma = \quad : \varepsilon^{\mathrm{el}} = \quad : \left[\varepsilon - \varepsilon^{\mathrm{th}} - \quad ^{\mathrm{lits}} : \sigma \right]. \tag{6.8}$$

Reformulation of Eq. (6.8) gives

$$\varepsilon = \left(\quad + \quad ^{\mathrm{lits}} \right) : \sigma + \varepsilon^{\mathrm{th}}, \tag{6.9}$$

where $\quad = \quad ^{-1}$ represents the elastic compliance tensor. Setting $\quad ^{\mathrm{lits}} = k \; ^{\mathrm{vol}}\varepsilon^{\mathrm{th}} / f_{c,0}$, the Thelandersson-approach given in Eq. (6.6) is recovered. For the special case of axisymmetric conditions (axial and radial stress and strain components), the overall compliance tensor in Eq. (6.9) becomes

$$+ \quad ^{\mathrm{lits}} = \frac{1}{E} \begin{bmatrix} 1 & -\nu \\ -\nu & 1 \end{bmatrix} + k \frac{\varepsilon^{\mathrm{th}}}{f_{c,0}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{6.10}$$

giving the axial strain ε_a in case of uniaxial loading, with the radial stress $\sigma_r = 0$, as

$$\varepsilon_a = \left(\frac{1}{E} + k \frac{\varepsilon^{\mathrm{th}}}{f_{c,0}} \right) \sigma_a + \varepsilon^{\mathrm{th}}. \tag{6.11}$$

Comparison between the experimental results presented in Sect. 6.2 with the results from Eq. (6.11) reveals a significant deviation between the model response and experimental results, especially in the low-temperature regime (see Fig. 6.11).

In order to improve the agreement between model response and experimental data, the level of loading $\sigma / f_{c,0}$ in Eq. (6.6) is reformulated, relating the stress to the

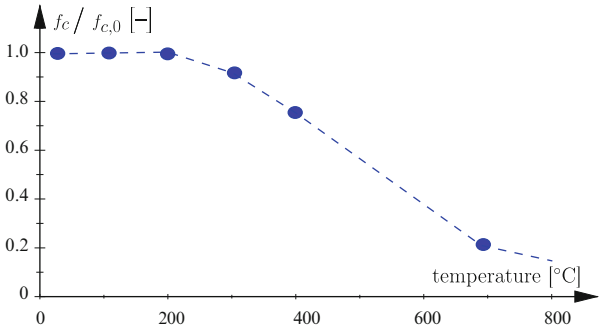

Fig. 6.12 Experimentally obtained normalized compressive strength of concrete as a function of temperature [23]

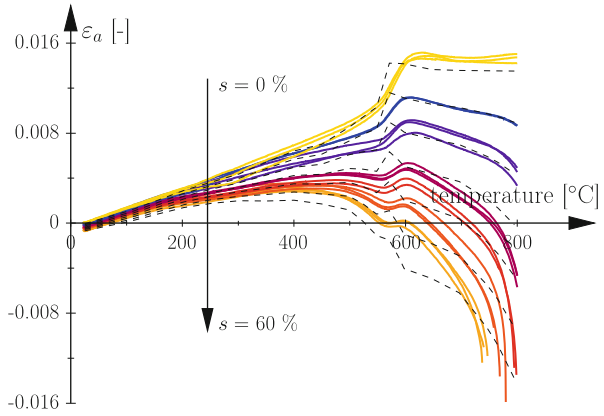

Fig. 6.13 Evolution of axial strain of concrete: modified Thelandersson-approach (Eq. (6.13) with $k = 0.4$) compared with experimental results ($f_c(T)$ taken from Fig. 6.12)

actual compressive strength $f_c(T)$ (see Fig. 6.12). Accordingly, $^{\text{lits}}$ in Eq. (6.7) becomes

$$^{\text{lits}} = k \; ^{\text{vol}}\varepsilon^{\text{th}} / f_c(T) \,, \tag{6.12}$$

giving the axial strain in case of uniaxial loading as

$$\varepsilon_a = \left(\frac{1}{E} + k \frac{\varepsilon^{\text{th}}}{f_c(T)} \right) \sigma_a + \varepsilon^{\text{th}} \,. \tag{6.13}$$

With this modification, the model response shows an improved agreement with the experimental results, especially in the low and medium temperature regime (see Fig. 6.13). The temperature dependent compressive strength $f_c(T)$ was investigated in the literature [1,12], highlighting a stress dependence of $f_c(T)$. For mechanically preloaded fire-exposed specimens the compressive strength was found to be higher

than for mechanically unloaded concrete. More recent investigations concerning the stress dependence of the compressive strength at 250 °C can be found in [19]. However, no stress dependence for the compressive strength $f_c(T)$ is considered in the proposed model up to now.

So far, the model was applied and validated by means of experimental data only for constant mechanical loading. In real-life applications, however, the amount of mechanical loading certainly changes with time, e.g., in case of unloading, the elastic deformation (with $\varepsilon^{\mathrm{el}} =$ $: \sigma$) vanishes. LITS deformations, on the other hand, account for the path-dependence of thermo-mechanical loading of heated concrete and must therefore remain. Accordingly, in contrast to the total formulation for the elastic strain $\varepsilon^{\mathrm{el}} =$ $: \sigma$, a differential form is adopted for LITS, reading

$$d\,\varepsilon^{\mathrm{lits}} = d \quad {}^{\mathrm{lits}} : \sigma \,, \tag{6.14}$$

with the actual stress tensor σ affecting the differential change of LITS via d $^{\mathrm{lits}}$. Replacing the differential changes in Eq. (6.14) by finite changes within the time increment $n+1$, one gets:

$$\begin{aligned} \sigma_{n+1} &= \quad {}_{n+1} : \left[\varepsilon_{n+1} - \varepsilon^{\mathrm{th}}_{n+1} - \left(\varepsilon^{\mathrm{lits}}_{n} + \Delta\varepsilon^{\mathrm{lits}}\right)\right] \\ \sigma_{n+1} &= \quad {}_{n+1} : \left[\varepsilon_{n+1} - \varepsilon^{\mathrm{th}}_{n+1} - \varepsilon^{\mathrm{lits}}_{n} - \Delta \quad {}^{\mathrm{lits}} : \sigma_{n+1}\right], \end{aligned} \tag{6.15}$$

where the incremental change of the LITS-compliance tensor is determined from Δ $^{\mathrm{lits}} =$ $^{\mathrm{lits}}_{n+1} -$ $^{\mathrm{lits}}_{n}$. Rewriting Eq. (6.15) for the case of stress-driven situations (such as in case of axisymmetric uniaxial loading) gives

$$\varepsilon_{n+1} = \quad {}_{n+1} : \sigma_{n+1} + \varepsilon^{\mathrm{th}}_{n+1} + \varepsilon^{\mathrm{lits}}_{n} + \Delta\varepsilon^{\mathrm{lits}} \,. \tag{6.16}$$

with

$$\Delta\varepsilon^{\mathrm{lits}} = \left({}^{\mathrm{lits}}_{n+1} - \quad {}^{\mathrm{lits}}_{n}\right) : \sigma_{n+1} \,. \tag{6.17}$$

It can be seen in Fig. 6.14, that the agreement of the proposed incremental model with experimental data is equally good as the total formulation (see Fig. 6.13).

In order to validate the proposed differential formulation of LITS, experiments with changing load levels were performed and compared to the respective model response, considering both the modified total (Eq. (6.13)) and the differential formulation (Eq. (6.16)):

1. Within the first experiment, the level of loading is increased in four steps (see Fig. 6.15a). Starting from $s = 10\,\%$, the load level is increased in three steps to 20, 30, and finally 40 %.
2. In the second experiment, the level of loading is first increased (from $s = 20$–40 %) and then reduced to zero loading ($s = 0\,\%$) (see Fig. 6.16a).

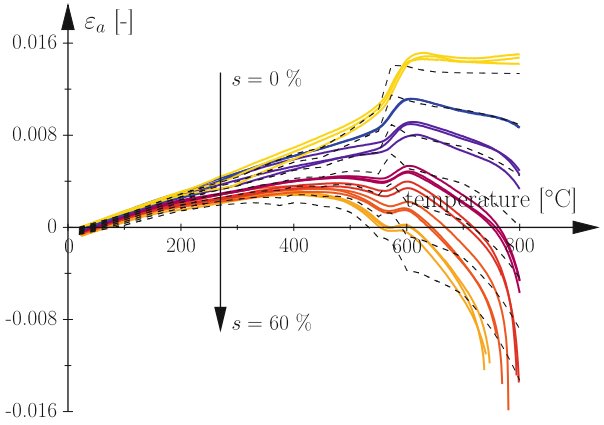

Fig. 6.14 Evolution of axial strain of concrete: differential formulation (Eq. (6.16) with $k = 0.4$) compared with experimental results

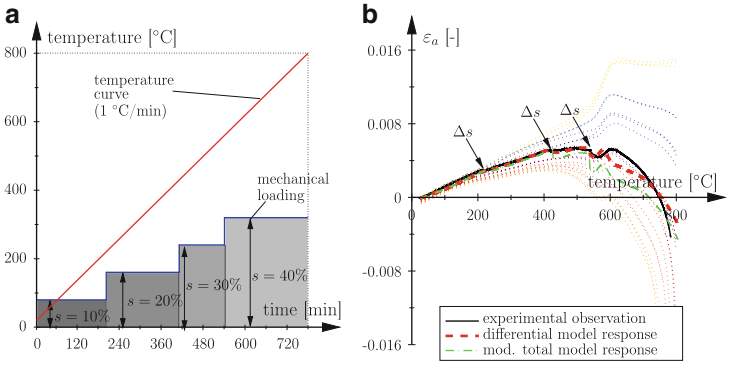

Fig. 6.15 Experiment 1: (**a**) temperature curve (1 °C/min) and mechanical loading (from 10, 20, 30, to 40 %); (**b**) evolution of axial strain

3. During the third experiment, the initial level of loading ($s = 40\,\%$) is decreased ($s = 10\,\%$) and finally increased ($s = 30\,\%$) (see Fig. 6.17a).
4. During the fourth experiment, the initial level of loading ($s = 50\,\%$) is linearly decreased to $s = 0\,\%$ between 400 and 670 °C and again linearly increased up to $s = 20\,\%$ at 770 °C (see Fig. 6.18a).

For all experiments, the better agreement with the experimentally obtained strain is found when using the proposed differential formulation (Eq. (6.16)). The response of the modified total formulation (Eq. (6.13)), strongly deviating from the experimental results, shows the largest error in case of mechanical unloading in the high-temperature regime.

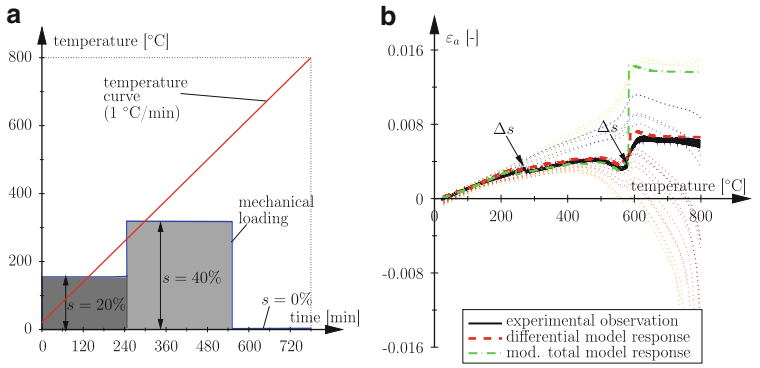

Fig. 6.16 Experiment 2: (**a**) temperature curve (1 °C/min) and mechanical loading (from 20, 40, to 0 %); (**b**) evolution of axial strain

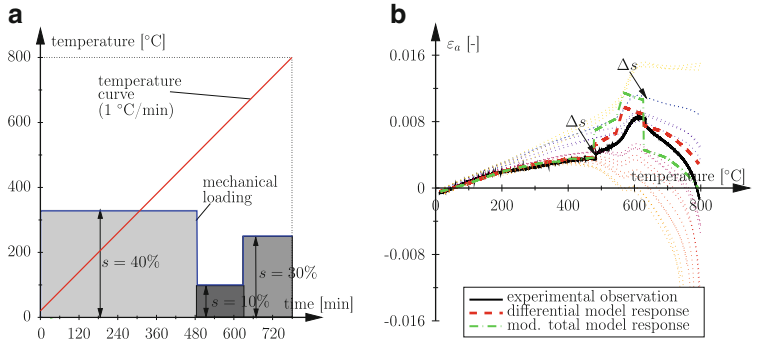

Fig. 6.17 Experiment 3: (**a**) temperature curve (1 °C/min) and mechanical loading (from 40, 10, to 30 %); (**b**) evolution of axial strain

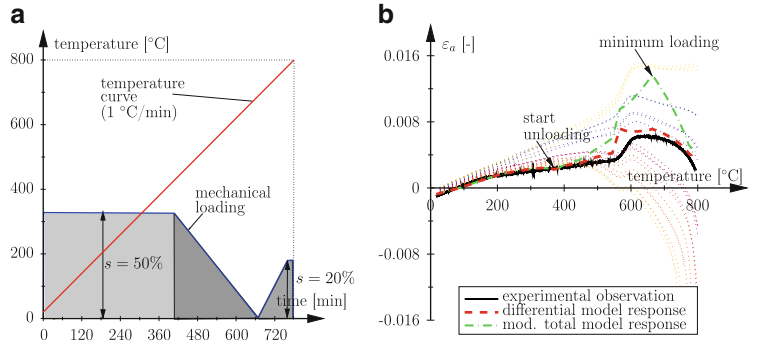

Fig. 6.18 Experiment 4: (**a**) temperature curve (1 °C/min) and mechanical loading (linear decrease from 50 to 0 % and linear increase from 0 to 20 %); (**b**) evolution of axial strain

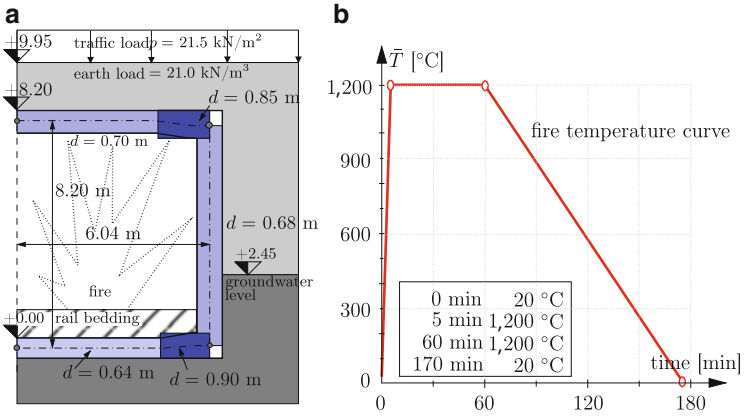

Fig. 6.19 Rectangular tunnel cross-section: (**a**) geometric properties and (**b**) applied temperature loading within the tunnel (applied to side wall and ceiling of the frame)

6.5 Finite-Element Implementation and Numerical Results

In this section, the differential formulation for the strain behavior of heated concrete outlined in the previous section is implemented into a finite element (FE) program [24]. In contrast to the stress-driven uniaxial stress situation encountered in the LITS experiments, strain increments are given by the underlying incremental-iterative solution procedure in nonlinear FE analysis, while the stress state σ_{n+1} at the end of the respective time increment needs to be determined. According to Eq. (6.18), σ_{n+1} is given by

$$\sigma_{n+1} = \left[\quad_{n+1} + \Delta \quad^{\text{lits}}\right]^{-1} : \left[\varepsilon_{n+1} - \varepsilon^{\text{th}}_{n+1} - \varepsilon^{\text{lits}}_{n}\right]. \tag{6.18}$$

In order to highlight the effect of the underlying LITS formulation, the proposed material model is used within the numerical analysis of a rectangular tunnel cross-section subjected to fire loading. The geometric properties of the considered tunnel cross-section are presented in Fig. 6.19a. The thermal loading within the cross-section is obtained from a coupled thermo-hydro-chemical analysis [30], using the prescribed temperature loading within the tunnel shown in Fig. 6.19b.

In order to determine the influence of LITS on the structural response, three different material models are considered:

- Model 1 (no LITS): No consideration of LITS;
- Model 2 (TOT): Modified total formulation of LITS based on the Thelandersson-approach [27] (Eq. (6.13));
- Model 3 (DIFF): Differential formulation of LITS (Eq. (6.18)).

While LITS are considered only in case of compressive loading of concrete, the tensile stresses are limited in all considered models (Model 1 to 3) using

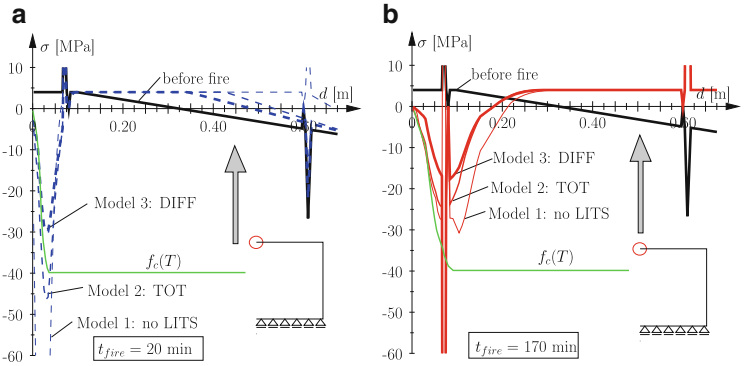

Fig. 6.20 Comparison of numerical results (model 1 to 3): stress distributions at (**a**) $t_{\text{fire}} = 20$ and (**b**) $t_{\text{fire}} = 170$ min

a Rankine failure criterion, with $f_t(T) = 1/10 f_c(T)$ (according to [24]). The micromechanics-based Young's modulus (see Fig. 6.8) and free thermal strain (see Fig. 6.9) are employed. For the simulation of the reinforcement, a 1-D elasto-plastic material model was chosen, considering degradation of stiffness and yield-strength according to [9] (see [23] for details).

In the underlying fire scenario, a cooling phase is included (see Fig. 6.19b). During the heating phase, the evolution of the material properties is determined based on the temperature dependence shown in Fig. 6.12. During cooling the material properties (Young's modulus, compression/tensile strength) are dependent upon the maximum temperature reached. Since LITS were observed to take place during the first heating only [12], LITS are considered only during heating. In the course of cooling, no change of LITS take place.

In Fig. 6.20, stress distributions at the top of the rectangular tunnel cross-section for different time instants ($t_{\text{fire}} = 0$, 20, 170 min) are presented.

Model 1 (no LITS) gives comparably high compressive stresses, even exceeding the compressive strength of concrete $f_c(T)$. Model 2 (TOT) results in a reduction in the stress build-up nevertheless, the stresses still exceed the compressive strength of concrete. Finally, Model 3 (DIFF) further reduces the compressive stresses which stay below the temperature-dependent compressive strength.

Figure 6.21 shows the deformation of the tunnel cross-section, with the deformation history in the symmetry axis at the top of the tunnel given in Fig. 6.21a and deformation patterns of the whole cross-section given in Fig. 6.21b.

The largest restraint occurs for Model 1 (no LITS), resulting in large regions with plastic deformations within the reinforcement. On the other hand, the model response for Model 3 (DIFF) shows no plastic deformations of the reinforcement at all since the stress build-up due to thermal loading is considerably reduced by LITS, resulting in less loading of the concrete and, thus, the reinforcement bars.

Finally, the influence of the different models on the evolution of bending moments is presented in Fig. 6.22. In case of Model 3, the thermally induced

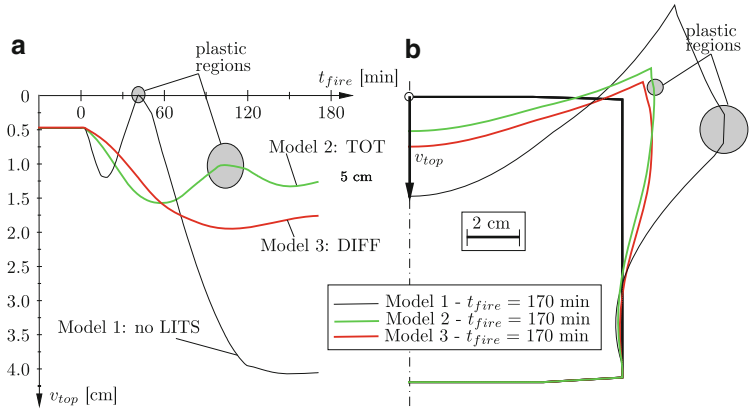

Fig. 6.21 Comparison of numerical results (model 1 to 3): (**a**) deformation history at top of the tunnel; (**b**) deformation pattern of the whole cross-section at $t_{\text{fire}} = 170$ min

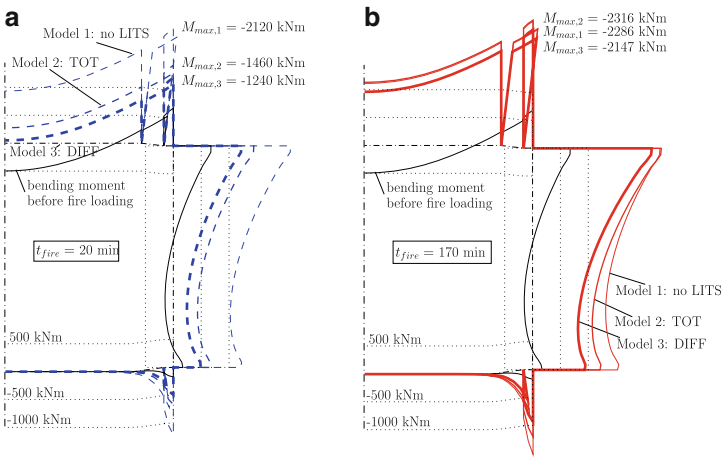

Fig. 6.22 Comparison of bending-moment distribution obtained with model 1 to 3 at (**a**) $t_{\text{fire}} = 20$ min and (**b**) $t_{\text{fire}} = 170$ min

bending moments are reduced, resulting from lower stresses within the cross-section. On the other hand, totally neglecting the effect of LITS leads to the highest bending moments. For Model 3, the maximum bending moment at $t_{\text{fire}} = 170$ min is reduced (see Fig. 6.22), indicating—based on the more realistic material model for heated concrete—a higher safety of the underlying tunnel design.

6.6 Concluding Remarks

Within this chapter, a differential strain formulation allowing the description of the path dependence of the strain behavior of concrete (LITS) in case of combined thermo-mechanical loading is presented. The underlying material parameters for concrete (Young's modulus, Poisson's ratio, free thermal strain) are determined using a micromechanical model. Based on experimental observations highlighting the influence of the level of loading on the strain behavior of heated concrete, the developed material model was validated. Finally, the effect of different modes of consideration of LITS on the structural response of a rectangular tunnel cross-section was assessed. Based on the obtained results, the following conclusions can be drawn:

- Effect of plasticity: The introduction of the differential formulation to consider LITS reduces compressive stresses induced by thermal restraint within concrete which then remain below the respective ultimate compressive strength. In case of tensile loading, LITS were not considered but a Rankine-type plasticity model was used for the simulation of tensile cracking of concrete.
- Loading of reinforcement: Due to the smaller stress build-up obtained from the proposed differential formulation, the stresses in the reinforcement are reduced, leading to less plastic deformations and, thus, an increase of the integrity of the tunnel-support structure.
- Reduction of bending moment: Originating from the reduced stress build-up obtained from the differential formulation, the maximum bending moment was reduced, indicating a higher structural safety of the tunnel-support structure when subjected to fire.

The presented approach of improved modeling of the material behavior of concrete under temperature loading, enabling a more realistic prediction of the thermally induced stress build-up within the concrete lining in the event of fire, provides a proper basis for the realistic structural safety assessment and design.

Acknowledgements This research was conducted with financial support by the Austrian Ministry for Transport, Innovation and Technology (bm.vit) within the KIRAS-project (Austrian security research program) 824781 "Sicherheit von Hohlraumbauten unter Feuerlast—Entwicklung eines Struktursimulationstools (Safety of underground structures under fire loading—Development of a structural simulation tool)". The authors want to take this opportunity to thank all members of the research consortium for the fruitful and inspiring cooperation throughout this research project, having ranged from fundamental research toward applied research dealing with the structural safety assessment of tunnels.

Appendix: Effective Prescribed Strains in Two-Phase Materials

According to [14], the effective strain E_{eff} is related to the prescribed strain $\bar{\varepsilon}$ in the material phases as:

$$K_{\text{eff}} E_{\text{eff}} = \langle A : K : \bar{\varepsilon} \rangle_V \, . \tag{6.19}$$

Considering a two-phase material with matrix m and inclusion i, with

$$\bar{\varepsilon} = \bar{\varepsilon}_m \text{ in } V_m \, , \; \bar{\varepsilon} = \bar{\varepsilon}_i \text{ in } V_i \, , \tag{6.20}$$

$\bar{\varepsilon}_i$ may be substituted by

$$\bar{\varepsilon}_i = \bar{\varepsilon}_m + \Delta \bar{\varepsilon}_i \, . \tag{6.21}$$

Rewriting Eq. (6.19) and considering Eq. (6.21) gives

$$K_{\text{eff}} E_{\text{eff}} = \langle A : K \rangle_V \bar{\varepsilon}_m + f_i \langle A : K \rangle_{V_i} \Delta \bar{\varepsilon}_i \, . \tag{6.22}$$

Considering $K_{\text{eff}} = \langle A : K \rangle_V$ in Eq. (6.22), one gets

$$E_{\text{eff}} = \bar{\varepsilon}_m + f_i \langle A \rangle_{V_i} \frac{K_i}{K_{\text{eff}}} (\bar{\varepsilon}_i - \bar{\varepsilon}_m) \, , \tag{6.23}$$

where $\langle A \rangle_{V_i}$ is given in [14].

References

1. Abrams, M.S.: Compressive strength of concrete at temperatures to 1600 F. Am. Concrete Inst. SP **25**, 33–58 (1971)
2. Alarcon-Ruiz, L., Platret, G., Massieu, E., Ehrlacher, A.: The use of thermal analysis in assessing the effect of temperature on a cement paste. Cement Concrete Res. **35**, 609–613 (2005)
3. Anderberg, Y., Thelandersson, S.: Stress and Deformation Characteristics of Concrete at High Temperatures: 2. Experimental Investigation and Material Behaviour Model. Technical Report 54. Lund Institute of Technology, Lund (1976)
4. CEB: Fire Design of Concrete Structures, Bulletin d'Information 208. CEB, Lausanne (1991)
5. Cruz, C.R., Gillen, M.: Thermal expansion of portland cement paste, mortar, and concrete at high temperatures. Fire Mater. **4**(2), 1–12 (1980)
6. DeJong, M.J., Ulm, F.-J.: The nanogranular behavior of C-S-H at elevated temperatures (up to 700 °C). Cement Concrete Res. **37**, 1–12 (2007)
7. Dweck, J., Ferrerira da Silva, P.F., Büchler, P.M., Cartledge, F.K.: Study by thermogravimetry on the evolution of ettringite phase during type II Portland cement hydration. J. Therm. Anal. Calorim. **69**, 179–186 (2002)

8. Ehm, C.: Versuche zur Festigkeit und Verformung von Beton unter zweiaxialer Beanspruchung und hohen Temperaturen [Experiments on strength and strain of concrete under biaxial loading at high temperatures]. Ph.D. thesis, University of Braunschweig, Braunschweig (1985)
9. EN1992–1–2: Eurocode 2 – Bemessung und Konstruktion von Stahlbeton- und Spannbeton-tragwerken – Teil 1-2: Allgemeine Regeln – Tragwerksbemessung für den Brandfall [Eurocode 2 – Design of concrete structures – Part 1-2: General rules – Structural fire design]. European Committee for Standardization (CEN) (2007)
10. Gawin, D., Pesavento, F., Schrefler, B.A.: Towards prediction of the thermal spalling risk through a multi-phase porous media model of concrete. Comput. Methods Appl. Mech. Eng. **195**(41–43), 5707–5729 (2006)
11. Jay, A.H.: The thermal expansion of Quartz by X-ray measurements. Proc. R. Soc. Lond. **37**(133), 195–215 (1985)
12. Khoury, G.A., Grainger, B.N., Sullivan, P.J.E.: Strain of concrete during first heating to 600°C under load. Mag. Concr. Res. **37**(133), 195–215 (1985)
13. Kührer, T.: Nachbrandfestigkeit von zementgebundenen Werkstoffen. Druckversuche und Thermogravimetriemessungen [Fire exposed cementitious materials, compressive strength and TG-measurements]. Technical Report, Vienna University of Technology, Vienna (2008)
14. Lackner, R., Pichler, C., Kloiber, A.: Artificial ground freezing of fully saturated soil: viscoelastic behavior. J. Eng. Mech. **134**(1), 1–11 (2008)
15. Lakshtanov, D.L., Sinogeikin, S.V., Bass, J.D.: High-temperature phase transitions and elasticity of silica polymorphs. Phys. Chem. Miner. **34**(1), 11–22 (2007)
16. Lee, J., Xi, Y., William, K., Jung, Y.: A multiscale model for modulus of elasticity of concrete at high temperatures. Cement Concrete Res. **39**, 754–762 (2009)
17. Mori, T., Tanaka, K.: Average stress in matrix and average elastic energy of materials with misfitting inclusions. Acta Metar. **21**, 571–574 (1973)
18. Nielsen, C.V., Pearce, C.J., Bićanić, N.: Improved phenomenological modelling of transient thermal strains for concrete at high temperatures. Comput. Concrete **1–2**, 189–209 (2004)
19. Petkovski, M.: Effects of stress during heating on strength and stiffness of concrete at elevated temperature. Cement Concrete Res. **40**, 1744–1755 (2010)
20. Pichler, C.: Multiscale characterization and modeling of creep and autogenous shrinkage of early-age cement-based materials. Ph.D. thesis, Vienna University of Technology, Vienna (2007)
21. Pichler, C., Lackner, R.: A multiscale micromechanics model for early-age basic creep of cement-based materials. Comput. Concrete **5**(4), 295–328 (2008)
22. Ring, T., Zeiml, M., Lackner, R., Eberhardsteiner, J.: Experimental investiagtion of strain behavior of heated cement paste and concrete. Strain **49**, 249–256 (2013)
23. Ring, T., Zeiml, M., Lackner, R.: Underground concrete frame structures subjected to fire loading: Part I-Large scale fire tests. Eng. Struct. **58**, 175–187 (2014)
24. Savov, K., Lackner, R., Mang, H.A.: Stability assessment of shallow tunnels subjected to fire load. Fire Saf. J. **40**, 745–763 (2005)
25. Schneider, U.: Concrete at high temperature: a general review. Fire Saf. J. **13**, 55–68 (1988)
26. Terro, M.J.: Numerical modeling of the behavior of concrete structures in fire. Am. Concrete Inst. **95**(2), 183–193 (1998)
27. Thelandersson, S.: Modeling of combined thermal and mechanical action in concrete. J. Eng. Mech. **113**(6), 893–906 (1987)
28. Thienel, K.-C.: Festigkeit und Verformung von Beton bei hoher Temperatur und biaxialer Beanspruchung – Versuche und Modellbildung [Strength and deformation of concrete at high temperature – experiments and modeling]. Technical Report 437, Deutscher Ausschuss für Stahlbeton, Berlin (1994)
29. Tsivilis, S., Kakali, G., Chaniotakis, E., Souvaridou, A.: A study on the hydration of portland limestone cement by means of TG. J. Therm. Anal. Calorim. **52**, 863–870 (1998)
30. Zeiml, M., Lackner, R., Pesavento, F., Schrefler, B.A.: Thermo-hydro-chemical couplings considered in safety assessment of shallow tunnels subjected to fire load. Fire Saf. J. **43**(2), 83–95 (2008)

Chapter 7
Scientific Computing in Urban Water Management

R. Sitzenfrei, M. Kleidorfer, M. Meister, G. Burger, C. Urich, M. Mair, and W. Rauch

Abstract Urban water management is concerned with the supply of drinking water to households and industry and the discharge of stormwater and waste water from the urban environment. The system is highly dynamic and driven by meteorology, urban development, change in land use and technological innovations. Key mechanisms in urban water systems are on the one hand the transport of water and substances in the environment and the pipe network and on the other hand the conversion of substances due to physical and biochemical processes. Urban water management thus requires computer simulations in time (ranging typically from hours to years) and space (one to three dimensions). With the models becoming more and more complex by simulation at detailed spatio-temporal scale and by simulating whole urban environments, the limits of traditional numerical methods have been reached. In this chapter three emerging topics in scientific computing in urban water management are discussed and the need for advanced software methods is exemplified.

7.1 Introduction

Urban water management is concerned with the supply of drinking water to households and industry and the discharge of stormwater and waste water from the urban environment. Key mechanisms in urban water systems are on the one hand the transport of water and substances in the environment and the pipe network and on the other hand the conversion of substances due to physical and biochemical

R. Sitzenfrei (✉) • M. Kleidorfer • M. Meister • G. Burger • C. Urich • M. Mair • W. Rauch
Unit of Environmental Engineering, University of Innsbruck, Technikerstr. 13, A6020 Innsbruck, Austria
e-mail: Robert.Sitzenfrei@uibk.ac.at; Manfred.Kleidorfer@uibk.ac.at; Michael.Meister@uibk.ac.at; Gregor.Burger@uibk.ac.at; Christian.Urich@uibk.ac.at; Michael.Mair@uibk.ac.at; Wolfgang.Rauch@uibk.ac.at

G. Hofstetter (ed.), *Computational Engineering*, DOI 10.1007/978-3-319-05933-4_7,
© Springer International Publishing Switzerland 2014

processes. Urban water management thus requires computer simulations in time (ranging typically from hours to years) and space (one to three dimensions).

The requirements for urban water management like hygiene, economics, environment protection, etc. resulted in traditional engineering design (centralized network systems based on pipes and nodes). The evolvement of such a system is highly dynamic and driven by meteorology, urban development, change in land use and technological innovations [62] or climate change [24].

Traditionally, in the design process and assessment of water networks, different parts of the networks were regarded separately and frequently even by analytical equations and/or empirical relations. But in the last decades and assisted with increasing computer power, the assessment of water networks is proceeding from investigations on different, separate parts (e.g. a single catchment with a combined sewer overflow (CSO) structure or waste water treatment plant (WWTP)) to a numerical, model-based view on the entire network system [48]. Going one step further, in the 1990s, integrated models were developed and applied. In these models different sub-systems/models (i.e. sewer, CSO, WWTP and receiving water) are combined to integrated approaches in order to assess water pollution in the receiving water (e.g. [20]).

Usually, in such approaches, the engineering system (i.e. network structure) and its boundary conditions (e.g. dry weather flow, drained area, etc.) are kept static. Therefore, the spatial dynamic drivers of urban water systems (i.e. cities) are not considered explicitly, but only as multipliers for expected future conditions (e.g. prospective demand, population growth, etc.)

New developments in data management and increasing availability of digital data enabled engineers to use GIS-software and raster-based spatial distributed data for their investigations. Among others, raster-based GIS-data can be used to obtain input parameters for numerical network simulations (e.g. topography, impervious area from processing ortho-photos, land-use and population densities). For example, Sitzenfrei et al. [60] presented a procedure for automatic generation of water distribution networks based on GIS data topography and population densities. Also, simulation engines are integrated in GIS-software environment to use the capabilities of GIS-software for visualization, data processing and data modification combined with different hydraulic simulation models. This even enables to investigate different infrastructure systems in a comprehensive approach (multi-utility, e.g. Mike Urban, [7, 59]). Only an interlinked digital description of a city enables new comprehensive investigations and the identification of coherences. With such an integrated "Digital City" [53] interlinked infrastructure systems can be investigated (e.g. water supply under consideration of water saving strategies or water reuse and the impact on the sewer system [55]). Further, this helps not only to test plausibility of data (intersection and alignment of different data, etc.) but also to complement insufficient data sets with e.g. stochastic approaches [31] or inverse modelling. But taking this approach one step further, the question arises: are the network-based descriptions and models still necessary respectively advantageous for up-coming modelling tasks? To face challenges of climate change and future developments, decentralized solution for on-site water reuse strategies are increasingly developed,

investigated and implemented [33]. Especially for simulation and analysis of the spatial dimensions of decentralized solutions (rain-water infiltration, water and rain-water reuse, etc.), the network-based models are not effective.

Spatially enhanced integrated modelling approaches (including infrastructure, land use and population models) allow novel insights how dynamic urban systems work and to identify system coherences [42]. Recent research focuses on the integration of urban simulation models in the assessments of water infrastructure systems to integrated urban simulation approaches (e.g. [8,56,58] or [45]). Therein, the infrastructure models (e.g. water distribution system or combined sewer system and WWTP) are coupled with urban simulation tools for dynamic simulation of population development under consideration of socio-economic issues. For these investigations and for an interactive consideration of time dynamics in the sub-models, spatially distributed information on parameters is required. Therefore, raster-based models (e.g. based on local water balances with regional interactions) are in this context more capable to model especially decentralized systems and to enable spatially distributed, time dynamic interactions. The transition from centralized systems to decentralized systems is an important part, respectively, and the coexistence and functionality of both systems has to be investigated (e.g. rainwater harvesting and water distribution system).

In this chapter we will focus on three issues that have been in the centre of scientific attendance for the last decade. The first topic, the estimation of possible solutions for water management in megacities requires the spatially distributed, dynamic and grid-based simulation of the evolution of public water infrastructure under consideration of changes (e.g. climate, global, environment, economy, land use). Currently, these simulations can be realized with the help of frameworks for integrated modelling like, e.g. "DynaMind"—a workflow engine especially designed for urban water management simulations.

Second topic is the utilization of multicore facilities in software for simulating the dynamics in water networks. The basic features of parallel coded network simulations are discussed for standard public domain software tools in the field that is SWMM for drainage systems and EPANET for water supply networks.

Third, smoothed particle hydrodynamics (SPH) is presented as an alternative numerical method to explore fluid flow phenomena in urban water management based on the simulation of particle movement that can easily be extended towards multiphase flow phenomena, solids transport and bioconversion processes. Thus SPH could potentially be the core numerical engine to simulate fluxes and processes in the complete water infrastructure on a very detailed level.

7.2 From Water Networks to an Integrated Assessment of Urban Water Systems

To identify different steps of model complexity and also to evaluate the according model requirements a literature review is used. Based on this, different levels of

modelling approaches are outlined/defined and their advantages and disadvantages are pointed out, respectively.

7.2.1 State-of-the-Art Modelling Approaches

Traditionally individual parts of the drainage systems were calculated by engineers independently with simplified or empirical equations (empirical Manning equation for open channel flows, time area method, etc.). Among others, the software tool SWMM enabled modelling of the entire sewage system and the tool is increasingly developed (starting from 1973 to the current version SWMM5 [51]). With increasing computer power but also with progressing understanding of the relevant mechanism in the different sub-parts of wastewater systems, integrated models are developed which couple different sub-systems of the urban (waste)water cycle (e.g. [20]). In the last few years, the requirement of integrated water management approaches increased and new modelling approaches were developed and applied. Hardy et al. [19] developed an integrated water management approach (UrbanCycle) to investigate urbanization in the context of efficiency of the implemented technical systems. Especially, for regions with high climatic variability, changes in boundary conditions can possibly produce highly inefficient technical solutions of the urban water management systems. Traditionally, investigations are performed with top–down approaches, but for interacting systems new approaches are required [19]. UrbanCycle is a modelling framework for an integrated view on water supply, wastewater and stormwater solutions which aims to model interacting systems from bottom up. Therefore, clusters for allotments represent the water cycle/reuse at that scale. For a performance assessment, these clusters are connected to headwork systems (e.g. main trunks, etc.). Doglioni et al. [8] developed another integrated framework to model interactions of the urban water systems with urban expansion. The developed integrated framework dynamically couples a land use change model, a sewer simulation model and a wastewater treatment plant (WWTP) model to an integrated approach. For the infrastructure (sewage system and WWTP) no dynamic update (redesign over time) was regarded. Therewith, the impact of urban expansion on the existing sewage system (node-based) and the WWTP were investigated. A multi-agent model combined with a cellular automata-based model was used to model the (raster-based) urban expansion and population dynamics. But for coupling of the raster-based information of the urban development model the spatially distributed information was abstracted to the node based representation of the sewage network and no information feedback and therefore no infrastructure adaptation was regarded.

All these integrated approaches couple different models with data generalization from raster-based to node-based information and vice versa, respectively and work on different modelling scale (i.e. allotment cluster, head network level). In general, such approaches bring up the problem of data abstraction and generalization. Especially for integrated models with population and land-use dynamics, informa-

tion feedback through the coupled models is crucial. There is also an increasing requirement of new modelling approaches from another research field. Brown et al. [3] formulated a transitions framework for urban water systems to describe the historical and future scenarios of water management in Australian cities. Therein, different steps of the transition from traditional urban water systems (centralized water supply and sewage) to integrated urban water cycles (fit for purpose water sources, i.e. water sources with different qualities are appropriately used) and sustainable water management are defined. The framework encompasses six transition steps to an adaptive, water sensitive city. For the transition to such water sensitive cities, integrated modelling approaches including dynamic socio-economic issues and decentralized solutions are required.

For environmental processes [44] described the need for integrated assessment and modelling of such systems. Interdisciplinarity is described to be the key to address environmental problems of the twenty-first century [43]. Modelling the water cycle with taking into account socio-economic processes is a challenging task. Especially, investigations based on agent based modelling techniques have the potential to manage such spatially distributed and dynamic systems [36]. Also, investigations on the impact of climate change have to be done on a large temporal scale. To estimate, e.g. the impact of climate change on our environment requires therefore the inclusion of the temporal change of demography and infrastructure in the investigations. For example, Barth et al. [2] investigated these aspects on the rural Upper Danube Catchment with a multi-actor simulation framework denoted DANUBIA including agent-based approaches. To assess the impact on the entire water cycle in that approach, scenario analyses were performed with this modelling framework. Therein a raster-based modelling concept (proxel concept) was used for a description of interdisciplinary interactions. Each raster cell (i.e. each proxel) is connected to other proxels through fluxes [29]. In the approach a 1 km proxel size is used for (mesoscale) modelling of land surface and socio-economic processes.

The European FP7 project “PREPARED enabling change” aims to develop a software tool for modelling an integrated urban water management cycle. This includes the technical water systems as well as socio-technical dynamics (urban development, socio-economic transition, etc.). The project aims to model interactions of water infrastructure including decentralized solutions, (multi-utility assessment) including urban development, dynamic adaptation of technical urban water systems under consideration of socio-economic transitions.

7.2.2 Raster-Based and Node-Based Models

With raster-based models, the available spatial information can be directly used (see Fig. 7.1). There is no need of data generalization or abstraction (abstraction for node-based models). This helps to cut down calculation time in terms of feedback loops (computation time of data conversion) but also assists to evaluate decentralized systems (e.g. rain-water harvesting).

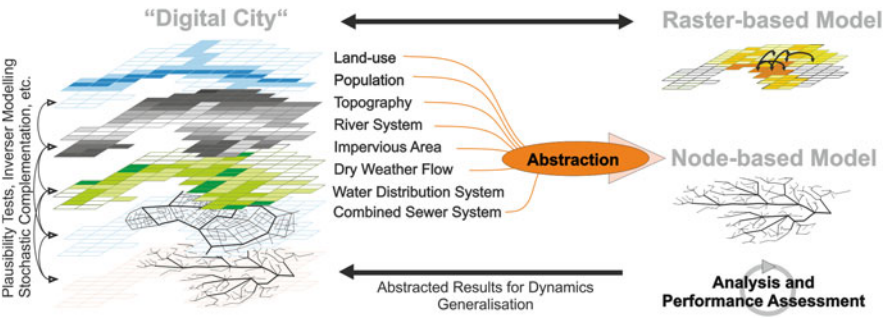

Fig. 7.1 Node-based and raster-based models in context with a "Digital City" description

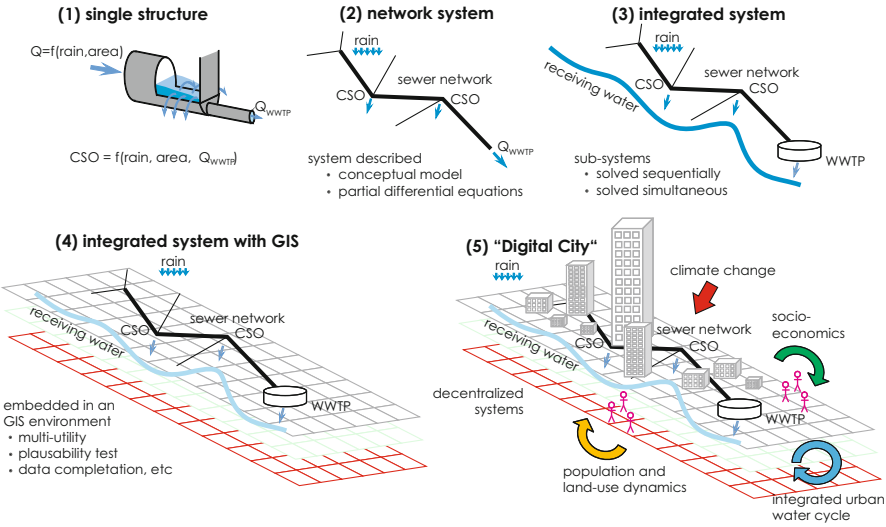

Fig. 7.2 Different modelling approaches for urban water systems

7.2.3 Definition of a Framework for Modelling Approaches

In the following, different steps of model complexity to evaluate urban water systems are identified. Therewith, the shift in approaches to assess water systems is described and discussed. Further, the theoretical framework of an integrated "Digital City" [53] to comprehensively assess urban water systems on a raster-based description of the investigation areas is characterized (see Fig. 7.2).

1. *Single structures* of systems are investigated (traditionally with "paper and pencil" based methods). Detailed information on a specific structure is required and therefore very case specific, local results are obtained. With these approaches no holistic view can be obtained. Traditionally, such investigations are performed due to either limited computer power or because of very specific questions

(e.g. specific design issues). In regulatory guidelines (e.g. in Austrian guideline for design of CSOs) there has already been a shift in the requirements for assessment of such systems. While the former guideline focuses on design of specific CSO structures [40], the new version aims already on an assessment of the entire combined sewer system performance [23, 41].

2. *Entire systems*/processes are investigated (e.g. sewer network, water distribution system or WWTP, etc.). But still, each system is assessed separately and only the performance of the specific system (network, etc.) is simulated. But there is no holistic view, and no broader assessment of the entire water systems (no system interconnections).
3. *Integrated* urban water methods couple models of different sub-processes to an integrated approach. By coupling different sub-systems to such an integrated assessment helps to understand and identify holistic system coherences like real time control for combined sewers and oxygen depletion in the receiving water. Also coupled water supply and urban drainage models can be used to assess low flow conditions [54].
4. *GIS-assisted integrated* infrastructure systems (different infrastructure models, are embedded in a GIS environment as, e.g. provided in the software Mike Urban or Hystem-Extran). The rising amount of available digital data enables engineers to use GIS-software and raster-based spatially distributed data for their investigations. Especially for data intersection, multi-utility interactions, data verification, plausibility tests these new approaches have comprehensive potential. Going one step further, population models are integrated in holistic modelling approaches to investigate dynamic interactions. For raster-based population models, this requires extensive calculation time for data conversions and dynamic feedback loops.
5. The *Digital City* approach denotes integrated urban systems (interlinked infrastructure and urban simulation models for population dynamics with socio-economics, etc.) combined with raster-based models and data management. This allows both the consideration of decentralized systems and spatio-temporal interactions and the dynamic feedback of population models to water infrastructure. The spatial resolution requirements to model such systems node-based are (especially for larger systems) at least a significant computational burden and sometimes even prohibitive for available computer power. Approaches including urban dynamics with data conversion (raster to node-based data conversion and vice versa) represent a pre-stage of such a "Digital City". Fully raster-based models respectively also fully vector based descriptions on the other hand enable comprehensive and extensive investigations of the urban system. For consideration of socio-economic processes in a detailed spatio-temporal model (e.g. impact of general conditions/constraints on the choice of technical solutions) such approaches are a prerequisite. For traditional network based system description such an assessment is only feasible with transfer functions/data conversion. One of the main advantages of the "Digital City" are the interfaces and linkage with GIS approaches that can be implemented with ease. Since there is a direct interface, neither data

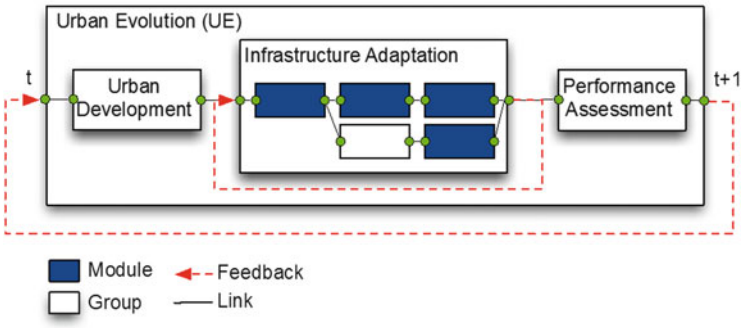

Fig. 7.3 Workflow of an integrated urban environment used in DAnCE4Water

conversion nor generalization (loss of spatial information) is required. The software implementation of concepts can be realized with fewer efforts which offer more opportunities to model interactions. The "Digital City" represents an easier way to model spatial correlations of different technical systems. Primarily in the context of decentralized solutions there are strong linkages between the drainage efficiency, groundwater recharge and high water impacts. The "Digital City" meets requirements of upcoming modelling tasks such as efficient integration of population models, but has yet not been applied.

7.2.4 Assessment Tools and Applications

As discussed by [6] traditional GIS systems are unsuited to model dynamic urban systems due to their limitations to represent time. Therefore, to model the evolution of a city spatially explicitly, new software tools are required. The open source software tool DynaMind [63] provides such a modelling environment to create dynamic urban simulations. Like in GIS the urban system is represented with simple geometric objects (nodes, edges, faces) and raster data. Linking of these data enables the representation of complex objects like buildings or combined drainage networks. These objects are altered by means of data encapsulated modules. To create a module, DynaMind provides easy to use interfaces (C++ and Python) to accessed/modified spatial data during the run time. DynaMind comes already with a set of modules for data import/export and basic GIS functionality (spatial joining, etc.) as well as more complex modules that enable the procedural generation of parcels, buildings or sewer and drainage systems [64]. It also provides interfaces to external hydraulic solvers like SWMM [51] and CityDrain3 [4]. These modules can be linked together to describe a complex workflow in the urban environment. Figure 7.3 conceptually shows the workflow of an application [49] to describe the evolution of the urban environment and its water infrastructure.

DynaMind enables the procedural evolution of cities and their water infrastructure under numerous future scenarios to identify possible development strategies.

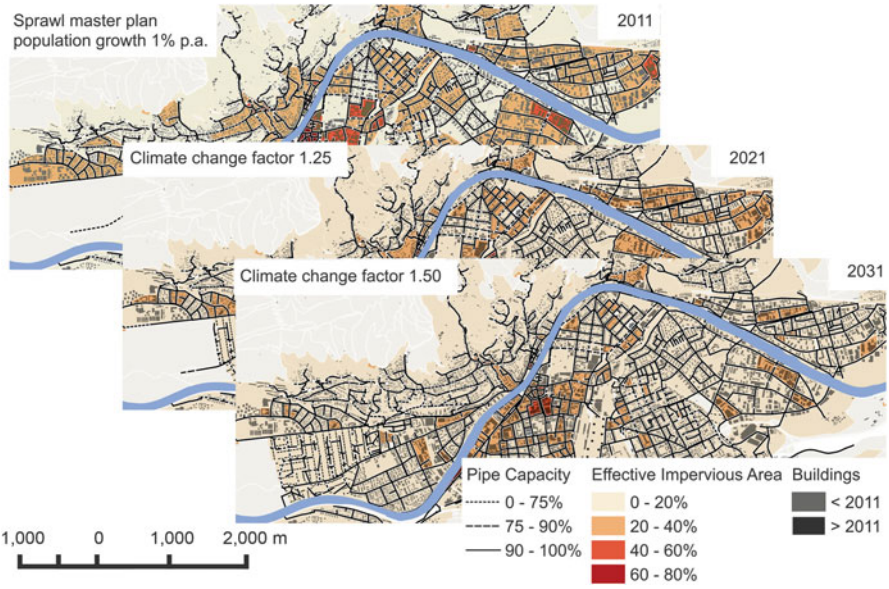

Fig. 7.4 Procedural evolution of Innsbruck, Austria with DynaMind

This can be used to test, e.g. the robustness of a climate change adaptation strategy. Figure 7.4 exemplarily shows one out of 1,200 realizations for the City of Innsbruck, Austria.

7.3 Utilization of Multicore Facilities in Software for Simulating Complexity and Dynamics in Urban Water Management

In the following section different strategies to improve the computational performance of urban water models are presented. This is required to take advantage of recent developments in information technologies as the development of multicore processors to deal with upcoming challenges for urban water systems.

7.3.1 Requirements for Simulations in Urban Water Management

Urban water management requires computer simulations in time (timeframe ranging typically from hours to years) and space (one to three dimensions). The models

typically include both physical/biochemical process descriptions as socio-economy considerations of water infrastructure planning and operation. Urban water models have to be calibrated and validated on measurement data and parameter values have to be determined during model calibration. This process is a mathematical optimization problem aiming to minimize the deviations between measured data and model output [26]. This means that multiple model runs are required before a model can be used in any planning process. With models and their applications becoming more and more complex by either tackling processes at a detailed spatial scale, simulating whole urban environments or performing numerous model runs for scenario or uncertainty studies, boundaries of traditional numerical solutions are reached.

For example, comprehensive simulation studies to determine the uncertainty bounds of model outputs (expressed as confidence intervals) require between 1,000 and 30,000 iterations [11]. Currently such studies are only possible for relatively simple models with a short model runtime (e.g. conceptual models with coarse spatial resolution). Uncertainties of more complex models are usually expressed in scenario uncertainties investigating only a limited number of different scenarios. Such scenarios can be future conditions as impact of climate change or urban development [24], or parameter scenarios [30]. Therefore for each analysed scenario one model run is required.

Depending on the application, different strategies for performance improvement are possible. One possibility is to try to reduce the number of required iterations by improving the calibration/uncertainty algorithm, e.g. by reducing the parameter space which has to be investigated or by improving parameter sampling strategies [10]. Another possibility is to try to reduce the computational time of the iterations by different parallelization strategies ranging from batch level parallelization to model level parallelization [32]. In the following different methods of performance improvements are presented.

7.3.2 Model Level Parallelization

With batch level parallel strategies, a high factor of scalability and efficiency can be achieved (an overview of batch level parallelism can be found in Sect. 7.3.3). Nevertheless, in certain scenarios batch level parallelism is not an option or cannot be used because of certain constraints. In such scenarios parallelization must be targeted at deeper levels. The parallelization level discussed in the following subsections is based on the models itself. Performance enhancements in this layer also benefit users of single model call scenarios.

Parallelization at the model level is typically more involved than batch level parallelization. This is because knowledge of the internal mathematical procedure is necessary. Changes to the source code of the model, which could potentially introduce new defects especially in the case of parallel and concurrent programming, are needed. It is often the case that the current mathematical formulation

or programming model does not allow to parallelize the model. The biggest obstacle in model level parallelization is that changes to the source code are needed and therefore the source code must be available. This is different to batch level parallelization where the whole application is treated as a black box and can be called from the operating system level. In this case no changes of the model itself are needed.

The following sections show three different scenarios of model level parallelizations. Each of them shows a different approach of parallelization which makes them very interesting candidates for describing model level parallelizations. The first one is the storm water management model (SWMM) from the US-EPA for hydrodynamic sewer modelling. It is used for urban rainfall run-off simulations. The second model is EPANET, again from the US-EPA. It is used for water distribution network simulations. Models from the US-EPA are publicly funded and therefore the source code is open source. The third one is CityDrain3 for conceptual sewer modelling. With its simplified mathematical formulations of an urban drainage system it is possible to run long-term effect simulations of urban drainage systems (several decades) in a short manner of simulation run-time.

7.3.2.1 Parallel Flow Routing in SWMM 5.0

Due to its open source code and robust model implementation SWMM 5.0 is a very popular tool for engineers and scientists in the field of urban drainage modelling [51]. SWMM solves the 1D shallow water equations for flow routing in sewers—also known as the Saint Venant Equations (SVE) [52]. Parallelization of this model was imagined to be very complex. Reason for this was that the complexity of the SVE did not allow to outline a parallel algorithm implementation beforehand. The second reason was that the code was totally unknown and that it was ported from Fortran. Further it has a long history of revisions and bug fixes.

With these preconditions a very pragmatic approach for parallelization was chosen. The first step was to find the code segments that contribute the most CPU time. A profiling tool showed that the method *findConduitFlow*, responsible for calculating flow through the conduits using a finite difference scheme for solving the SVE, takes the most time. This function is called for every conduit in the system in a loop. Because the order of calculations for the conduit was not critical (the order was as taken from the input file) it seemed as if the calculations were independent and therefore a possible candidate for parallelization. After a review of the mathematical formulations it was clear that the flow was calculated based on boundary conditions upstream and downstream. These boundary conditions are calculated beforehand and therefore the loop around *findConduitFlow* could be and was parallelized.

After several iterations of finding and fixing concurrent memory accesses, introduced by the parallelization, a speedup of around ten on a twelve core machine was achieved. Contrary to initial estimates and despite the uncertain preconditions of the project very good results were achieved in the manner of weeks.

7.3.2.2 Implementation of Parallel Solvers in EPANET 2.0

The EPANET model for the calculation of water distribution systems is based on a graph of nodes with a certain demand and links (pipes) with a corresponding roughness of the represented pipe. Together with reservoirs and tanks as boundary conditions a system of non-linear equations is formulated in a Jacobian matrix and solved using the iterative Newton–Raphson method. The pressure at each node is the result of such a simulation. The pressure of the node influences the flow through the pipes and vice versa. At each iteration step the Jacobian matrix needs to be solved until pressure and pipe flow are stable [50].

Solving of the Jacobian matrix is the most time demanding task in EPANET. Profiling assured this although the updating of the coefficients, which involves a lot of pipe flow calculations, takes more time than expected. A lot of fast and parallel solvers, even for graphical processing units (GPUs), are available for solving such symmetric positive definite systems. Speeding up EPANET was imagined to be as easy as replacing the hand crafted old solver with a call to a new parallel and highly optimized one. Because such systems are highly parallel a GPU solver was targeted.

Seven solvers, including parallel sparse direct and iterative solvers for multicore CPUs and many-core GPUs, were tested on a range of artificial and real world water distribution networks. The outcome of this research is that the solver currently implemented in EPANET, a solver that was published in a book 32 years ago [14], is still the fastest one.

Linear systems from graphs are typically sparse. The algorithmic complexity of a sparse solver does not only depend on the problem size, which is the case for dense solvers. The complexity depends on the sparsity and the sparse pattern of the problem. Systems from water distribution networks, although, are very sparse. The ratio between the size of the system and the number of non-zeros is typically around two. The fact that water distribution systems are very sparse and typically very small, dimensions in the range of 10^4, makes them not a good target for high performance solvers which aim at systems that begin at dimensions of 10^6.

7.3.2.3 CityDrain3: Parallel Conceptual Sewer Modelling

CityDrain3 (CD3) is the successor of CITY DRAIN II (CD2) a very popular conceptual integrated urban drainage modelling (IUDM) toolkit. Although CD2 is, as SWMM, a simulation toolkit for urban drainage modelling (UDM), the modelling approach is very different. CD2 uses a lumped, conceptual cause–effect approach. CD2 is used for long-term simulation for which such a modelling approach is favoured due its lower computational requirements.

CD2 was implemented using Matlab/Simulink access to the internal simulation core and therefore parallelization of it was not possible. Because of this and the fact that a CD2 version free of Matlab/Simulink has additional advantages, it was rewritten into CD3 which follows the same modelling principles but uses C++ as its implementation base.

In CD3 the wastewater cycle is modelled as a directed acyclic graph where each node represents an element of the wastewater cycle and links represent data/flow transfer between nodes. Links have therefore no computational aspects assigned. A node can be e.g. a sewer, a catchment, a wastewater treatment plant or a river stretch. Because of the conceptual nature the precondition of a node is the outflow of its upstream connected nodes. A parallelization strategy in the same manner as in SWMM is therefore not possible.

Several strategies were implemented to exploit parallelization in such conceptual IUDM simulations. The first one exploits the fact that a wastewater system is often in the shape of a tree with lots of independent streams that eventually merge at the WWTP. At each source, typically a catchment, a thread can be started. Although this offers a way of parallelization it is very limited with regards to parallel workload. A second strategy exploits the fact that parallelization can be pipelined through the time steps. This is possible because the length of a time step is fixed and known before hand [4].

The rewrite of CD2 from an interpreted general purpose simulation framework into a tailor made, native and parallelized rewrite in C++ made CD3 up to 40 times faster.

7.3.3 Performance Improvement by Batch-Level Parallelism

In urban water management modelling the chosen parallelization technology and especially the level of which parallelization is realized in the source code is strongly depending on the modelling aim and existing modelling software used. In the previous Sect. 7.3.2 already existing and newly developed software tools and their parallelization strategy were described. Here the performance improvement according to computational efficiency and speedup on multi/many-core systems within one model simulation run was the motivation.

Another interesting research field in urban water management is to assess the sensitivity of system components according to specific performance indicators. Under the scope of this book following two different applications can be identified which are:

- Assess the sensitivity and impact of a model parameter (e.g. roughness of conduits within a hydrodynamic sewer model) on model simulation results (e.g. water level at junctions) [25].
- Assess the vulnerability and consequences of existing systems according to hazardous events (e.g. pipe bursts within a water supply system due to deep temperatures [35] or a sewer pipe collapse due to deterioration [27, 34]). Moreover cascading effects can be assessed where the first hazardous event (e.g. failure of a source and therefore change in pressure regime) is the trigger for another hazardous event (e.g. the pipe burst) [57].

From the programmers and model developers perspective this application can be realized by (1) modelling the needed adaption within an original model (e.g. pipe burst of one specific pipe), (2) simulate the model and (3) access the consequences of the adaption with global performance indicators (PI) by comparing simulation results from the adapted model with the original model. Repeating steps (1)–(3) for all components within a system, vulnerable/senstive sites according to a specific hazard can be identified.

One might immediately realize that testing each component within a system against such hazardous events needs many different model runs. As each test is independent from each other all model simulations can be run in parallel moreover this parallelism is in theory embarrassingly parallel (batch-level parallelism). Many existing model software products in this field (e.g. EPANET2 and SWMM5) have grown over time and therefore often have no parallel implementation. In this kind of application one huge advantage is that the original model simulation code can be used and at the same time multicore systems can be utilized. The only limiting factor is data communication during the evaluation of all PIs which leads to a non-linear speedup.

Performance tests showed that this parallelization strategy in combination with the software presented earlier has a speedup of 12 by using twelve threads at batch-level and one thread at the model level. By using one thread at the batch-level and 12 threads at model level a speedup of only four can be achieved. More investigations with other model simulation software products (e.g. EPANET) showed that this parallelization strategy is a good alternative to speeding up the previously described applications. Moreover if the model software comes already with a parallel implementation (e.g. parallel version of SWMM 5.0, Sect. 7.3.2.1—Model level parallelism) and at the same time parallel executing these models, investigations showed that the best CPU-load efficiency can be achieved by only applying parallelism at the batch-level [32].

7.4 SPH: An Alternative Numerical Method to Explore Fluid Phenomena

7.4.1 Motivation and Aim

SPH is a computational fluid dynamics (CFD) method for solving fluid flows. In Layman's terms in SPH a fluid is represented by a myriad of small spheres which are referred to as particles. As the movement of particles is governed by the continuum equation of fluid dynamics, the overall picture resembles the true hydrodynamic phenomena. By statistically weighting the influence of each particle's neighbourhood (see Fig. 7.5), the equations of motion reduce to a set of ordinary differential equations which are easy to understand and implement [37].

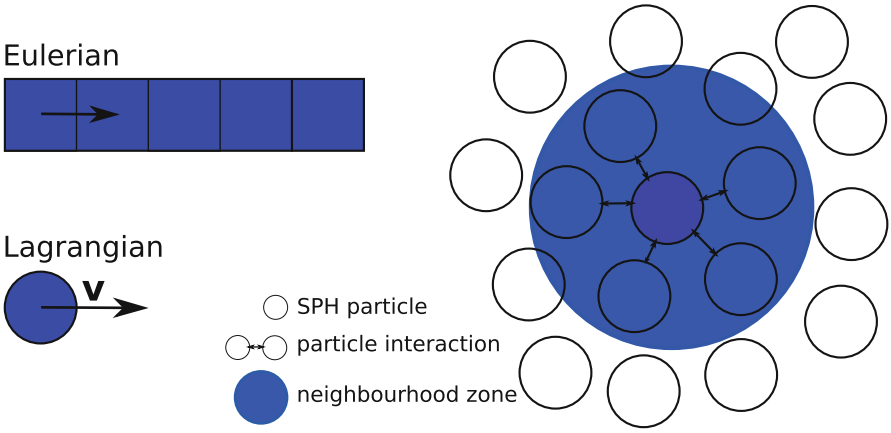

Fig. 7.5 Comparison between Eulerian and Lagrangian movement (*left*) for an SPH particle with five neighbours (*right*)

SPH was introduced, at the same time, by [15, 28] to solve astrophysical problems. In contrast to conventional grid based CFD methods, SPH is a fully Lagrangian meshless method such that each particle is free to move and carries physical parameters like mass, velocity and density. SPH has been applied to a wide range of problems in the fields of material science, oceanography and volcanology. However, the core application area of SPH is fluid mechanics, in particular transport phenomena [61], free surface [16, 38] and multiphase flows [5].

Compared to conventional CFD methods, SPH has various advantages owing to its Lagrangian nature (see Fig. 7.5). Namely, advection is treated exactly and conservation laws of mass, linear respectively angular momentum and energy are satisfied. In addition, SPH is a physically correct numerical scheme and can be formulated without empirical parameters such that the effort of calibration is minimized. Hence, once the SPH model is set up, it is reasonably simple to account for complex hydrodynamic phenomena like multiphase flow and transport of solid objects. However, the physical correctness of the method requires comparable large computational demand, which can be reduced by relaxing physical requirements. For example, for some practical applications it is sufficient to approximate incompressible fluids by slightly compressible analogues. Through this approach, which is referred to as weakly compressible SPH, the solution of a pressure Poisson equation is substituted by a simple equation of state and hence computational cost is significantly reduced.

Nonetheless, further reduction in simulation time is required for practical applications of SPH. This is achieved by parallelization of the method, which is simplified by the fact that the numerical scheme itself is highly parallel. SPH has already been implemented on highly parallel computing devices like graphics processing units [18]. In particular, an efficient parallel solution for finding neighbours, which is the process that requires most computational power, was found [17, 22].

7.4.2 SPH for Sewer Modelling

Over the last three decades, urban drainage modelling evolved from simple models to high complexity [47]. While state of the art methods for one dimensional hydrodynamic simulations in pipe networks exist, recently more complicated CFD methods have been applied to simulate specific structures [12]. However, modelling of pollution transport and sewer solids is still an unresolved issue. Both deterministic and conceptual models failed to convincingly explain the underlying phenomena (see e.g. [9]). In this respect there is a perspective for a novel, deterministic numerical method as represented by SPH.

SPH has several advantages which makes it a viable alternative to solving the simplified St. Venant equations, which are used in state of the art sewer hydraulic simulation models. First of all, the method is inherently three dimensional, while a reduction to two dimensions is simple but only motivated by limitations in computational power. Therefore, complex hydraulic structures can be easily modelled. Secondly, as the continuum equations of fluid mechanics can be used as governing SPH equations, it is possible to model pressure effects in pipes which are currently bypassed by the Preissman Slot [46]. Thirdly, extension of SPH to multiphase flow and solid transport phenomena is much simpler than the conventional Eulerian methods. Especially, the application of SPH to the latter field gives a whole new angle to tackle the problem of pollution transport in drainage systems. However, the challenge of huge computational burden for simulating SPH sewers remains. In particular, it is unclear whether the SPH method is applicable for real world pipe networks, but stringent parallel coding and use of novel technology like graphics processing units could open a pathway. Based on present results we foresee a huge potential of the method, whilst significant obstacles still need to be tackled.

7.4.3 SPH for Wastewater Treatment Simulations

As with sewer modelling, multiphase and transport phenomena are the key challenges for numerical simulations of wastewater treatment processes. Since conventional CFD methods are not particularly suitable for these problems, currently the fluid dynamics are neglected in the well-established activated sludge models (ASM) [21]. Even though the biological kinetics are successfully modelled with this approach, local effects are neglected. Hence, a wastewater tank is assumed to be completely mixed at all times and therefore the hydraulics are effectively uncoupled from biological processes. Whilst the development of SPH is not yet advanced enough to accurately simulate air, sludge and water phases at the same time, it is required to separate the discussion of aeration and sedimentation tanks.

Aeration processes can be modelled as two-phase air water flow, but this is challenging since huge density differences cause rapid movement at the phase

interface which gives rise to instabilities in the SPH formalism. Recently, a simple two-phase SPH algorithm has been proposed to cure this problem [39]. In combination with adding an oxygen concentration parameter, which is evolved by an advective diffusion equation [1], the local dissolved oxygen concentration is accounted for correctly. As this key parameter governs the differential equations of the ASM model, the local oxygen concentration provides a coupling interface between the local hydraulics and the biological kinetics. This approach improves the present ASM model and is the first step to advance to a full-scale three-phase model.

Similar to aeration tanks, sedimentation processes are well described by two-phase SPH. In contrast to air water flow the solid phase is not modelled as a weakly compressible fluid phase, but sediments are considered as a slightly compressible pseudo-Newtonian fluid. Thereby, the Newtonian constitutive equation has to be modified [13] and a yield criterion is required to correctly account for sediment-fluid scouring at the phase interface. Both the Mohr-Coulomb and the Drucker-Prager criterion yield satisfactory results, but the latter method is slightly preferred [13].

7.5 Conclusions and Outlook

Scientific computing in urban water management is widespread. This chapter mainly summarizes current research activities at the Unit of Environmental Engineering within the framework of the research center “Computational Engineering” at the University of Innsbruck focusing on currently challenging issues. The first topic of the chapter reviews increasing complexity of assessing urban water systems respectively describes the shift to city scale analysis. In particular it is outlined how increasing computer power over the last decades changed the way of how system analysis in urban water management is performed. In traditional engineering approaches the complexity of the problem is reduced in order to obtain an applicable mathematical problem description. For that an in depth understanding of the that particular (sub-)system is necessary. The application of such a description but also simulation models can usually be applied in research and practice. Increasing computer power enables us to integrate and couple models with more and more complexity. Different existing models and extensive amount of data can be used for comprehensive analysis which produces an effusive amount of results data. With that the complexity of the engineering task is shifted to analysis of the result data. Such tasks are therefore usually research applications. Nonetheless, such analysis deepens the system understanding and helps also to obtain system coherences which have been usually overlooked. The second topic demonstrates the utilization of multicore facilities in software for simulating such complex systems related to urban water management. In that section it is outlined which parallelization approaches are required in order to speed up different kinds of simulation models in urban water management. This work aims to reduce computation time for existing research tasks and also practical applications. The third topic discusses alternative

numerical methods SPH to explore fluid phenomena in urban water management. That approach can easily be extended towards multiphase flow phenomena, solids transport and bioconversion processes and shows therefore great potential in future. Thus SPH could potentially be the core numerical engine to simulate fluxes and processes in the complete water infrastructure on a very detailed level.

Acknowledgements This work was funded by the Austrian Science Fund (FWF) in the project *DynaVIBe* P23250, project *DynAlp* funded by the Austrian Climate and Energy Fund (project number B175093) and by the EU-Framework-programme *Prepared: enabling change* under the contract number 244232. The authors gratefully acknowledge the financial supports.

References

1. Aristodemo, F., Federico, I., Veltri, P., Panizzo, A.: Two-phase SPH modelling of advective diffusion processes. Environ. Fluid Mech. **10**, 451–470 (2010)
2. Barth, M., Hennicker, R., Kraus, A., Ludwig, M.: DANUBIA: an integrative simulation system for global change research in the upper Danube basin. Cybern. Syst. Int. J. **35**(7), 639–666 (2004)
3. Brown, R.R., Keath, N., Wong, T.H.F.: Urban water management in cities: historical, current and future regimes. Water Sci. Technol. **59**(5), 847–855 (2009)
4. Burger, G., Fach, S., Kinzel, H., Rauch, W.: Parallel computing in conceptual sewer simulations. Water Sci. Technol. **61**(2), 283–291 (2010)
5. Colagrossi, A., Landrini, M.: Numerical simulation of interfacial flows by smoothed particle hydrodynamics. J. Comput. Phys. **191**, 448–475 (2003)
6. Crooks, A., Castle, C., Batty, M.: Key challenges in agent-based modelling for geo-spatial simulation. Comput. Environ. Urban Syst. **32**, 417–430 (2008)
7. DHI: MIKE URBAN Users Manual (2008)
8. Doglioni, A., Primativo, F., Laucelli, D., Monno, V., Khu, S.T., Giustolisi, O.: An integrated modelling approach for the assessment of land use change effects on wastewater infrastructures. Environ. Model. Softw. **24**(12), 1522–1528 (2009)
9. Dotto, C.B.S., Kleidorfer, M., Deletic, A., Fletcher, T.D., McCarthy, D.T., Rauch, W.: Stormwater quality models: performance and sensitivity analysis. Water Sci. Technol. **62**(4), 837–843 (2010)
10. Dotto, C.B.S., Kleidorfer, M., Deletic, A., Rauch, W., McCarthy, D.T., Fletcher, T.D.: Performance and sensitivity analysis of stormwater models using a Bayesian approach and long-term high resolution data. Environ. Model. Softw. **26**(10), 1225–1239 (2011)
11. Dotto, C.B.S., Mannina, G., Kleidorfer, M., Vezzaro, L., Henrichs, M., McCarthy, D.T., Freni, G., Rauch, W., Deletic, A.: Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. Water Res. **46**(8), 2545–2558 (2012)
12. Fach, S., Sitzenfrei, R., Rauch, W.: Determining the spill flow discharge of combined sewer overflows using rating curves based on computational fluid dynamics instead of the standard weir equation. Water Sci. Technol. **60**(12), 3035–3043 (2009)
13. Fourtakas, G., Rogers, B.D., Laurence, D.: Modelling sediment resuspension in industrial tanks using SPH on GPUs. In: Proceedings of International Conference SPHERIC SPH workshop, pp. 310–316, Trondheim (2013)
14. George, A., Liu, J., Ng, E.: Computer Solutions of Sparse Linear Systems. Academic, Orlando (1994)
15. Gingold, R.A., Monaghan, J.J.: Smoothed particle hydrodynamics: theory and application to non-spherical stars. Mon. Notices R. Astron. Soc. **181**, 375–389 (1977)

16. Gomez-Gesteira, M., Rogers, B.D., Dalrymple, R.A., Crespo, A.J.C.: State-of-the-art of classical SPH for free-surface flows. J. Hydraul. Res. **48**, 6–27 (2010)
17. Goswami, P., Schlegel, P., Solenthaler, B., Pajarola, R.: Interactive SPH simulation and rendering on the GPU. In: Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA '10, pp. 55–64. Eurographics Association, Aire-la-Ville (2010)
18. Harada, T., Koshizuka, S., Kawaguchi, Y.: Smoothed Particle Hydrodynamics on GPUs. In: Proceedings of Computer Graphics International, pp. 63–70 (2007)
19. Hardy, M.J., Kuczera, G., Coombes, P.J.: Integrated urban water cycle management: the urbancycle model. Water Sci. Technol. **52**(9), 1–9 (2005)
20. Harremoes, P., Rauch, W.: Integrated design and analysis of drainage systems, including sewers, treatment plant and receiving waters. J. Hydraul. Res. **34**(6), 815–826 (1996)
21. Henze, M., Gujer, W., Takashi, M., Loosdrecht, M.V.: Activated Sludge Models ASM1, ASM2, ASM2d and ASM3. IWA, London (2000)
22. Kalojanov, J., Slusallek, P.: A parallel algorithm for construction of uniform grids. In: HPG '09: Proceedings of the 1st ACM conference on High Performance Graphics, pp. 23–28. ACM, New York (2009). doi:10.1145/1572769.1572773
23. Kleidorfer, M., Rauch, W.: An application of austrian legal requirements for cso emissions. Water Sci. Technol. **64**(5), 1081–1088 (2011)
24. Kleidorfer, M., Möderl, M., Sitzenfrei, R., Urich, C., Rauch, W.: A case independent approach on the impact of climate change effects on combined sewer system performance. Water Sci. Technol. **60**(6), 1555–1564 (2009)
25. Kleidorfer, M., Leonhardt, G., Mair, M., McCarthy, D., Kinzel, H., Rauch, W.: Calimero-A model independent and generalised tool for autocalibration. In: 8UDM & 2RWHM the 8th International Conference on Urban Drainage. The 2nd International Conference on Rainwater Harvesting and Management. IWA, Tokyo (2009)
26. Kleidorfer, M., Möderl, M., Fach, S., Rauch, W.: Optimization of measurement campaigns for calibration of a conceptual sewer model. Water Sci. Technol. **59**(8), 1523–1530 (2009)
27. Kleidorfer, M., Möderl, M., Tscheikner-Gratl, F., Hammerer, M., Kinzel, H., Rauch, W.: Integrated planning of rehabilitation strategies for sewers. Water Sci. Technol. **68**(1), 176–183 (2013)
28. Lucy, L.B.: A numerical approach to the testing of the fission hypothesis. Astron. J. **82**, 1013–1024 (1977)
29. Ludwig, R., Mauser, W., Niemeyer, S., Colgan, A., Stolz, R., Escher-Vetter, H., Kuhn, M., Reichstein, M., Tenhunen, J., Kraus, A., Ludwig, M., Barth, M., Hennicker, R.: Web-based modelling of energy, water and matter fluxes to support decision making in mesoscale catchments–the integrative perspective of glowa-danube. Phys. Chem. Earth Parts A/B/C **28**(14–15), 621–634 (2003)
30. Mair, M., Sitzenfrei, R., Kleidorfer, M., Möderl, M., Rauch, W.: Gis-based applications of sensitivity analysis for sewer models. Water Sci. Technol. **65**(7), 1215–1222 (2012)
31. Mair, M., Rauch, W., Sitzenfrei, R.: Improving Incomplete Water Distribution System Data. Procedia Engineering (2013)
32. Mair, M., Sitzenfrei, R., Kleidorfer, M., Rauch, W.: Performance improvement with parallel numerical model simulations in the field of urban water management. J. Hydroinformatics. doi:10.2166/hydro.2013.194. URL http://www.iwaponline.com/jh/up/pdf/jh2013287.pdf
33. Mankad, A., Tapsuwan, S.: Review of socio-economic drivers of community acceptance and adoption of decentralised water systems. J. Environ. Manag. **92**(3), 380–391 (2011)
34. Möderl, M., Kleidorfer, M., Sitzenfrei, R., Rauch, W.: Identifying weak points of urban drainage systems by means of VulNetUD. Water Sci. Technol. **60**(10), 2507–2513 (2009)
35. Möderl, M., Hellbach, C., Sitzenfrei, R., Mair, M., Lukas, A., Mayr, E., Perfler, R., Rauch, W.: GIS based applications of sensitivity analysis for water distribution models. In: World Environmental and Water Resources Congress 2011, pp. 129–136. American Society of Civil Engineers, Palm Springs (2011)
36. Moglia, M., Perez, P., Burn, S.: Modelling an urban water system on the edge of chaos. Environ. Model. Softw. **25**(12), 1528–1538 (2010)

37. Monaghan, J.J.: Smoothed particle hydrodynamics. Ann. Rev. Astron. Astrophys. **30**, 543–574 (1992)
38. Monaghan, J.J.: Simulating free surface flows with SPH. J. Comput. Phys. **110**, 399–406 (1994)
39. Monaghan, J.J., Rafiee, A.: A simple SPH algorithm for multi-fluid flow with high density ratios. Int. J. Numer. Methods Fluids **71**, 537–561 (2013)
40. ÖWAV-RB 19: Guideline for design and construction of combined sewer overflows. Österreichischer Wasser- und Abfallwirtschaftsverband, Wien (1987)
41. ÖWAV-RB 19: Guideline for the design of combined sewer overflows. Österreichischer Wasser- und Abfallwirtschaftsverband, Wien (2007)
42. Pahl-Wostl, C.: Information, public empowerment, and the management of urban watersheds. Environ. Model. Softw. **20**(4), 457–467 (2005)
43. Pahl-Wostl, C.: The implications of complexity for integrated resources management. Environ. Model Softw. **22**(5), 561–569 (2007)
44. Parker, P., Letcher, R., Jakeman, A., Beck, M.B., Harris, G., Argent, R.M., Hare, M., Pahl-Wostl, C., Voinov, A., Janssen, M., Sullivan, P., Scoccimarro, M., Friend, A., Sonnenshein, M., Barker, D., Matejicek, L., Odulaja, D., Deadman, P., Lim, K., Larocque, G., Tarikhi, P., Fletcher, C., Put, A., Maxwell, T., Charles, A., Breeze, H., Nakatani, N., Mudgal, S., Naito, W., Osidele, O., Eriksson, I., Kautsky, U., Kautsky, E., Naeslund, B., Kumblad, L., Park, R., Maltagliati, S., Girardin, P., Rizzoli, A., Mauriello, D., Hoch, R., Pelletier, D., Reilly, J., Olafsdottir, R., Bin, S.: Progress in integrated assessment and modelling. Environ. Model. Softw. **17**(3), 209–217 (2002)
45. Polebitski, A.S., Palmer, R.N.: Seasonal residential water demand forecasting for census tracts. J. Water Res. Plann. Manag. **136**(1), 27–36 (2010)
46. Preissmann, A., Cunge, J.: Calcul des intumescences sur machines electroniques. In: Proceedings of the 9th IAHR Congress, Dubrovnik, pp. 656–664 (1961)
47. Rauch, W., Bertrand-Krajewski, J.L., Krebs, P., Mark, O., Schilling, W., Schütze, M., Vanrolleghem, P.A.: Deterministic modelling of integrated urban drainage systems. Water Sci. Technol. **45**, 81–94 (2002)
48. Rauch, W., Kleidorfer, M., Fach, S.: From the pencil to the processor: change in the modelling of urban sewerage systems. Österreichische Wasser- und Abfallwirtschaft **62**(3–4), 43–50 (2010)
49. Rauch, W., Bach, P.M., Brown, R.R., Deletic, A., Ferguson, B., de Haan, J., McCarthy, D.T., Kleidorfer, M., Tapper, N., Sitzenfrei, R., Urich, C., de Haan, F.J.: Modelling transitions in urban drainage management. In: Proceedings of the Ninth International Conference on Urban Drainage Modelling (2012)
50. Rossman, L.A.: EPANET Version 2 Users Manual. US Environmental Protection Agency (USEPA), Cincinnati (2000)
51. Rossman, L.A.: Storm Water Management Model: User's Manual Version 5.0. National Risk Management Research Laboratory: U.S. Environmental Protection Agency, Cincinnati (2010)
52. Saint-Venant, A.D.: Theorie du mouvement non permanent des eaux, avec application aux crues des rivieres et a lintroduction de marees dans leurs lits. Comptes rendus des seances de lAcademie des Sciences **36**, 174–154 (1871)
53. Sitzenfrei, R., Rauch, W.: From water networks to a "Digital City": a shift of paradigm in assessment of urban water systems. In: 12th International Conference on Urban Drainage. Porto Alegre (2011)
54. Sitzenfrei, R., Rauch, W.: Investigating Transitions of Centralized Water Infrastructure to Decentralized Solutions an Integrated Approach. Procedia Engineering (2013)
55. Sitzenfrei, R., Fach, S., Kinzel, H., Rauch, W.: A multi-layer cellular automata approach for algorithmic generation of virtual case studies: VIBe. Water Sci. Technol. **61**(1), 37–45 (2010)
56. Sitzenfrei, R., Fach, S., Kleidorfer, M., Urich, C., Rauch, W.: Dynamic virtual infrastructure benchmarking: DynaVIBe. Water Sci. Technol. Water Supply **10**(4) (2010)
57. Sitzenfrei, R., Mair, M., Möderl, M., Rauch, W.: Cascade vulnerability for risk analysis of water infrastructure. Water Sci. Technol. **64**(9), 1885–91 (2011)

58. Sitzenfrei, R., Möderl, M., Mair, M., Rauch, W.: Modeling dynamic expansion of water distribution systems for new urban developments. In: World Environmental & Water Resources Congress. American Society of Civil Engineers, Albuquerque (2012)
59. Sitzenfrei, R., Möderl, M., Fritsch, E., Rauch, W.: Schwachstellenanalyse bei Mischwasseranlagen für eine sichere Bewirtschaftung. Österreichische Wasser- und Abfallwirtschaft **64**(3–4), 293–299 (2012)
60. Sitzenfrei, R., Möderl, M., Rauch, W.: Automatic generation of water distribution systems based on GIS data. Environ. Model. Softw. **47**, 138–147 (2013)
61. Tartakovsky, A.M., Meakin, P., Scheibe, D., West, R.M.E.: Simulations of reactive transport and precipitation with smoothed particle hydrodynamics. J. Comput. Phys. **222**, 654–672 (2007)
62. Urich, C., Bach, P.M., Hellbach, C., Sitzenfrei, R., Kleidorfer, M., McCarthy, D.T., Deletic, A., Rauch, W.: Dynamics of cities and water infrastructure in the DAnCE4Water framework. In: 12th International Conference on Urban Drainage. Porto Alegre (2011)
63. Urich, C., Burger, G., Mair, M., Rauch, W.: DynaMind: a software tool for integrated modelling of urban environments and their infrastructure. In: Proceedings of the 10th International Conference on Hydroinformatics HIC 2012 (2012)
64. Urich, C., Sitzenfrei, R., Kleidorfer, M., Rauch, W.: Klimawandel und Urbanisierung: Wie soll die Wasserinfrastruktur angepasst werden? Österreichische Wasser- und Abfallwirtschaft **65**, 82–88 (2013)

Chapter 8
Numerical Simulations in Hydraulic Engineering

R. Gabl, B. Gems, M. Plörer, R. Klar, T. Gschnitzer, S. Achleitner, and M. Aufleger

Abstract The main focus of the chapter is to present various case studies, showing the link between Computational Fluid Dynamics (CFD) and traditional scale model tests in the laboratory. The goal is to illustrate the possibilities and limitations when coupling these two different methods in the context of hydraulic engineering applications. The topics range from hydraulic investigations where numerical simulations are a vital tool for model validation (optimisation and quantification of local head losses, the capacity of a spillway and as a third example impulse waves caused by an avalanche), to modelling of debris flow and log jam processes, including bed load transport issues. The use of such hybrid approaches can contribute to cost-saving and realisation of more complex investigations in shorter time.

8.1 Introduction

Numerical methods in the field of Computational Fluid Dynamics (CFD) are a very powerful and important tool for hydraulics and hydraulic engineering. Case dependent 1D-, 2D- and 3D-numerical approaches are applied. Various academic and commercial codes are available.

Subsequently, two different 3D-numerical software solutions will be presented: ANSYS-CFX and FLOW-3D. Both are commercial codes and each of them has its special application area and advantages. Open source codes are rare in this area, but for example OpenFOAM or TELEMAC-3D could be a suitable choice.

R. Gabl (✉) • B. Gems • M. Plörer • R. Klar • T. Gschnitzer • S. Achleitner • M. Aufleger
Unit of Hydraulic Engineering, University of Innsbruck, Technikerstr. 13, A6020 Innsbruck, Austria
e-mail: Roman.Gabl@uibk.ac.at; Bernhard.Gems@uibk.ac.at; Manuel.Ploerer@uibk.ac.at; Robert.Klar@uibk.ac.at; Thomas.Gschnitzer@uibk.ac.at; Stefan.Achleitner@uibk.ac.at; Markus.Aufleger@uibk.ac.at

G. Hofstetter (ed.), *Computational Engineering*, DOI 10.1007/978-3-319-05933-4_8,

© Springer International Publishing Switzerland 2014

In general, the Navier–Stokes-equation is the foundation of the numerical modelling concept. Because of the necessary huge amount of calculation power, direct numerical simulations (DNS) of this equation are only possible for special research problems. For most engineering cases, each of the values that has to be calculated is split into a mean part (time average) and its fluctuation. To reduce the effect of the non-linearity, only the mean values of these equations are solved. In these so called Reynolds-averaged-Navier–Stokes (RANS) equations, the average of the fluctuation is zero, but the Reynolds-stress tensors have to be added. To close this system, further equations have to be provided by the turbulence model [58, 73]. Therefore, the complete turbulence is represented by the turbulence model and all results are mean values. In between of DNS and RANS the large-eddy simulation (LES) could be classified. Therewith, only a sub-grid scale-model is used and every bigger Eddy spectrum is resolved [32].

In contrast, the 1D-numerical solutions are based on a mean value of a section on a flow path. To solve the governing mass and momentum equations, the method of characteristics (MOC) is a very popular and often used approach for pipeline systems [27, 28]. Hydraulic System [8] and WANDA [10] are the two most commonly used software for transient pipe flow (water hammer, surge chamber oscillation) at the Unit of Hydraulic Engineering at the University of Innsbruck. The finite difference (FD) and finite volume (FV) methods are alternative ways [27]. Amongst others, the software HEC-RAS (Hydrologic Engineering Centers River Analysis System), provided by the US Army Corps of Engineers, uses the FD-method to calculate free surface flow. The software includes unsteady flow, sediment transport and water temperature modelling [23, 70].

Based on the 1D-approach, each value is calculated for one section. For the two presented case studies in Sects. 8.5 and 8.6, a 2D-numerical software has to be used. With the help of the commercial software HYDRO_AS-2D pure hydraulic investigations are conducted. In case of bed load transport processes, the software HYDRO_GS-2D (Version 3.0) is applied. Both are based on the Finite Volume Method (FV) to solve the Shallow Water Equations [53]. Morphological changes and bed load transport are modelled using a multi-fraction multi-layer approach. Mass balancing is performed between three layers: a top mixing layer, an intermediate subsurface layer and a bottom layer. The grain size distributions in the mixing and subsurface layers are determined according to *Hirano* [36]. Bed load transport is calculated with a multi-fraction application of the Meyer-Peter and Müller equation [48] including a hiding function as introduced by *Hunziker* [38,39]. The fast and reliable code is able to describe grain sorting and bed armouring processes as well as embankment collapses. The effect of transversal bed slopes and secondary flows on the sediment transport capacity is modelled by well-known empirical approaches. To create a case study model a multitude of different parameters and boundary conditions has to be defined, e.g. a three-dimensional surface mesh, the total and skin roughness coefficients and the initial grain size distributions for each node and layer based on field measurements. The 2D-numerical analysis has successfully been applied to assess such different aspects as for example the

deposition and flushing of man-made reservoirs, bed morphology changes during flood events or the long-term morphodynamic evolution of up to 50 km long alpine river reaches [42, 43].

8.2 Asymmetric Orifice

In general, the main goal of a hydraulic design is to reduce the head losses in the system in an economic way. Therewith, the transported discharge through the pipeline or the energy production can be maximised. But for specific cases, a local head loss is wanted. An example for this could be orifices, which cause a defined amount of local head loss and are added into the flood discharging tunnel in order to dissipate energy [41, 46, 74]. A second application is the measurement of discharge in pipes [1].

Another use for the orifices is to throttle a surge tank of a hydro power plant. Thereby, the surge tank oscillation is limited in a practical way and the needed volume in the chambers of the surge tank can be reduced [28, 40]. These hydraulic parts are sub-classified in (a) symmetric orifice, (b) asymmetric orifice and (c) reverse flow throttle (also known as vortex chamber diode [31]). The last two mentioned types of construction provide a different head loss depending on the flow direction and will be concentrated on.

Every change of the discharge at the turbine has an effect on the flow regime of the power plant. Especially a fast shut down causes a water hammer, which runs up the penstock and is reflected at the next free surface. The pressure waves put a lot more strain on the system than the normal use under static pressure. Thus, the building costs for a long headrace tunnel can be far smaller if a free surface is provided by an added surge tank. In addition, the kinetic energy that differs in the headrace tunnel and the penstock can be compensated by an up- or down-lift of the free surface as potential energy in the surge tank. Hence, a periodically changing flow between the surge tank and the reservoir is initiated, which is a slow mass oscillation in comparison to the fast pressure waves of the water hammer. The movement is damped by the friction in the system and so a new equilibrium state is reached in the hydraulic system [40, 52]. To avoid unstable configurations, different criteria were developed in the past. One is the criterion from Thoma [69], which limits the minimum area of the cross section of the surge chamber.

Modern high-head hydro plants are designed to ensure a free operation management. Even the most disadvantageous cases of multiple changes in the hydraulic system should not lead to an overload. The overall aim for the design is to minimise the required volume in each chamber of the surge tank and to achieve a new equilibrium state in the hydraulic system as fast as possible. For this reason, the asymmetric behaviour of the loss quantity with respect to the flow direction through the throttle is a very useful effect [28]. Therefore, the upward loss is limited so that the head race tunnel is not overstressed. For the reverse flow situation, increased losses are added to the system to throttle the oscillation.

8.2.1 Investigation Area

The presented investigation is part of a research project with the energy producer TIWAG-Tiroler Wasserkraft AG. In addition to the rebuilding of the penstock after nearly 50 years of operation, a new surge tank is added to the high-head power plant Kaunertal (Fig. 8.1). The overall hydraulic system was simulated with the help of a TIWAG in-house 1D-numerical code. This global simulation was refined by 3D-numerical simulations. In this context, three different parts were investigated:

1. The numerical optimised behaviour of the **asymmetric orifice** is also validated with the help of a scale model test.
2. Based on the change from the existing reverse flow throttle [65, 66] to an asymmetric orifice and other boundary conditions, the capacity to store more water in the **upper chamber** has to be increased. Therefore, to the existing upper chamber of the surge tank a new tunnel is added. It is connected to the existing tunnel at two points and so a nearly circular system has been construed. The transient filling and emptying processes including the reflected waves had to be checked. For this free surface problem the 3D-numerical software FLOW-3D was used.
3. The flow conditions in the **connection** of the existing headrace power tunnel to the newly built penstock and the lower chamber of the surge tank are numerically simulated (ANSYS-CFX) to look for further optimisation options.

Further information for the last two points is provided in [19,20]. The presented case study is focused on the quantification and optimisation of an asymmetric orifice. This throttle is placed on top of a 90°-elbow between the vertical shaft and the circular lower chamber of the surge tank (Fig. 8.1). The radius of curvature of the elbow is chosen with 7.0 m and the cross section from a diameter of 5.0 m (lower chamber) to 4.0 m, which is the starting section of the asymmetric orifice. The minimal diameter is 3.1 m and after the throttle, the flow path is expanded to 6.3 m (diameter of the shaft). The flow through the orifice is limited with 140 m^3/s. All investigations are conducted under steady-state conditions. After this investigation, the values are used in the global 1D-numerical simulation as a local head loss coefficient [16].

8.2.2 Basic Equations

For real incompressible fluid the Bernoulli equation:

$$z_1 + \frac{p_1}{\rho \cdot g} + \frac{\alpha_1 \cdot v_1^2}{2 \cdot g} = h_{E1} = \text{constant} = h_{E2} = z_2 + \frac{p_2}{\rho \cdot g} + \frac{\alpha_2 \cdot v_2^2}{2 \cdot g} + h_v \tag{8.1}$$

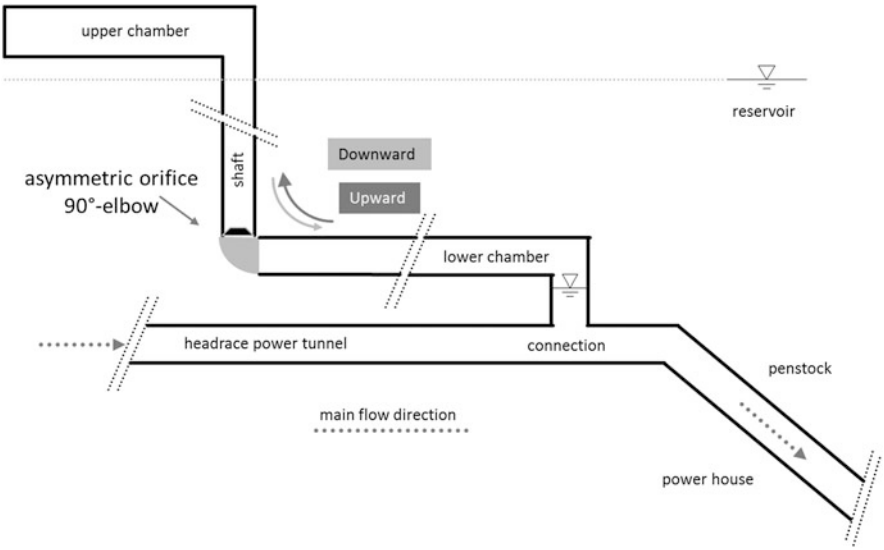

Fig. 8.1 Simplified sketch of the surge tank including the flow directions

states that the total energy h_E is constant along a steady continuous streamline. h_E is the sum of elevation z, pressure head (which is equal to pressure p divided by density ρ and gravity acceleration g) and the kinetic energy based on the velocity v. The indexes 1 and 2 in this equation mark the different points along the streamline [3, 50]. If the sections are numbered in flow direction, h_v is equivalent to the difference between the energy heights h_{E1} and h_{E2}.

The kinetic energy flux coefficients α_1 and α_2 are correction factors for the non-uniformity of the realistic cross-sectional velocity distribution. For normal pipe-flow, the value of α can range between 1.0 and 2.0, the last value representing a fully developed laminar (parabolic) flow field [71]. To calculate the coefficient, the complete velocity profile $v(A)$ has to be known:

$$\alpha = \frac{1}{A \cdot v_{\text{mean}}^3} \int_A v(A)^3 dA \tag{8.2}$$

The measuring of the velocity could not be conducted in the presented scale model test. Because of the high pressure in the model (up to 7 *bar*) no transparent part could be integrated, to use laser Doppler anemometry (LDA) or Particle image velocimetry (PIV) [51]. As an assumption, α_1 and α_2 are assumed to 1.0 [−]. As part of the Post-Processing of the numerical results, these values are calculated and used for the quantification of the local head loss coefficient in both model concepts [11].

The elevations of the sections before and after the orifice differ in nature (Fig. 8.1). For the investigation the symmetry plane is turned into the datum plane

(90° rotation around the axis of the lower chamber). Therewith, z_1 is equal to z_2 and the difference is zero.

The head loss h_v can be split into the part, which is caused by friction between the two sections, and the actual local head loss of the structure. The observed area is situated as closely as possible around the orifice, so friction is comparably very small and can be disregarded. Based on this assumption, the loss h_v in Eq. (8.1) represents the local loss of the orifice. Only in rare cases the head loss is directly calculated because it also depends on the discharge Q. In general, one of the following parameters is used: (a) ζ [−] and (b) χ (s^2/m^5). The connection between these two coefficients is given by:

$$h_v = \zeta \cdot \frac{v_2^2}{2 \cdot g} = \chi \cdot Q^2 \tag{8.3}$$

The value χ is often used for 1D-simulations of surge tanks, but in literature normally ζ is used, for which the velocity v_2 at the downstream section is further needed for the calculation of h_v. In case of higher Reynolds-Numbers, both values are constants and only depend on the geometry.

In the case of an asymmetric orifice the local head loss coefficient is depending on the flow direction. Hence, the parameter λ_{DownUp} [−] is defined by the ratio of the χ-values and quantifies the asymmetric behaviour. Using the ζ-coefficients, an additional factor depending on the two cross sections in the lower chamber and the shaft has to be introduced as follows:

$$\frac{\chi_{\text{down}}}{\chi_{\text{up}}} = \lambda_{\text{DownUp}} = \frac{\zeta_{\text{down}}}{\zeta_{\text{up}}} \cdot \frac{r_{\text{Shaft}}^4}{r_{\text{Lower chamber}}^4} \tag{8.4}$$

The value λ_{DownUp} is the main optimisation parameter and should be greater than 2.5 [−] for the newly built asymmetric orifice. For the previous throttle λ_{DownUp} reached up to nearly 50.0 [−] [65, 66], but needed a start-up time which can cause additional transient effects in the hydraulic system.

The following investigation concentrates on the local head loss coefficient ζ in one direction. Combining the basic equations (8.1) and (8.3) with respect to the assumptions for the elevation and the kinetic energy flux coefficient, the local head loss coefficient ζ can be defined as:

$$\zeta \cdot \frac{v_2^2}{2 \cdot g} = h_V = \frac{p_1 - p_2}{\rho \cdot g} + \frac{v_1^2 - v_2^2}{2 \cdot g} \tag{8.5}$$

$$\Rightarrow \zeta = \left(\frac{p_1 - p_2}{\rho} + \frac{v_1^2 - v_2^2}{2} \right) \cdot \frac{2}{v_2^2} \tag{8.6}$$

Due to the rotation into the datum plane, the value of ζ can be calculated independently of gravity with Eq. (8.6). The velocity v and the pressure p at both sections of the pipes are therefore needed. The differential pressure Δp is defined

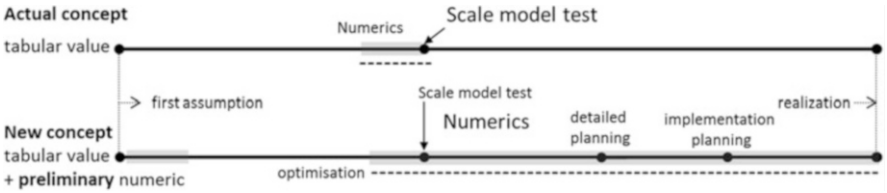

Fig. 8.2 Scale model test including the additional orifices

as $(p_1 - p_2)$ and can be measured with a higher accuracy than each pressure separately [16].

Using the principle of mass conservation (continuity equation):

$$Q = v_1 \cdot A_1 = v_2 \cdot A_2 \tag{8.7}$$

the difference $\left(v_1^2 - v_2^2\right)$ can be reduced to a geometrical relation between the cross sections A_1 and A_2. Therewith, Eq. (8.6) can be converted and the loss coefficient is calculated as follows:

$$\zeta = \left(\frac{\Delta p}{\rho \cdot Q^2} + \frac{\left(A_1^{-2} - A_2^{-2}\right)}{2} \right) \cdot 2 \cdot A_2^2 \tag{8.8}$$

All in all, Eq. (8.6) is easy to use in the 3D-numerical simulation but for the scale model test equation (8.8) is more convenient since the measured values are: (a) differential pressure Δp and (b) discharge Q. It is essential for the calculation of the local head loss coefficient to know the observational error of each parameter. Further investigation on this topic can be found in [16].

8.2.3 Modelling Concept

At the beginning of the optimisation process the fixed boundary conditions have to be defined. In this particular case the diameter of the lower chamber and the shaft should not be modified. The 90°-elbow is optimised separately at the end. Thus, the only variable is the geometry of the last two segments of the orifice. For this, a first assumption is to be made. Three different tools can be used:

- Tabular values based on the literature or standards
- Scale model test in the laboratory
- Numerical simulations

As part of the research project, a new concept of the complete process was formulated and tested (Fig. 8.2). For both, the new and the existing concepts, the starting point are tabular values. If a standardised throttle is used, the exact values

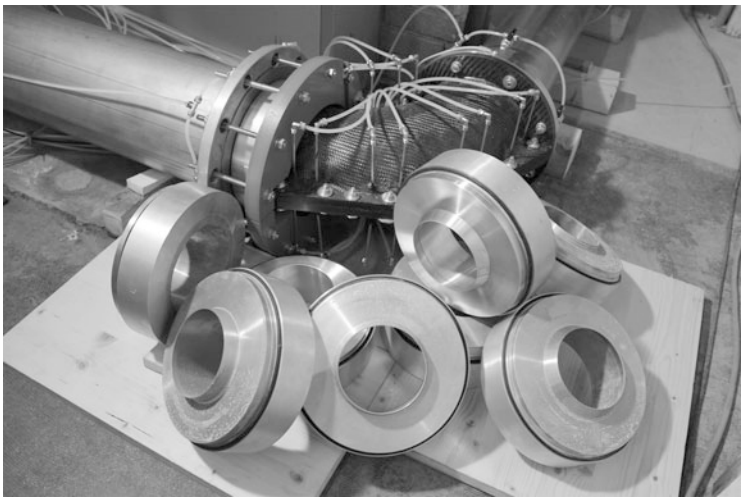

Fig. 8.3 Scale model test including the additional orifices

for the local head loss coefficient can be found in literature. Especially asymmetric orifices are non-standard parts and so only a rough estimate can be made based on these tabular values.

As a next step of the actual concept, these first assumptions are tested in physical laboratory tests [34, 41]. More and more, numerical simulations are used as a prearrangement and validation of the scale model test [37]. The new concept changes the weighting of these two tools. The main model concept should now be the 3D-numerical simulations, which are validated with the scale model test. The big advantage of this approach is that the numerics can easily accompany the complete planning and constructing process.

In addition to this, at an early stage of the investigation, the preliminary numerics can offer a wide range of geometry variations and help finding a suitable orifice. Thus, a simplified model can be used. In this particular case, the 90°-elbow was reduced to a change in the diameter of the pipe. Only a segment of this axially symmetric geometry was simulated with ANSYS-CFX [2]. Based on the fully parameterised geometry, the influence of each length and angle on the local head loss coefficient could be analysed.

Based on these variations, hypotheses for the future design of asymmetric orifices could be found [16]. Nine different orifices (Fig. 8.3) are defined and tested in parallel in the scale model test and with ANSYS-CFX. For this validation experiment, the numerical simulation of the model including the elbow was done on the scale of the laboratory test (scale 1:25). Due to the symmetry the geometry could be reduced to a half model. Each part was verified separately and the comparison led to a very good agreement of the two model approaches. Hence, the use of the 3D-numerics as well as the hypotheses for the design could be proved and the optimisation target $\lambda_{\mathrm{DownUp}}$ (8.4) of 2.5 [−] could be reached.

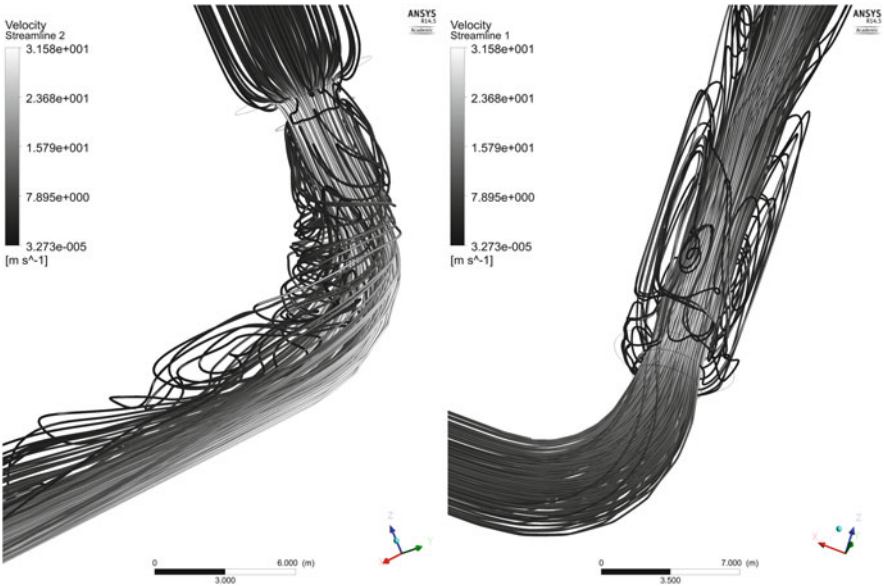

Fig. 8.4 Exemplary numerical results of the optimised asymmetric orifice in downward (*left*) and upward (*right*) flow direction—nature scale, ANSYS-CFX

To check the laboratory results in nature scale, additional numerical simulations were conducted, which allow to asses scale effects in the laboratory tests. Furthermore, the influences of modifications in the construction of the elbow and the outer part of the orifice were checked. So the complete detail design process could be supported by the 3D-numerics. Exemplary results (streamlines) of the last simulations are shown in Fig. 8.4. The now optimised asymmetric orifice is currently under construction and in 2015 the complete project should be finished.

8.2.4 Conclusion and Further Research

3D-numerical simulations of such complex and unique hydraulic parts offer an effective way to quantify the local head loss and check the flow conditions. The model can be modified and adapted to the need of each planning phase. The investigation can be done as part of a new design or of a revitalisation project. Compared to only using assumptions based on tabular values, better results can be reached.

The simple way to vary the geometry in the numerical model can lead to a cost-effective optimisation process. Each simulation, independent of the discharge or geometry changes, needs nearly the same time to calculate. In comparison to this, the scale model test offers a wide range of discharge variations. If the model

is built up, only a few seconds are needed to measure a different flow situation. Based on this fact and the long lasting experience with scale model tests, it is highly recommended to regularly validate the numerical results based on laboratory tests.

The results of each simulation, which is conducted as part of the presented research project, are based on the RANS equations. It could be proved that the mean value of pressure fluctuation could be simulated very accurately [16], but if the minima and maxima at the time dependent values are needed, the results based on the measurements of the scale model tests have to be used. Other numerical models, for example LES, could simulate these fluctuations, but would require a lot more calculation power and are part of future research.

8.3 Spillway

To prevent an overtopping of the dam caused by an incoming flood in the reservoir, spillways are used as a safe passage into the downstream river section. These hydraulic structures can be categorised according to the function (emergency or service use), to the mode of control (uncontrolled or gated) as well as to typical hydraulic criteria. Most spillways are built as side channel, overfall, chute, tunnel or a combination of the mentioned structures. In rare cases, a siphon or a shaft can be part of a spillway [50, 68].

The quantification of the design flood is an essential part for the dam safety. All of the different methods are based either upon historical records of maximum observed floods or on rainfall analysis, which is converted to runoff. Details can vary depending on the country. In Germany, for example, two discharges have to be proved based on the DIN 19700-11 (Sect. 4.3). For large dams, the smaller BHQ_1 is calculated with a return period of 1,000 years and the BHQ_2 with 10,000 years. The BHQ_1 should pass the reservoir without damage even if the outlet with the highest capacity has to stay closed ($n - 1$ rule) [68].

For Austrian dams the BHQ (*Bemessungshochwasser*) is regulated similarly to Germany, but with a return period of 5,000 years. The second bigger flood, also called SHQ (*Sicherheitshochwasser*), is identical to the probable maximum flood (PMF) [6]. This is depending on the flood producing catchment areas and is not a fixed value. With regard to design purpose, it is considered to have the same 5,000 year rainfall return period, but acting on somewhat more unfavourable pre-conducting as the BHQ. Therefore, it has to be reviewed and if necessary corrected [50].

8.3.1 Investigation Area

The presented case study is an uncontrolled side-channel spillway of the balancing reservoir Enzingerboden located in Austrian Alps owned by the ÖBB-Infrastruktur

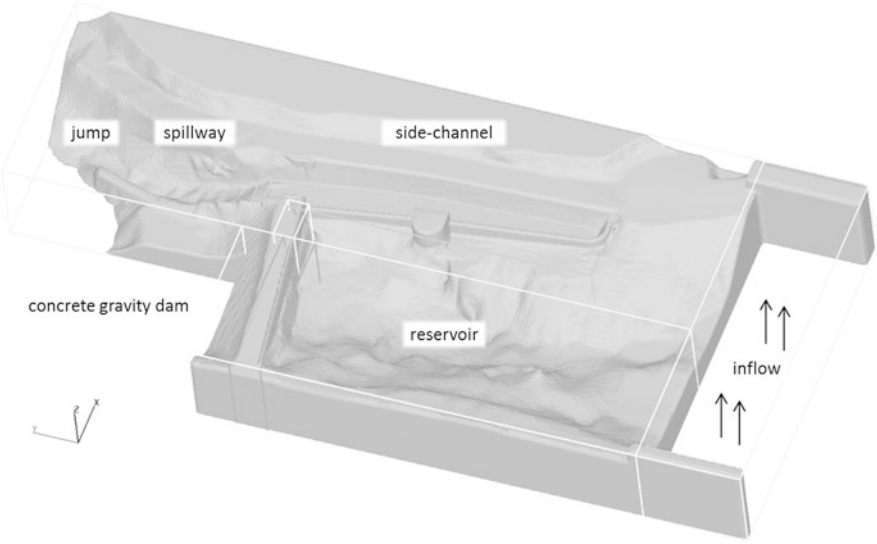

Fig. 8.5 Discretised geometry of the numerical model—FLOW-3D

AG. The capacity of the spillway has to be increased in about additional 100 % of the actual maximum flow rate to reach the new PMF.

The main part of the spillway is a side-channel, which is placed next to the dam (Fig. 8.5). The weir is nearly perpendicular to the dam axis and split into two sections by the entrance to the bottom outlet of the reservoir. Under current conditions, an unsubmerged overflow could not be maintained for the given new design flow. Hence, it is necessary to widen the limiting channel and the opening in the direction of the dam axis. Afterwards, the steep declivity leads to a fast increase in velocity of the water and at the end of this channel a jump is located.

8.3.2 Modelling Concept

As a first step, the given geometry was tested with 3D-numerics. The new design flood would cause an overtopping of the dam, thus a redesign of the spillway is an inevitable consequence. Therefore, a hybrid concept based on two different model assumptions was used [9, 21]:

1. To find a new geometry of the spillway with respect to maximum allowable water levels in the reservoir, different options are first investigated with the help of **3D-numerical simulations** (Fig. 8.5).
2. For detailed optimisations of the structure and validation of the numerics a **scale model test** was built up (Fig. 8.6).

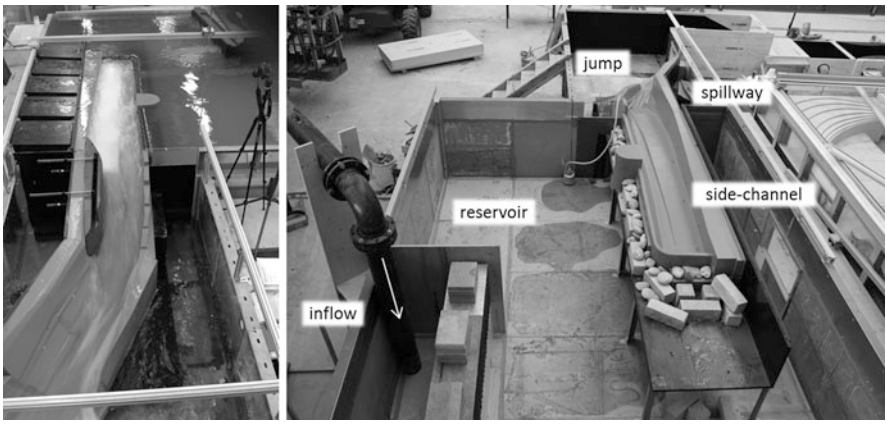

Fig. 8.6 Scale model test—operating state (*left*) and without water (*right*)

For the 3D-numerics the software FLOW-3D [12] was used. This software is limited to structured rectangular mesh blocks. To simulate a complex geometry, different blocks can be strung together and nested. The latter is used to refine the grid locally. In contrast to other software solutions such as ANSYS-CFX [2], which was used in Sect. 8.2, the discretisation of the geometry in FLOW-3D (also known as fractional area/volume method or short FAVOR [12]) has to be checked carefully. The restricted meshing is a disadvantage of FLOW-3D but it is compensated by a very good simulation of free surface flow. The software does not need to simulate the air-flow above the free surface and only calculates the velocity field of the water. To track the surface as a sharp interface moving through the used computational grid, the Volume of Fluid (VOF) method is used. Furthermore, in case of moving objects (Sect. 8.4) or small changes in the geometry the grid can stay the same and so a comparison can easily be made between two variations of the geometry [16].

In Fig. 8.5 the discretised geometry of the 3D-numerical investigation is shown, which considers a rectangular section of 150 to 70 m of the reservoir. Six mesh blocks with nearly 4 million cells (50 % of them are active in the simulation) are defined. The upstream part of the reservoir and the inflow are simulated with an inflow block, where the discharge is provided from the bottom of the block (Fig. 8.5). Hence, the water level in the reservoir is depending on the capacity of the investigated spillway geometry. Different local adaptations such as higher inclines of different parts in combination with an expansion of the channel were investigated. At the end of this optimisation process, a geometry could be defined for which the SHQ with 211 m^3/s could safely pass the dam [21]. No further optimisations were investigated. In addition to a preliminary geometry further findings could be made:

- The velocity in the reservoir before the weir section is very small. Hence, the geometry of the reservoir has nearly no impact on the capacity of the side channel.

- Based on the same reason, the dam geometry can be simplified to a vertical wall.
- The area behind the dam and especially the foundation stay dry in all investigated cases.
- All optimisations are based on the adding of material.

The first three points allowed a simplification of the model boundaries of the scale model test. As a result, the lab scale could be improved from 1:25 to the used scale of 1:20 (Froude-model [34]) without additional costs. The complete geometry, which has to be manufactured, could be reduced to essential and mostly pre-fabricated parts. Hence, a fast and cost-effective build-up was possible. The last point of the itemisation is essential for the operation of the model. The existing sealing level and the ground structure can stay untouched during the complete optimisation process. Each built-in component and the combinations could be tested under various flow conditions.

In Fig. 8.6 the picture at the left shows one of the investigated so called noses. In this area of the slope, existing anchors are situated, which possibly should not be removed. The scale model test could demonstrate that these built-in components even improve the behaviour of the spillway and direct the water so that a smaller heightening of the side walls in the spillway is needed. Beside these optimisations, the comparison of both model assumptions was conducted. To use the water level in the reservoir, the maximum difference between both concepts is smaller than 5 cm in nature scale. In comparison to the up-scaled measurement inaccuracy and especially to possible wave heights in the reservoir under such extreme circumstances, this difference is insignificant.

8.3.3 Conclusion and Further Research

The capacity of the spillway could be increased by more than 100 % of the previous discharge. Therefore, the advantages of both models could be of use. The 3D-numerical model offered a reliable predesign and helped to minimise the start up effort for the geometric optimisations in physical modelling. Based on the hands-on optimisation in the scale model test a good and cost-effective way to enlarge the spillway could be found. The validation based on the comparison of numerics and scale model test led to a high safety in the chosen design. Especially for such safety-relevant structures such as a spillway, this combined use can lead to the requested very high level of safety.

8.4 Avalanche into a Reservoir

Based on the hydro power production, dams play an extremely important role in alpine catchments. Reservoirs reach the maximum filling very often after the snow melt in summer. Other smaller dams are used for the production of artificial snow

in the winter. These reservoirs are filled during the summer time and reach the peak water level in early winter. To guarantee the full energy storage and water amount, all reservoirs are filled up to a maximum water level considering the necessary and defined freeboard. To prevent an overtopping by an incoming flood, spillways are used (Sect. 8.3). Especially in mountain areas, reservoirs are endangered by potential landslides and surrounding avalanche tracks. Such impacts may generate impulse waves, which can flood the dam in the worst case [49]. An efficient way to prevent the effect of impulse waves is to drop the water level in the reservoir. For hydro power plants, the hydrology in the winter months and the energy production results in a lower water level. Keeping the water level low results in a decreased energy production, which affects equally pump storage stations and run-of-river plants. In addition, also the costs for artificial snow production can increase as a consequence because a bigger reservoir is needed for the same amount of snow. To minimise the flooding risks and to optimise the permissible water level, the impact and the impulse wave in the reservoir are to be predicted as accurately as possible.

Fundamental research on this topic was conducted at the ETH Zürich in the last years [13, 14, 33, 49, 75] and a manual to calculate such impulse waves was released [35]. The effect of seven different governing parameters for impulse waves generation was tested in a physical model [33]:

- water depth of the water body
- slide thickness
- slide impact velocity
- bulk slide volume
- bulk slide density
- slide impact angle
- grain size diameter

The particle size as well as the length of the sliding mass is negligible [13,33,75]. Apart from these main trigger parameters of impulse waves, the wave generation is also depending on the avalanche track in correlation to the dam and on the reservoirs topography. Furthermore, the superposition of reflexions could lead to a bigger wave height than the initial one. In the majority of cases, a single empirical calculation is therefore not sufficient and further investigations based on scale model tests or 3D-numerics are needed [15, 18].

8.4.1 Investigation Area

The presented project area is the reservoir of a run-of-river plant, which was tested at a scale of 1:35 (Froude-model [34]) in the laboratory. The investigated structures were two weir fields with an intake and a fish pass at the orographically right side of the reservoir. In addition to the hydraulic and sedimentation tests, impulse waves due to an avalanche impact had to be evaluated. In this particular case, one main direction of the impact could be defined as well as the slide angle with 30°.

Figure 8.7 shows the scale model test including the channel to simulate the avalanche path. The presented investigation was also used as a calibration

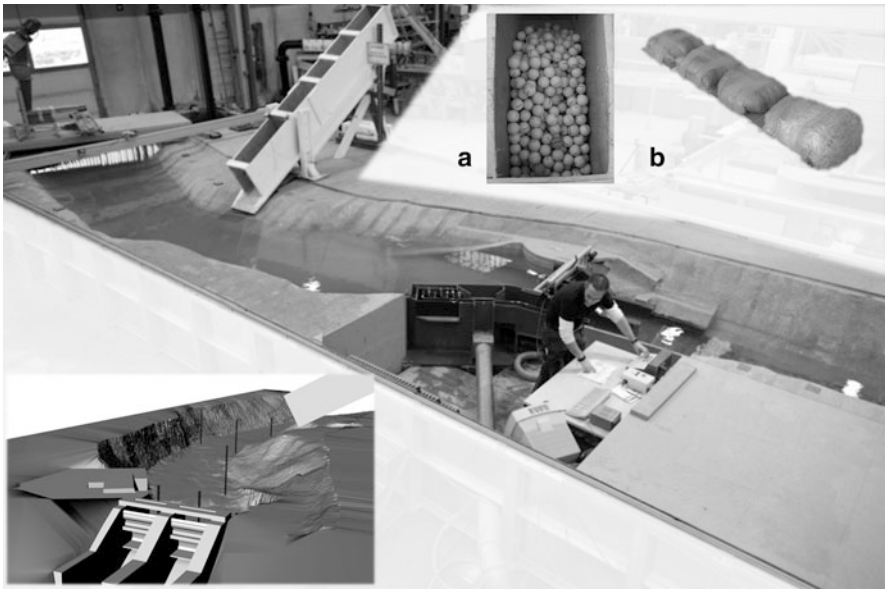

Fig. 8.7 Physical scale model (*big picture*) including the two set-ups (**a**) golf balls and (**b**) skateboard; model for the numerical investigation (*small picture*)

experiment for the 3D-numerical investigation of such impacts. The simulations with FLOW-3D are based on laser scan data of the built-up laboratory model. For comparison of the two model assumptions, water levels at the opposite side of the impact and in front of the weir were chosen. The points are marked in the small picture in Fig. 8.7 with the help of staves [18].

8.4.2 Modelling Concept

In the scale model test and the 3D-numerics, two different test set-ups for the avalanche were defined [17]:

Scale model test:
- P1 golf balls
- P2 skateboard

Numerics:
- N1 particles with initial water
- N2 general moving object

The physical set-up P1 uses 900 single golf balls with a density of about 1,140 kg/m^3 and can simulate with its porosity of about additional 25 % an avalanche density of 600 kg/m^3 (Fig. 8.7a). This represents a porous slide mass, which could be varied by the volume of the golf balls in the start reservoir. A density of 700 kg/m^3 could be reached with the second one, for which a compact moving

block was mounted on a skateboard (set-up P2). For this the modelled avalanche height is fixed. The upscaled length of the simulated avalanche head is 170 m for P1 and 100 m for P2. For both hypotheses, the slide impact velocity could be determined by varying the distance between the gate in the speed-up channel and the water surface. All geometric values were used as fixed parameters.

In comparison to the golf balls, three different fluids (water in the reservoir, air and moving snow on the hillside, which was modelled with the golf balls) would have to be simulated in the numerical model. FLOW-3D is limited to two different fluids. Thus, up to 4,800 particles are used as a numerical approach N1, which have identical density but only one-third of the diameter. First tests showed that the interaction between particles is not simulated correctly and so the main challenge was to model the correct movement of particles. Therefore, additional initial water has to be provided in the channel, which was varied between 0.8 l up to the best results with about 38 l in model scale.

The skateboard in the physical scale model test is represented as general moving object (GMO) [7, 12] within the numerical set-up N2. Flow-3D offers two different models for the calculation of such a GMO. In the prescribed mode, the complete curve of motion is predefined and only the interaction with the water is calculated. A two-way interaction is incorporated in the coupled mode. With the latter, in theory the real behaviour of the floatable skateboard would be simulated, but tests showed an abnormal bounce after the first contact with the water. Hence, the prescribed movement was used with a velocity of 3.4 m/s (nearly 20 m/s in nature).

8.4.3 Conclusion and Further Research

Exemplary results for the physical model are shown in Fig. 8.8 and for numerical simulation in Fig. 8.9. Based on the high impact Froude-numbers, in all cases the impacts result in a forward collapsed crater [13]. A lower crater height was generated in case P1 and N1. It is assumed that this is a result of the single impulse impact of each golf ball (P1) and the mixture of water and particles (N1). In comparison to this, P2 and N2 have only one big impact, which generates a nearly doubled wave height at the opposite bank of the reservoir. The interferences of reflected waves reduce the difference between those two model pairs until the wave reaches the weir.

The main conclusion of the investigation was that both physical avalanche models P1 and P2 could be simulated with a very good agreement using numerical assumptions. The difference between P1 and N1 as well as P2 and N2 was in the range of few centimetres after the upscale into nature scale. In contrast to this, the maximum elevation of the free surface near the weir ranges from nearly 1 (P1 and N1) to 0.6 m in nature scale. Hence, the uncertainty associated to the simulation of such an avalanche impact can be quantified in the range of decimetres. Consequently further calibration data is needed to simulate such an avalanche impact into a reservoir more reliably.

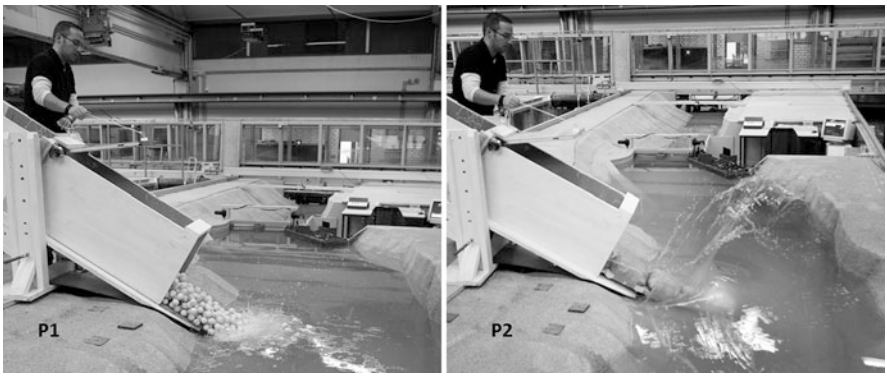

Fig. 8.8 Physical set-up golf balls (*left*) and skateboard (*right*)

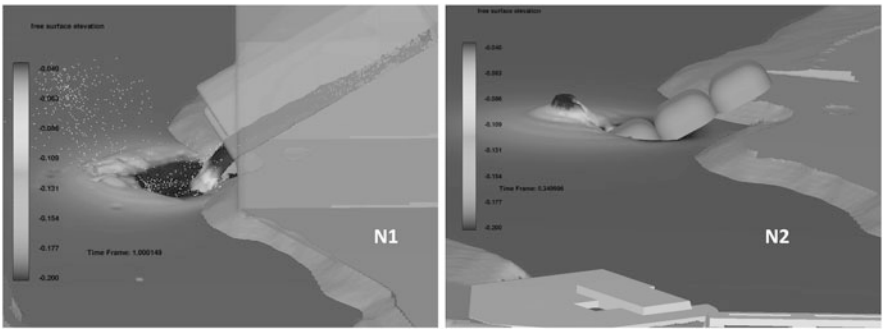

Fig. 8.9 Numerical set-up particles (*left*) and general moving object (*right*)

8.5 Log Jam Processes

In catchments with considerable forest cover, the entry of large woody debris into torrents and rivers is a frequently appearing process under torrential hazards conditions. Erosion on the embankments, landslides, storms and avalanches represent essential mechanisms causing the input of fresh wood into channels [44]. Both, fresh and dead wood [62]—the latter is already located in the channel or in the near surroundings (e.g. died off wood, industrial wood)—can be moved in the channel if the flow forces and water depths enable mobilisation. For specific hydraulic conditions and wood characteristics, driftwood is threatened to be blocked at narrow channel sections such as gorges and bridges. The arised log jam consequently causes a reduction in the cross-sectional area in the channel and, thus, damming and backflow affects the upstream region of the clogged cross section. Accordingly, adjoining settlement area and infrastructure are exposed to a considerable higher flood risk compared to clear-water conditions without any relevance of driftwood [23, 26, 62].

A comprehensive understanding of all driftwood-related processes within the context of protective hydraulic engineering initially requires research in the mechanisms of wood entry into the channel, of its mobilisation and of driftwood transport [47]. Further focusing on the impacts at bridges and narrow spots in the channel, the knowledge of substantial process parameters regarding topography, hydraulics and driftwood characteristics is required for the quantification of flood risk and the analysis of protective measures [5, 64]. Damming and backflow effects, but also clogging probabilities for a variety of specific bridge structures and topographic conditions need to be determined, each for critical hydraulic conditions and specific driftwood characteristics [22, 29, 30, 63, 67]. Finally determining the log-jam-induced flood risk of buildings and infrastructure, flood plain modelling and the estimation of potential damages are required [23, 26].

Essential requirements for the analysis of bridge clogging processes are a physically correct modelling of the logs moving within flow and an adequate simulation of the three-dimensional flow characteristics conditions near by the bridge structure. On the contrary, flood plain simulations for large settlement areas require a modelling approach on a larger spatial scale. Three-dimensional flow effects in the channel mostly have no significant influence there. Considering the capabilities and the constraints of possible modelling approaches, the simulation of bridge clogging requires a physical scale model, locally focusing on the processes at the bridge and in the channel upstream. However, flood plain modelling is generally accomplished with 2D-hydro- or morphodynamic models.

8.5.1 Modelling Concept

Aiming at a holistic and precise simulation of log jam processes and its effects on flood risk, a hybrid modelling concept is required, comprising and linking an experimental model and a 2D-numerical model according to the scheme presented in Fig. 8.10 [23, 26].

The physical scale model is applied according to Froude's similarity law. The log jam characteristics in terms of clogging probabilities and correlations of backflow and damming are initially determined for a specific set of experimental arrangements (topography, hydraulics, driftwood characteristics). According to the scheme illustrated in Fig. 8.10 and the model log characteristics in Fig. 8.11, these test series are accomplished with the use of artificial logs. They can be either configured uniformly, ensuring simplified and repeatable experimental conditions, or feature a more natural non-uniform structure. In order to deliver clogging probabilities, the tests have to be conducted under the same conditions several times. Since the transport of floating elements such as driftwood is not implemented within current 2D-numerical models, bridge structure configuration has to be altered in the experimental model, which best possibly imitates the delivered log jam characteristics under clear water conditions. This is done by lowering the lower bridge deck level or rather occluding a part of the cross-sectional area with baffle

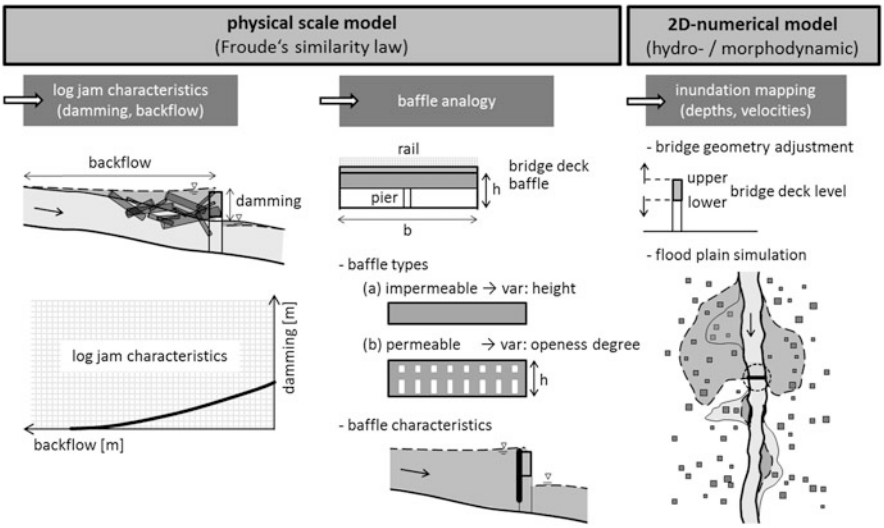

Fig. 8.10 Modelling concept comprising a physical scale model and a hydro-/morphodynamic numerical model (modified after [23, 26, 30])

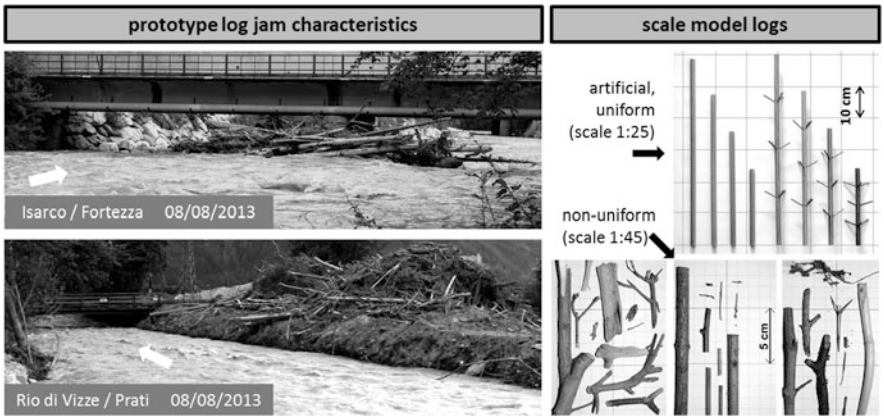

Fig. 8.11 Comparison of prototype and scaled log characteristics

elements. When using an impermeable baffle structure, its height has to be varied until a match with the log jam tests is gained. For the case of a permeable baffle, the openness factor has to be fitted (Fig. 8.10).

Within the numerical model HYDRO_AS-2D [53], a bridge deck is generally defined by determining the upper and lower bridge deck levels. Accordingly, pressure and weir flow conditions are considered when the water level reaches the respective deck level. For both flow conditions, the calculation is done empirically with the following equations [53]:

$$Q_{\text{pressure flow}} = c \cdot A \cdot b\sqrt{2 \cdot g \cdot \Delta h} \tag{8.9}$$

$$Q_{\text{weir flow}} = \frac{2}{3} \cdot \sqrt{1 - \left(\frac{h_{UW}}{h_{OW}}\right)^{16}} \cdot \mu \cdot b \cdot \sqrt{2 \cdot g \cdot h_W^3} \tag{8.10}$$

Therein, Q is the discharge (m^3/s), A is the wetted cross section area (m^2), b is the width of the bridge deck (m) and Δh denotes the difference in water levels upstream (h_{OW}) and downstream (h_{UW}) the bridge (m). Further, h_W is the upstream water level (m), related to the upper bridge deck level, or rather the corresponding kinetic head (m). The parameters c and μ are dimensionless run-off and weir flow coefficients.

With the results from experimental modelling with the artificial logs and the baffle analogy, flood plain modelling is accomplished with a 2D-numerical model, in which the bridge geometry is adjusted accordingly. The numerical model either exclusively simulates clear water conditions (e.g. HYDRO_AS-2D [53]) or additionally considers sediment transport processes (e.g. HYDRO_GS-2D [54]). Within experimental modelling, sediment transport processes are not yet considered.

8.5.2 Results

Concerning the experimental tests, accomplished in the hydraulic laboratory at the University of Innsbruck [29, 30], the following parameters/conditions lead to a general increase in clogging probabilities:

- increasing log lengths
- increasing number of branches
- increasing number of supplied logs, entrainment as a cluster
- increase in water level or rather decreasing freeboard at the bridge structure
- increasing channel gradient in case of pressure (and weir) flow conditions at the bridge
- decreasing channel gradient for the case that the bridge is not dammed (under clear water conditions)

The most influential conditions therein are the presence of branches and, particularly, the manner of wood input at the upstream model boundary. The more logs are supplied concurrently as a cluster the higher is the clogging probability of at least one single log. Within the mentioned scale model tests [29, 30], an inclinable flume with a rectangular shape and gradients between 0 and 3 % was set up. The model scale was 1:25. Different water levels and flow conditions (Froude numbers) were tested. The effect of piers was further analysed. Concerning the logs characteristics, both, uniform logs with and without branches were supplied, each as single logs and as a cluster (Fig. 8.11). Each test was accomplished several times.

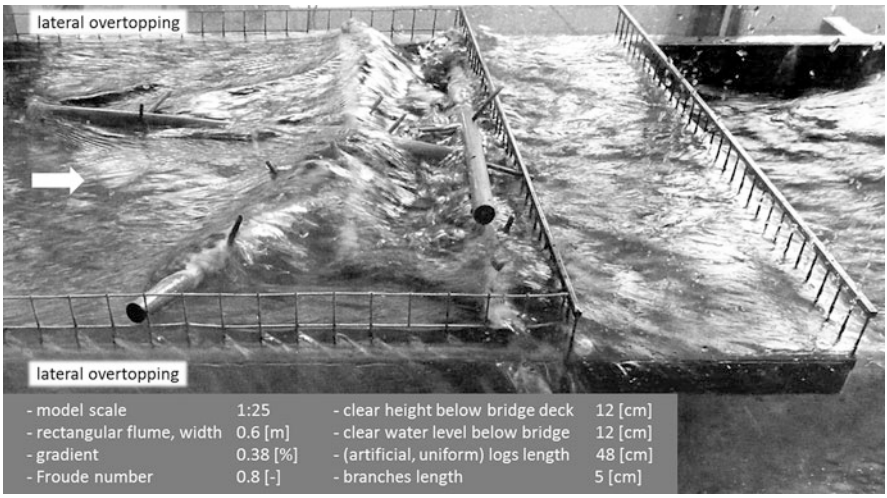

Fig. 8.12 Impression from experimental modelling

Focusing on the hydraulics, the damming is mainly influenced by the channel gradient and, by the number of logs clogged at the bridge. On the contrary, the logs dimensions are of minor significance. In case of supercritical initial flow conditions a flow transition appears once a log is clogged at the bridge. For subcritical initial conditions the extent of damming reaches further upstream.

Figure 8.12 exemplarily shows a clogging situation that appeared under subcritical flow conditions with a supplied cluster of logs with branches. There was no pier situated at the bridge for this case.

Clogging scenarios may cause a considerable increase in the extent of flooding and, thus, in the amount of flood related damages [23, 26]. Depending on the hydrological impact and on the channel capacity, log jams increase the extent of flooding upstream of the bridge section, whereas downstream, the flood plain rather decreases since the upstream flood plain has a retention effect. According to this, the impact of log jam processes on specific elements at risk within the settlement area is as well depending on the location of the elements compared to the location of the clogged bridge structure.

8.5.3 Conclusion and Further Research

The simulation of log jam processes within the hybrid modelling concept illustrated in Fig. 8.10 allows a well-founded determination of the increased flood risk due to the clogging of bridge structures or narrow spots in the river channels. Within the physical scale model a detailed insight in the clogging process is provided. However, the application of a 2D-numerical model enables an extensive impact assessment on

a large spatial scale. The conduction of the modelling concept not only allows the estimation of flood risk and potential damages due to clogging processes, it also is the basis for protective engineering measures, for instance structural measures at the bridge, practical evacuation plans etc.

Further research activities related to log jam processes deal with the following issues for instance:

- enlargement of the data base for bridge clogging resulting from the physical scale model (trapezoidal flume, intrusion of fine particles and bushes in the entrained cluster of logs, different bridge structures, influence of channel surface roughness, etc.)
- influence of sediment transport processes on clogging probabilities and structure of the clogged logs
- further development of numerical models (3D-numerical models, floating elements, etc.)
- field observation of log jam related processes—data base for model validation
- collecting research results of all relevant research institutions for providing a common and accessible data base
- analyses of specific protective engineering measures in the physical scale model and as well in the numerical model

8.6 Sedimentation and Flushing of Alpine Water Intake Reservoirs

Reservoirs and retention measures in hydro power projects affect the sediment balance in a river. Interrupting the natural flow conditions, sediments are held back and sedimentation arises in the reservoir storage. In the Alps mean annual storage losses are between 0 and 5 % of the stored water volume [72].

Several measures can help to avoid or to reduce storage loss. Most common methods are sediment bypass, off stream reservoirs or a simple periodic excavation of the stored material. In narrow V-shaped valleys the pass through method (sediment flushing) is an appropriate technique for sediment management [56].

8.6.1 Investigation Area and Data Base

The presented research has been realised with the support of the TIWAG-Tiroler Wasserkraft AG who has projected two water intakes with arch dams in the alpine valley Ötztal. One of the intakes is projected at the river Gurgler Ache and the second one is planned at the Venter Ache. Due to the topography (narrow, V-shaped valleys) the structures are planned as arch dams with overfall spillways and bottom outlets. The two reservoirs are fed by glaciated headwaters and the mean annual

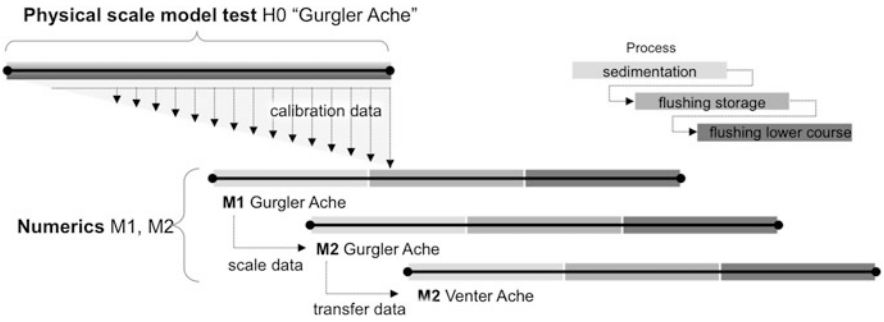

Fig. 8.13 Model concept

sedimentation process is an important issue. In addition, sediment supply during extreme events, which is accompanied by large bed load amounts, has to be controlled by an effective sediment management. Owing to the topography and the dam construction with bottom outlet, sediment management via flushing is enabled.

Sedimentation and flushing processes of the water intake Gurgler Ache were investigated with a 1:30 scaled hydraulic model (H0). The results of the model test are used to calibrate a 2D-numerical simulation of the water intake in model scale (M1). The parameters of the calibrated simulation are transferred to a nature scaled simulation (M2). After validation of the model Gurgler Ache, all parameters are used for a numerical simulation of the second water intake Venter Ache in nature scale (Fig. 8.13). Due to the similar dam design, topography and bed load conditions at the Gurgler Ache and Venter Ache, this method is a permissible approach. The aim within this method is to avoid a second, cost intensive lab model.

In the experiments and simulations two types of scenarios are modelled. For the investigation of mean yearly reservoir sedimentation the half annual bed load and the annual bed load were flushed in the reservoir. Furthermore, sedimentation processes were modelled which are associated with large flood flow events occurring within short times. Therefore discharge hydrographs of two historic extreme events (August 1987 and September 1999) were used. From available discharge series the return periods of the event 1999 and 1987 were estimated as 100 years (HQ100) and 50 years (HQ50), respectively [24]. For all given sub-catchments the bed load transport was calculated, using an approach for transport in steep torrents according to Rickenmann [59–61] and Bathurst [4]. Bed loads were routed downstream with a balance routing scheme according to Gems [23, 25]. The results show that the estimated annual bed load at the location of the projected reservoir Gurgler Ache is 8,300 m^3 and the half annual bed load is 4,150 m^3. The extreme event 1999 results in a load of 16,326 m^3; at the event 1987 calculations show a bed load of 22,647 m^3, which is almost three times the annual bed load.

The grain size distribution of the bed load is based on a "line by number analysis" and excavation analysis sampling sediments in the river bed in the study area. In hydraulic model tests a minimum grain-size of 0.5 mm has to be maintained to

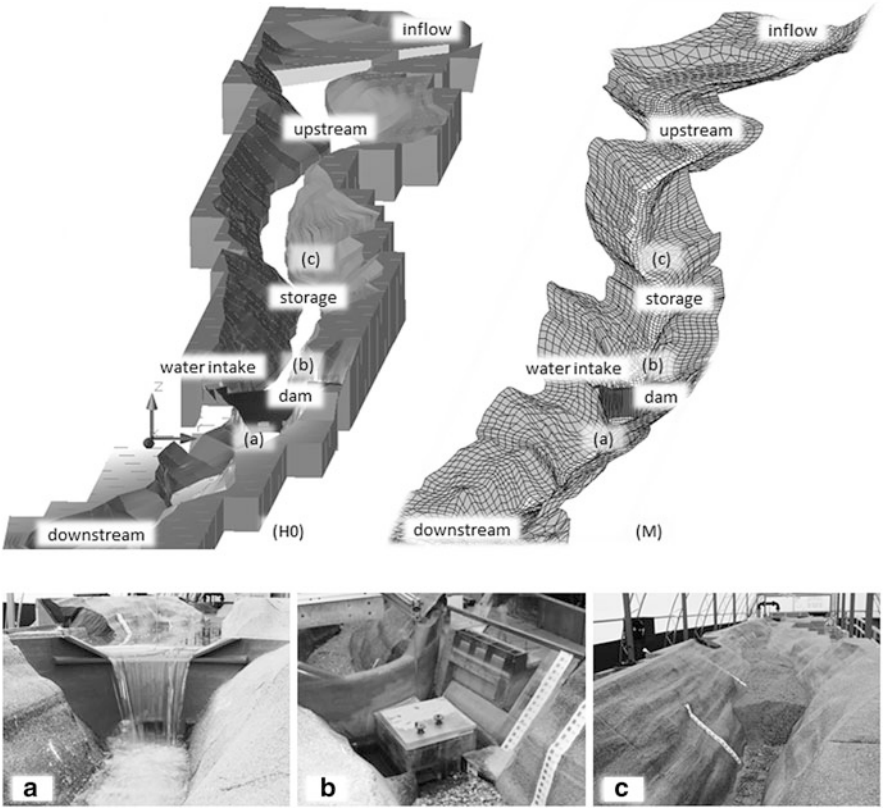

Fig. 8.14 Scheme hydraulic model (H0)(*left*); numerical simulation (M)(*right*); (**a**) overfall spillway, (**b**) bottom outlet and water intake, (**c**) storage facing headwater

avoid cohesive and other disturbing effects, which would influence the test results. With the scope of numerical simulation, the influence of fine sediments on the sedimentation and flushing process can be evaluated [55].

8.6.2 Modelling Concept

The lab model is based on the Froude-model approach which is usually used in modelling hydraulic structures like weirs, open channels or at free surface flow where gravity forces dominate. The Froude-model implements that the Froude number in the model is the same as in nature scale [34, 57]. The arch dam structure includes the operating structures of overfall spillway, bottom outlet and water intake. Besides the river stretch about 760 m upstream (headwater) and roughly 600 m downstream of the dam is modelled (Fig. 8.14) [56].

Table 8.1 Parameter settings for the numerical simulation

Abbrev.	Dimension	Definition	M1	M2
M_l		Model scale	30	1
τ_{crit}	[−]	Relative critical bed shear stress	0.03	0.047
ST	[−]	Sediment transport	MPM[a]	MPM[a]
p_r	[−]	Porosity	0.37	0.37
k_{1b}	$(m^{1/3}/s)$	Roughness river bed (bed value)	45	28
k_{1t}	$(m^{1/3}/s)$	Roughness river bed (total value)	60	28
k_{2b}	$(m^{1/3}/s)$	Roughness terrain (bed value)	70	45
k_{2t}	$(m^{1/3}/s)$	Roughness terrain (total value)	70	45
τ_{max}	(N/m^2)	Maximum bed shear stress	500	500

[a] MPM; sediment transport according to Meyer-Peter and Müller

The 2D simulations were realised using the numerical software Hydro_ AS-2D/ Hydro_ GS-2D. For discretisation in time, the second order Runge–Kutta scheme is used, which is a Predictor–Corrector scheme. The implemented algebraic turbulence model calculates the eddy viscosity depending on shear velocity and water depth. The friction slope is determined by the Darcy–Weisbach equation [53,54]. The software HYDRO_GS-2D uses a fractionated multi grain approach of the transported bed load. For sediment transport the software uses the bed load transport approach according to Meyer-Peter and Müller [45]:

$$m_g = 5 \cdot \rho_f \cdot \sqrt{\rho' \cdot g \cdot d_m^3} \cdot \left[\tau^* \cdot \frac{Q_s}{Q} \cdot \left(\frac{k_{\mathrm{st}}}{k_r} \right)^{3/2} - \tau_c^* \right]^{3/2} \tag{8.11}$$

Thereby, the amount of transported sediment m_g is a function of the density of the sediments ρ_f, the gravitational acceleration g and the mean grain size of the transported sediments d_m. Q defines the discharge and Q_s is the effective transport discharge. k_{st} is the Strickler roughness of the bed, while k_r defines the Strickler roughness of the transported grain. The bed shear stress is given by τ^* and τ_c^* is the relative critical bed shear stress according to Shields.

As mentioned above, the aim of the presented modelling concept is to use calibrated parameters from the numerical simulations of the intake Gurgler Ache for the second water intake Venter Ache. In a first step, the calibration was done by modelling the water intake in model scale. The grain size distributions of the added sediments are equal to those in the hydraulic model test, which do not contain most finest sediment classes. The roughness parameters were taken according to the terrain material and the river bed in the laboratory model. Roughness parameters of the calibrated model type M1 were transferred into nature scale to model M2 (Table 8.1). In addition, in the simulations with model type M2 the full grading curves, including the fine sediments, were set for the bed load.

The calibration of the sedimentation process in the numerics was done using results from the hydraulic model tests with half annual bed load and the extreme

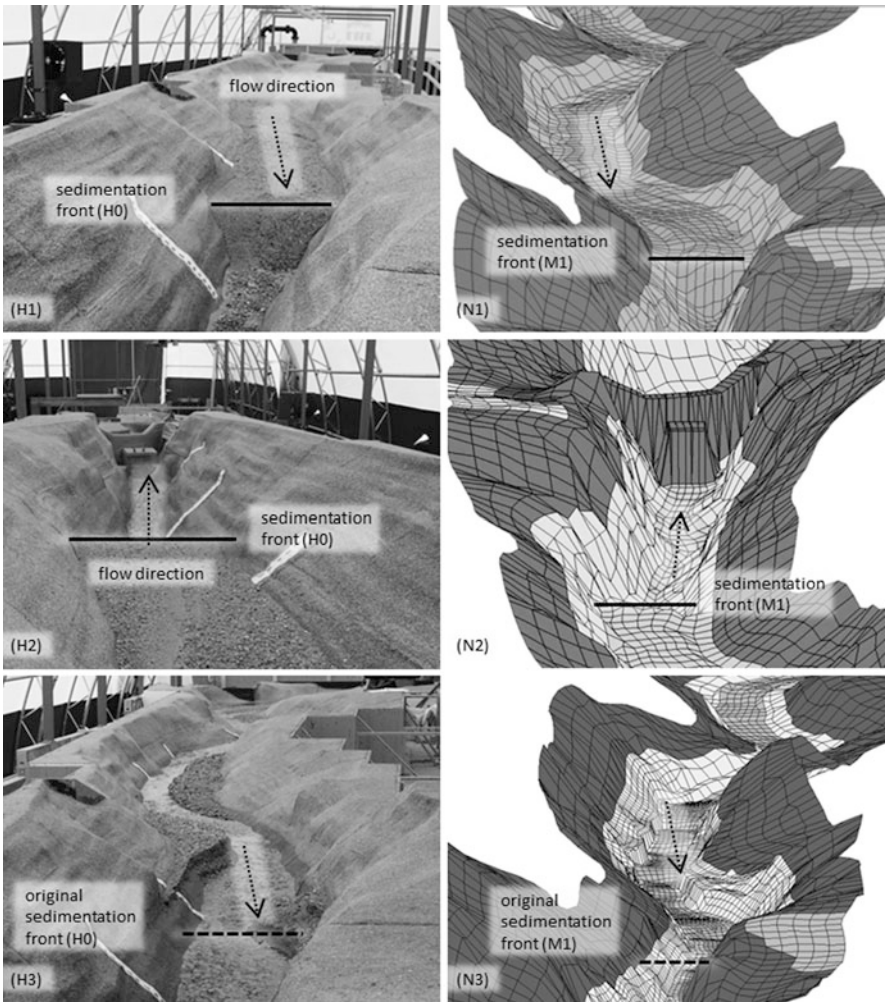

Fig. 8.15 Comparison between the physical scale model test H0 (*left column*) and the numeric simulation M1 (*right column*)

event in the year 1987. Within the calibration steps, following criteria were evaluated (Fig. 8.15) [55]:

1. Relative aggradation heights
2. Distance between the dam and sedimentation front
3. Absolute altitude of the sedimentation front
4. Longitudinal inclination of the sediment body

For calibrating the flushing process in the simulations observed flushing rates of the model test were used. The validation was done using a scenario with annual bed load volume and the extreme event in the year 1999.

8.6.3 Results and Further Research

Numeric simulation became an important tool in hydraulic engineering and is a vital extension to a hydraulic scale model test. The present study shows that the sedimentation process at reservoirs can be simulated using 2D-numerics. The results of the numeric simulations match very well with the results of the physical scale model test. The first obtained results in modelling the flushing process are promising, although the reservoir flushing is a highly unsteady process and the influence of three dimensional effects cannot be modelled with a 2D-simulation. Nevertheless, there are still differences between the observed flushing rates in the model test and the numerics. The next step will be a further investigation by analysing the sediment transport in the 2D simulations to achieve required improvements in the flushing process [55].

References

1. American Society of Mechanical Engineers: ASME-MFC-3M-2004: Measurement of Fluid Flow in Pipes Using Orifice, Nozzle, and Venturi. American National Standard, New York (2004)
2. ANSYS: ANSYS-CFX: User's Manual and Solver Modeling Guide (2010)
3. Aufleger, M., Joos, F., Jorde, K., Kaltschmitt, M., Lippitsch, K.: Stromerzeugung aus Wasserkraft. In: Kaltschmitt, M., Streicher, W., Wiese, A. (eds.) Erneuerbare Energien. Springer, Berlin (2013). doi:10.1007/978-3-642-03249-3_8
4. Bathurst, J.C.: Measuring and modelling bed load transport in channels with coarse bed materials. In: Richards, K. (ed.) River Channels: Environment and Process, pp. 272–294. Blackwell, Oxford (1987)
5. Bouska, P., Gabriel, P.: Results of a research project on flood protection of bridges. In: 33rd IAHR Congress, Vancouver (2009)
6. Leitfaden zum Nachweis der Hochwassersicherheit von Talsperren, Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft und TU Wien, Institut für Wasserbau und Ingenieurhydrologie, Karlsplatz, Vienna (2009)
7. Dargahi, B.: Flow characteristics of bottom outlets with moving gates. J. Hydraul. Res. **48**(4), 476–482 (2010)
8. De Souza, P., Boillat, J.-L., Schleiss, A.: Hydraulic System Modélisation des systèmes hydrauliques à ècoulements transitoires en charge. Communication Laboratoire de constructions hydrauliques, Ecole polytechnique fédérale de Lausanne, Lausanne (2004). LCH N16
9. De Cesare, G., Pfister, M., Daneshvari, M., Bieri, M.: Herausforderungen des heutigen wasserbaulichen Versuchswesens mit drei Beispielen. WasserWirtschaft **102**(7–8), 71–75 (2012)
10. Deltares: WANDA 4.2: User's Manual (2013)

11. Dobler, W., Zenz, G.: Particle image velocimetry of a Y-bifurcator. In: 34th IAHR World Congress, Brisbane (2011)
12. Flow Science Inc.: FLOW-3D Version 9.4 User's Manual (2010)
13. Fritz, H.: Initial phase of landslide generated impulse waves. Doctoral thesis, ETH Zürich, Nr. 14871 (2002). doi: 10.3929/ethz-a-004443906
14. Fuchs, H.: Solitary impulse wave run-up and overland flow. Doctoral thesis, ETH Zürich, Nr. 21174 (2013). doi: 10.3929/ethz-a-009787661
15. Fuchs, H., Pfister, M., Boes, R., Perzlmaier, S., Reindl R.: Impulswellen infolge Lawineneinstoß in den Speicher Kühtai. WasserWirtschaft **101**(1–2), 54–60 (2011)
16. Gabl, R.: Numerische und physikalische Untersuchung des Verlustbeiwertes einer asymmetrischen Düse im Wasserschloss. Innsbruck university press (IUP), Innsbruck (2012)
17. Gabl, R., Kapeller, G., Aufleger, M.: The effect of avalanche impulse waves in reservoirs. In: 33rd IAHR World Congress, Vancouver (2009)
18. Gabl, R., Kapeller, G., Aufleger, M.: Lawineneinstoß in einen Speichersee: Vergleich numerisches und physikalisches Modell. WasserWirtschaft **5**, 26–29 (2010)
19. Gabl, R., Achleitner, S., Gems, B., Neuner, J., Aufleger, M.: Numerische Berechnung von Hochdruckanlagen: global betrachtet: lokal verbessert. Österreichische Wasser- und Abfallwirtschaft (2013). doi: 10.1007/s00506-013-0104-4
20. Gabl, R., Achleitner, S., Neuner, J., Aufleger, M.: 3D-numerical refinements to simulate high-head power plants. In: 35th IAHR World Congress, Chengdue (2013)
21. Gabl, R., Achleitner, S., Sendlhofer, A., Höckner, T., Schmitter, M., Aufleger, M.: Optimierter Einsatz und Kombination von 3-D-Numerik und physikalischer Modellierung. WasserWirtschaft **5**, 128–131 (2013)
22. Gantenbein, S.: Verklausungsprozesse: experimentelle untersuchungen. Diploma thesis, ETH Zürich (2001)
23. Gems, B.: Entwicklung eines integrativen Konzeptes zur Modellierung hochwasserrelevanter Prozesse und Bewertung der Wirkung von Hochwasserschutzmassnahmen in alpinen Talschaften: Modellanwendung auf Basis einer regionalen Betrachtungsebene am Beispiel des Oetztales in den Tiroler Alpen. innsbruck university press (IUP), Innsbruck (2012)
24. Gems, B., Achleitner, S., Huttenlau, M., Thieken, A., Aufleger, M.: Flood control management for an alpine valley in Tyrol: an integrated hydrological-hydraulic approach. In: 33rd IAHR World Congress, Vancouver (2009)
25. Gems, B., Achleitner, S., Plörer, M., Schöberl, F., Huttenlau, M., Aufleger, M.: Bed-load transport modelling by coupling an empirical routing scheme and a hydrological-1-D-hydrodynamic model: case study application for a large alpine valley. Adv. Geosci. (2012). doi:10.5194/adgeo-32-23-2012
26. Gems, B., Sendlhofer, A., Achleitner, S., Huttenlau, M., Aufleger, M.: Verklausung von Brücken: Evaluierung verklausungsinduzierter Überflutungsflächen durch Kopplung eines physikalischen und numerischen Modells. In: 12th Congress Interpraevent, Grenoble (2012)
27. Ghidaoui, M., Zhao, M., McInnis, D., Axworthy, D.: A review of water hammer theory and practice. Appl. Mech. Rev. (2005). doi: 10.1115/1.1828050
28. Giesecke, J., Mosonyi, E.: Wasserkraftanlagen: Planung, Bau und Betrieb. Springer, Berlin (2009)
29. Gschnitzer, T., Gems, B., Aufleger, M., Mazzorana, B., Comiti, F.: Verklausung von Brücken durch Schwemmholz: physikalischer Modellversuch. Wasserbausymposium TU Graz, Verlag der Technischen Universität Graz (2012)
30. Gschnitzer, T., Gems, B., Aufleger, M., Mazzorana, B., Comiti, F.: Physical scale model tests on bridge clogging. In: 35th IAHR World Congress, Chengdu (2013)
31. Haakh, F.: Vortex chamber diodes as throttle devices in pipe systems: computation of transient flow. J. Hydraul. Res. **41**(1), 53–59 (2003)
32. Hanjalić, K., Launder, B.: Modelling Turbulence in Engineering and the Environment: Second-Moment Routes to Closure. Cambridge University Press, Cambridge (2011)
33. Heller, V.: Landslide generated impuls waves: prediction of near field characteristics. Doctoral thesis, ETH Zürich, Nr. 17531 (2007). doi: 10.3929/ethz-a-005512384

34. Heller, V.: Scale effects in physical hydraulic engineering models. J. Hydraul. Res. **49**(3), 293–306 (2011)
35. Heller, V., Hager, W.H., Minor, H.-E.: Rutscherzeugte Impulswellen in Stauseen, Grundlagen und Berechnung. Mitteilungen der Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie (VAW), Nr. 206 (2008)
36. Hirano, M.: River bed degradation with armouring. Proc. Jpn. Soc. Civ. Eng. **195**, 55–65 (1971)
37. Huber, B.: Physikalischer Modellversuch und Cfd-Simulation einer asymmetrischen Drossel in einem T-Abzweigstück. Österreichische Wasser- und Abfallwirtschaft (2010). doi: 10.1007/s00506-010-0170-9
38. Hunziker, R.P.: Fraktionsweiser Geschiebetransport. In Vischer, D. (ed) Reports of the Laboratory of Hydraulics, Hydrology and Glaciology of the Swiss Federal Institute of Technology Nr. 138, Zürich (1995)
39. Hunziker, R.P., Fardel, A., Garbani Nerini, P.: HYDRO_GS-2D Mehrkorn: Modell und Versuchsbeschreibung: HYDRO_GS-2D Reference Manual (2009)
40. Jaeger, C.: Fluid Transients in Hydro-Electric Engineering Practice. Blackie, London (1977)
41. Jianhua, W., Wanzheng, A., Qi, Z.: Head loss coefficient of orifice plate energy dissipator. J. Hydraul. Res. **48**(4), 526–530 (2010)
42. Klar, R., Achleitner, S., Umach, L., Aufleger, M.: Long-term simulation of 2D fractional bed load transport: benefits and limitations of a distributed computing approach. In: 10th International Conference on Hydroinformatics (HIC), Hamburg (2012)
43. Klar, R., Umach, L., Achleitner, S., Aufleger, M.: Adaptive roughness approach for 2D long-term morphodynamic simulation. In: International Conference on Fluvial Hydraulics: River Flow, San José, Costa Rica (2012)
44. Lange, D., Bezzola, G.R.: Schwemmholz: Probleme und Lösungsansätze. VAW-Mitteilungen, Nr. 188 (2006)
45. Lecher, K., Lühr, H.P., Zanke U.: Taschenbuch der Wasserwirtschafte Hydromechanik Band 4: Hydraulische und numerische Modelle. Parey, Berlin (2001)
46. Lin, X.: Application of Energy Dissipation of Multi-Orifice in Xiaolangdi Project. Waterpower '99 (1999). doi: 10.1061/40440(1999)109
47. Mazzorana, B., Hübl, J., Zischg, J., Largiader, A.: Modelling woody material transport and deposition in alpine rivers. Nat. Hazards (2010). doi: 10.1007/s11069-009-9492-y
48. Meyer-Peter, E., Müller, R.: Eine Formel zur Berechnung des Geschiebetriebs. Schweizerische Bauzeitung **67**(3), 29–32 (1949)
49. Müller, D.R.: Auflaufen und Überschwappen von Impulswellen an Talsperren. Doctoral thesis, ETH Zürich, Nr. 11113 (1995). doi: 10.3929/ethz-a-001469940
50. Nalluri, C., Featherstone R.: Civil Engineering Hydraulics. In: Marriott, M. (ed) Wiley-Blackwell, Chichester (2009)
51. Nitsche, W., Brunn, A.: Strömungsmesstechnik. Springer VDI, Berlin (2006)
52. Novak, P., Moffat, A.I.B., Nalluri, C., Narayanan, R.: Hydraulic Structures. Taylor & Francis, London (2005)
53. Nujic, M.: HYDRO_AS-2D – Ein zweidimensionales Strömungsmodell für die wasserwirtschaftliche Praxis – Benutzerhandbuch (2009)
54. Nujic, M.: HYDRO_GS-2D – 2D-Geschiebetransportmodell, Arbeitsblatt Geschiebe (2009)
55. Plörer, M., Achleitner, S., Neuner, J., Aufleger, M.: Sedimentation and flushing of Alpine water intake reservoirs: 2D numerical simulation based on calibration data of a hydraulic model test. In: 35th IAHR World Congress, Chengdu (2013)
56. Plörer, M., Achleitner, S., Neuner, J., Mayer, R., Lumassegger, S., Aufleger, M.: Sedimentation and flushing of alpine water intake reservoir: hydraulic model test. In: 10th International Conference on Fluvial Sedimentology, Leeds (2013)
57. Pohl, R., Martin H.: Technische Hydromechanik Band 4: Hydraulische und numerische Modelle. Huss, Berlin (2008)
58. Pope, S.: Turbulent Flows. Cambridge University Press, Cambridge (2006)
59. Rickenmann, D.: Hyperconcentrated flow and sediment transport at steep slopes. J. Hydraul. Eng. **117**(11), 1419–1439 (1991)

60. Rickenmann, D.: Empirical relationships for debris flows. Nat. Hazards **19**, 47–77 (1999)
61. Rickenmann, D.: Comparison of bed load transport in torrents and gravel bed streams. Water Resour. Res. **37**, 3295–3305 (2001)
62. Rimböck, A.: Schwemmholzrückhalt an Wildbächen: Grundlagen zu Planung und Berechnung von Seilnetzsperren. Berichte des Lehrstuhls und der Versuchsanstalt für Wasserbau und Wasserwirtschaft, Nr. 94 (2003)
63. Schmocker, L., Hager, W.H.: Probability of drift blockage at bridge decks, J. Hydraul. Eng. (2011). doi: 10.1061/(ASCE)HY.1943-7900.0000319
64. Schmocker, L., Weitbrecht, V.: Driftwood: risk analysis and engineering measures. J. Hydraul. Eng. (2013). doi: 10.1061/(ASCE)HY.1943-7900.0000728
65. Seeber, G.: Das Wasserschloss des Kaunertalkraftwerkes der TIWAG: Ein neuer Typ eines rückstromgedrosselten Kammerwasserschlosses. Schweizerische Bauzeitung **88**(1), 1–8 (1970)
66. Seeber, G.: Druckstollen und Druckschächte: Bemessung, Konstruktion, Ausführung. ENKE, Stuttgart/New York (1999)
67. Sendlhofer, A.: Systematische Versuchsreihen zur Überprüfung der Verklausungssicherheit von Brücken. Diploma thesis, Unit of Hydraulic Engineering, University of Innsbruck (2010)
68. Strobl, T., Zunic, F.: Wasserbau: Aktuelle Grundlagen: Neue Entwicklungen. Springer, Berlin/Heidelberg (2006)
69. Thoma, D.: Zur Theorie des Wasserschlosses bei selbständig geregelten Turbinenanlagen. R. Oldenbourg, München Berlin (1910)
70. U.S. Army Corps of Engineers (USACE): HEC-RAS, River Analysis System, User's Manual v4.1. USACE-Hydologic Engineering Center (2010)
71. Ward-Smith, A. J.: Internal Fluid Flow: The Fluid Dynamics of Flow in Pipes and Ducts. Clarendon Press, Oxford (1980)
72. White, R.: Evacuation of Sediments from Reservoirs. Thomas Telford, London (2001)
73. Wilcox, D.: Turbulence Modeling for CFD. DCW Industries, La Canada, Califonia (2000)
74. Zhang, Q., Chai, B.: Hydraulic Characteristics of Multistage Orifice Tunnels. J. Hydraul. Eng. **127**(8), 663–668 (2001)
75. Zweifel, A.: Impulswellen: Effekte der Rutschdichte und der Wassertiefe. Doctoral thesis, ETH Zürich, Nr. 15596 (2004). doi: 10.3929/ethz-a-004770787

Chapter 9
A Genetic Algorithm Approach for the Rigorous Registration of Arbitrary Laser Scanner Point Clouds

K. Hanke and S. Schenk

Abstract Terrestrial laser scanners have achieved great popularity in the last decade. Their easy on-site application and the possibility of flexible and high quality post processing added to their success in several fields such as architectural, archaeological, and heritage documentation. We present a method for handling the automatic registration of point clouds which are characterized by significant noise level, generally imperfect geometry and occlusions. Hereby we combine and extend already existing and established methods to facilitate the registration of point clouds without prior pre-processing. Our approach consists—similar to other methods—of three steps which are scan analysis, pair-wise registration, and global registration. To handle the abovementioned datasets we propose to use imperfect and subdivided features, and to implement Genetic Algorithms (GAs). At the same time our approach can be seen as extension to already known Genetic Algorithms used for the registration of point clouds. By implementing an adapted version of a Genetic Algorithm in the classical registration process between rough alignment and fine registration, we are able to maintain robustness and computational performance also when registering point clouds of bigger objects characterized by a notably increased number of points, a significant noise level, and occlusions. We show and discuss the successful application of the algorithm on a scene which does not consist of classical geometric primitives such as planes.

K. Hanke (✉) • S. Schenk
Institute for Basic Sciences in Engineering, Unit of Surveying and Geoinformation, University of Innsbruck, Technikerstr. 13, A6020 Innsbruck, Austria
e-mail: Klaus.Hanke@uibk.ac.at; Stefan.Schenk@tiscali.it

G. Hofstetter (ed.), *Computational Engineering*, DOI 10.1007/978-3-319-05933-4_9,
© Springer International Publishing Switzerland 2014

9.1 3D Data Acquisition and Laser Scanning

Terrestrial laser scanners have got a significant standing in the three-dimensional recording of complex objects. By using a laser beam a terrestrial laser scanner can digitally capture the surrounding of the measurement device with the modest expenditure of time. The result is a collection of closely spaced three-dimensional points, representing the original object's surface with a discrete point cloud (see Fig. 9.1). Due to the steady technical improvement of laser scanners and the related software for analysis of the recorded data, the number of possible application areas is growing rapidly [63].

Laser scanners were developed and used mainly for digital inventories and documentation of plants and machinery [21] as well as in the field of manufacturing and quality control. Nowadays, however, they are increasingly found also in the documentation of architecture or archaeology [24] and the preservation of cultural heritage for different purposes.

Objects are generally recorded contactless in form of a "cloud" of dense discrete points on the measured surface using a focused laser beam. According to [50, p. 23] methods using light for object acquisition can be divided into active and passive ones. Within the passive process the capturing device (e.g. a digital camera) does not send out any radiation itself, whereas in an active approach radiation is emitted in a controlled way (e.g. laser) and its reflectance is being measured. Active methods are further divided in triangulation- and time-based methods. The latter can then once more be differentiated into pulse- and phase-based approaches. A description of the various methods and instruments may be found in [30, p. 17].

As for terrestrial laser scanning methods it is generally distinguished between static and dynamic laser scanning [30, p. 12]. Static working methods are characterized by the fact that during the acquisition process the instrument remains in a fixed position, while a high point density and a fairly high precision can be achieved. In dynamic laser scanning, however, the instrument is mounted at a mobile platform (such as a car or train) and the data acquisition is performed during the motion of this platform. For small objects, turntables as well as robot arms are in use. A discussion of the pros and cons of laser scanning is given in [22].

A typical terrestrial laser scanner is designed for average distances of about 2–200 m and consists of a range of different components. It is generally mounted on a tripod allowing a rotation of 360° around the vertical axis. The laser scanner and the whole data acquisition are controlled by a connected (ragged) mobile computer. An appropriate generator delivers the necessary power for all components and is therefore essential for all outdoor applications.

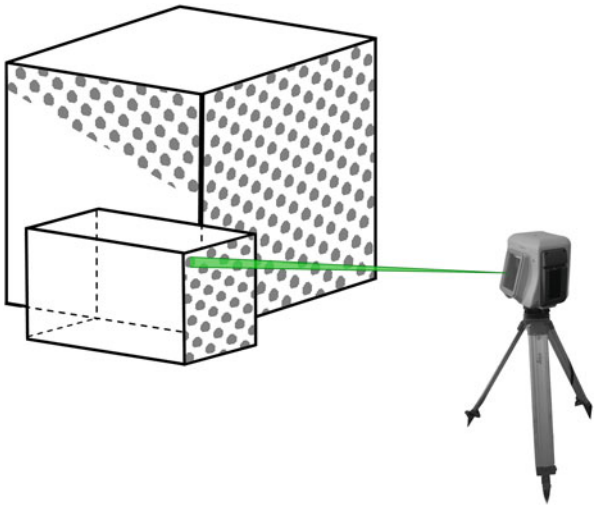

Fig. 9.1 Typical terrestrial laser scanner setup

9.2 Registration of Point Clouds

Laser scanning is typically a polar measurement method. Hereby 3D vectors to the object's surface are measured from a single survey station (position) of the measurement instrument (laser scanner); by doing so the spatial angles and distances from the instrument to the object point are acquired. However, only the part of the object that is visible from the instrument's position can be acquired at one time. To complement the data and get an overall recording of the object several standpoints of the instrument are necessary to avoid hidden parts and occlusions in the final data set. To merge these different point sets ("point clouds") to a joint point cloud of the surface in a common coordinate system a so-called "registration" is necessary (see Fig. 9.2). This means that a spatial transformation from the individual local coordinate systems of the gathered data into an overall global coordinate system needs to be done. As this task is a non-linear operation the work flow is generally separated into a rough alignment of the respective single point clouds and a subsequent fine registration.

According to [31], this task can be seen as a search for the corresponding Euclidean motion m and can be expressed as a translation vector $\mathbf{t}$ and rotation matrix $\mathbf{R}$ in the form of

$$m(\mathbf{x}) := \mathbf{x}' = \mathbf{t} + \mathbf{R} \cdot \mathbf{x} \,. \tag{9.1}$$

Hereby $\mathbf{x} \in \mathbb{R}^3$ represents the given coordinates in terms of a three-dimensional vector and $\mathbf{x}'$ its correspondence after the transformation procedure into the global coordinate system. Since laser scanner point clouds generally are already measured

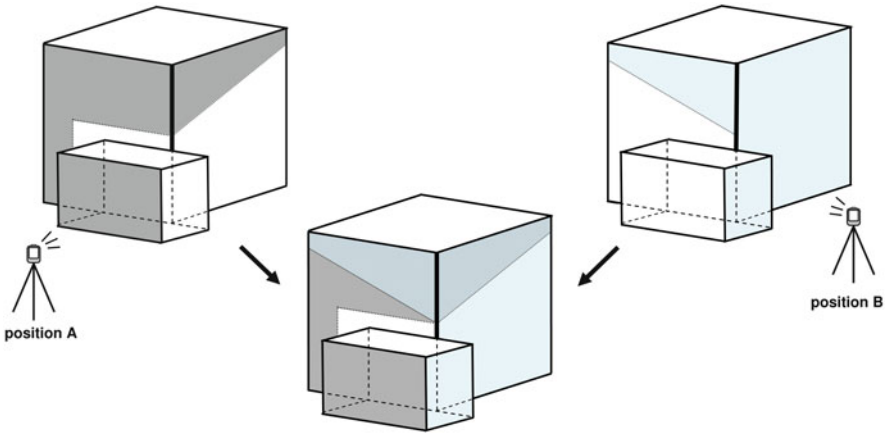

Fig. 9.2 Point cloud registration

in the real world scale of 1:1, no further scaling has to be considered. Similarly to [51, p. 48] using the vectors

$$\mathbf{x}' = [x', y', z']^T,\ \mathbf{x} = [x, y, z]^T,\ \mathbf{t} = [t_x, t_y, t_z]^T \tag{9.2}$$

and $r_{ij}, i = 1 \ldots 3, j = 1 \ldots 3$ as the elements of the 3×3 rotation matrix $\mathbf{R}$, Eq. (9.1) can be shown in matrix notation in the form

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} + \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} . \tag{9.3}$$

In the course of the registration the coordinate transformation (Eq. (9.1) or (9.3)) must be performed and all respective parameters determined for the captured point clouds. Therefore, a translation vector $\mathbf{t}$ and a rotation matrix $\mathbf{R}$ have to be determined separately for each point cloud; so all point clouds can be aligned optimally and the original object sufficiently and accurately represented in its entirety.

A robust and widely used method for the determination of the required parameters is to use artificial objects in the form of special planar signs, spheres, or cylinders in the immediate vicinity of the object to be recorded. During the data acquisition process at least three well spatially distributed targets have to be scanned additionally from each position. For this step usually a higher point density and accuracy is used which may take a lot more time. The centers of these targets are then used as tie points for the determination of the required transformation parameters in $\mathbf{t}$ and $\mathbf{R}$.

By using a sequential procedure, point clouds can be joined one after another or the positions of all point clouds to each other can be determined simultaneously.

Although this method is considered to be state-of-the-art, it has unique disadvantages in a number of practical applications.

Setting up the laser scanner and the artificial targets for the scanning of convoluted or complex objects and surfaces (especially with non-horizontal set-ups of the laser scanner) may result, for example, in a test of patience to place the three required artificial targets for each position in a way that they are also visible and measurable from the next position of the laser scanner. In some cases the immediate surrounding of an object is accessible only with difficulty or not at all (e.g. for objects in danger zones), which entirely prevents a wise distribution of artificial targets over or around the object to be scanned or renders it at least highly difficult. A further disadvantage is the relatively high cost, not only for the sensible placing of the targets but also for selecting and scanning them with a high density of points and increased accuracy.

These steps require in many cases much more time than scanning the actual object. If the abovementioned procedure is not feasible or provides a "false" result (because an artificial target was, e.g., moved inadvertently), at least three possible matching pairs of points in each pair of point clouds have to be hand-picked manually. These can be used to carry out the rough alignment process afterwards. Especially if many point clouds are to be registered and in these no distinctive corners and edges are clearly identifiable, this procedure is extremely time consuming and prone to errors, and should be used only as a last resort.

The goal of the here presented approach is to provide a concept, which—without the prior knowledge of approximate solutions and without the use of artificial targets but by using a Genetic Algorithm—will provide an automatic registration of overlapping and unordered point clouds, which are characterized by a significant noise level, imperfect and incomplete geometry, as well as occlusions and shadowing effects.

9.3 Automatic Registration

Automatic registration strategies are generally classified either into "rough alignment" or "fine registration." Genetic Algorithms are, however, a special case as they can be adapted to both rough alignment as well as fine registration.

Rough alignment doesn't need previous knowledge about the relative position of point clouds, whereas fine registration takes the result of rough alignment and refines the solution. To encounter the complexity of non-linear registration equations and enormous data volume, most algorithms are specialized on one of these two tasks. One of the challenges of rough alignment is that point clouds can be of any size, shape, or property. It is obvious that with automatic strategies only point clouds sharing common areas can be registered.

According to [53] algorithms can further be classified into "pair-wise registration" or "global registration" (multi-view registration). Pair-wise registration algorithms are specialized in handling two point clouds at the time, whereas a global registration algorithm registers more or even all point clouds simultaneously.

9.3.1 Genetic Algorithms

Genetic Algorithms are known to be a nature-inspired heuristic search method [34, p. 11] reaching great popularity through the work of [32]. According to [55, p. 46] a Genetic Algorithm (GA) is a computational model of natural evolution where stronger individuals in a competitive environment are more likely to survive than weaker ones. The basic principle—the "survival of the fittest"—implements the concepts of mutation, cross-over, and (natural) selection and was originally introduced by the English naturalist Charles Darwin (1809–1882).

With Genetic Algorithms, "individuals" represent single solutions to a given optimization problem, whereas a group of solutions is referred to as "population." The fitness (quality) of the individuals is determined by their chromosomes. Mathematically, such a chromosome is a vector $\mathbf{X} = [x_1, \ldots, x_n]$ representing the characteristic information of the solution as binary or real numbers. Among the several advantages of Genetic Algorithms is their flexible adaptation to different optimization tasks and their ability to find approximated solutions also in search spaces where other methods may fail. Further, the risk of getting stuck at local optima can be significantly reduced. It has, however, to be added that Genetic Algorithms are generally known to be a fairly intensive technique from a computational point of view.

9.3.1.1 Search in Correspondence Space

Among the several approaches regarding the use of Genetic Algorithms for point cloud registration, two common concepts can be identified: those algorithms performing the search in correspondence space and those searching for transformation parameters in transformation space. A Genetic Algorithm performing the search in correspondence space is described in [7]. Hereby a single chromosome (individual) consists of a multi-index linking the point indices of two point clouds. Such an approach is especially useful when free form-shapes are given and no other ideal geometrical shapes are available for registration. The quality of each individual is calculated by a so-called "fitness-function" f which is like a cost function and it is based in this case on the pair-wise match quality of all linked points; more details can be found in [7]. According to [53] such Genetic Algorithms can achieve good results in the rough alignment of point clouds; results that can accordingly be used for fine registration. It is, however, highly computationally expensive whenever a large number of points are used.

9.3.1.2 Search in Transformation Space

Whereas searching in correspondence space is based on finding corresponding point indices, searching for transformation parameters provides a different way of

encoding a registration solution in the form of $\mathbf{X} = [\alpha_x, \alpha_y, \alpha_z, \Delta_x, \Delta_y, \Delta_z]$. Hereby $\alpha_x, \alpha_y, \alpha_z$ represent the three Euler angles of the rotation matrix $\mathbf{R}$ and $\Delta_x, \Delta_y, \Delta_z$ the components of the translation vector $\mathbf{t}$. Such a representation is proposed in similar way for example in [11]; hereby the fitness-function is based on the median of Euclidean distances. An advanced strategy is discussed in [55]. Given two point clouds A and B, the fitness-function f is defined by

$$f_{(A,B)} = \frac{1}{N} \sum_{i=1}^{N} \rho(r_i) \tag{9.4}$$

where ρ is the residual term

$$\rho(r_i) = \begin{cases} r_i & \text{if } r_i < d \\ d & \text{if } r_i \geq d \end{cases} . \tag{9.5}$$

N represents the number of points, d a given threshold and r_i the squared Euclidean distance between a point in point cloud A and its nearest neighbor in point cloud B. This fitness-function is robust against outliers and therefore suitable for a lot of cases. To speed up application, a hill-climbing strategy is implemented; moreover the so-called "surface interpenetration measure (SIM)" helps the Genetic Algorithm to face fine registration. Overall this approach leads to very good results although only a few thousand points can be handled effectively.

Some authors extend the search in transformation space by searching also for a scaling factor S [13] or an unknown overlap factor ξ. Latter is presented by [40]; hereby the mean squared error $e(\xi)$ between two point clouds is minimized while the overlap ξ has to be maximized.

9.3.2 Rough Alignment

Rough alignment is used to register point clouds in such a way that a fine registration algorithm can successfully refine the resulting solutions.

One of the fastest ways to register point clouds is to use a principal component analysis and match both the barycenters and the main axes of the point clouds. This method, however, works only if the point clouds share a high degree of overlap. Only partially overlapping point clouds and occlusions will, in most cases, lead to completely wrong solutions.

Fischler and Bolles [17] presented the RANSAC (random sampling consensus) algorithm to detect robustly outliers in given measurement datasets. Chen et al. [9] extended this method enabling it to find three pairs or corresponding points in two point clouds which can subsequently be used for their registration. Though the so-called DARCES algorithm (data-aligned rigidity-constrained exhaustive search) includes, moreover, geometric compatibility restrictions rendering the approach

more efficient, the amount of points is still fairly limited in comparison with other approaches.

Several authors proposed to calculate the so-called point signatures which are more robust against noise. To identify the signature of a determined point, Chua and Jarvis [12] intersect the surface of the point cloud with a virtual sphere constructed in the considered point. The perpendicular distances of the so created curve and the tangent plane in the examined point can be used for point signature generation. Similarly [19] use the "integral volume descriptor" which can be determined by calculating the volume between the virtual sphere and the point cloud's surface. By doing this for several sphere radii, the informative value of the point signature can be increased.

The term "spin image" was coined by [36] and describes a method for generating a two-dimensional frequency matrix employable for fast point comparison. To create this matrix in a single point **p**, the perpendicular distances from all neighbor-points to the tangent plane in **p** as well as to the normal vector of **p** are used. Such point signatures lose, however, part of their significance in the presence of noise or when part of the neighborhood is not equally visible from other stations.

Another possibility for point cloud registration arises by using linear elements (features). Linear features are especially found in architectural datasets and can be detected directly [62], by intersecting planes [56] or by looking at the borders of planar elements [37].

Alternatively, the planes themselves are also common features for registration purposes. Three corresponding pairs of them are necessary for the successful registration of two point clouds. Planes have the advantage that they are quite robust against noise, but the three plane pairs have to be linearly independent. Such a strategy can be found in [6]. He et al. [29] proposed to include the barycenters of plane patches into the registration process, too. This way only two corresponding plane patch pairs are needed. To reduce the number of possible combinations, the area of the plane patches, angles or distances can be taken into account. Problems may arise if plane patches are partly occluded as different barycenters will emerge. However, by applying the here proposed concept of imperfect and subdivided features, this restriction can be evaded.

In several cases additional information about the scanner set-up is available. According to [61] the zenith direction may be used to reduce the number of needed plane pairs from two to one. In some cases the same information can also be recovered from the point cloud contents themselves (e.g., by a dominant ground plane). Such information can considerably speed up the registration process. Also the here proposed approach includes the possibility to use such additional data in the algorithm.

Von Hansen [61] suggested subdividing point clouds into 3D raster cells and calculating for each grid cell its dominant plane with a RANSAC-algorithm as seen, e.g., in [58]. By doing so, also large point clouds can be handled very efficiently. The here presented registration approach can be seen as further development of this method.

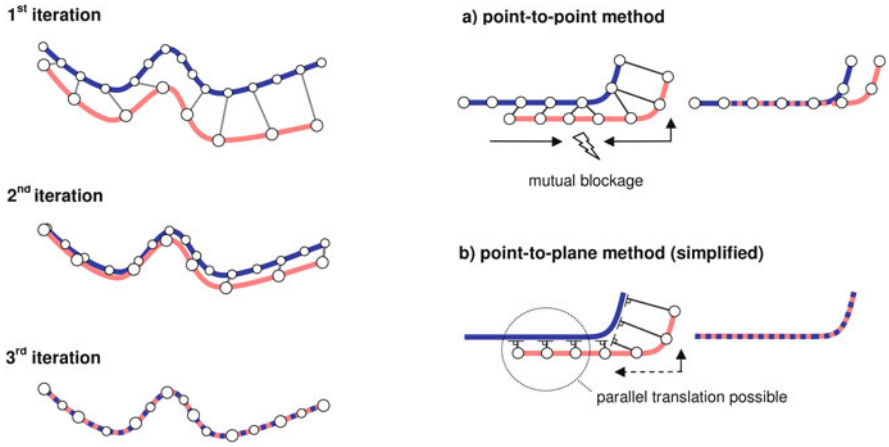

Fig. 9.3 The ICP-algorithm and its two variants

Especially in the fields of industry a lot of geometrical figures in form of planes, cylinders, toroids and spheres can be found. Rabbani et al. [48] presented a registration strategy for such environments including the modeling of the geometric figures as well as simultaneous registration. In cultural heritage documentation, however, such figures are less prominent.

9.3.3 Fine Registration

The goal of fine registration is to obtain the most accurate registration solution possible [53]. One of the most popular fine registration methods employs the Iterative-Closest-Point algorithm (ICP) presented by [4]. Starting from a given initial estimation for each point in a first point cloud, the nearest neighbor in a second point cloud is looked for. According to [33] these correspondences can be used to realign the point clouds in such a way that the sum of the squared point-to-point distances between the point clouds is minimized. Executing these steps iteratively, a convergence of the solution can be expected (see Fig. 9.3). The quality of the registration result, however, depends on several parameters such as the initial estimation and the specific properties of the point clouds. A similar algorithm, not focusing on point-to-point but point-to-plane distances, is presented in [8]. This point-to-plane algorithm adopts so-called "normal shooting" and calculates the squared distances from a point in A along its normal to a surface in B [28, pp. 36–37]. There are, however, also simplified versions calculating, e.g., the distance between a point in A and the tangent plane constructed in its nearest neighbor point in B. To calculate the realignment matrix after each iteration, no closed-form solution is available for the point-to-plane algorithm; therefore,

e.g., the Levenberg–Marquardt algorithm or a linearization has to be applied. The basic advantage of point-to-plane distances is that, while refining the registration, translation along the tangent plane of the surface will occur with more likeliness than with the point-to-point approach (see Fig. 9.3). This can make the algorithm more robust regarding outliers and decrease the number of iterations necessary for convergence. An overall quality or behavior judgment is, however, difficult because the point-to-point as well as the point-to-plane ICP have strengths and weaknesses which also depend on the specific properties of the given point clouds.

Since the publications of [4] and [8] a lot of attempts were made to render algorithms faster, more flexible and more robust. Rusinkiewicz and Levoy [52] summarized the most renowned variants of ICP-algorithms which differ in point selection, matching, weighting, point rejection, error metric, and minimization of the error metric. Although all of those suggestions can be implemented into one algorithm, this does not necessarily lead to higher quality results.

A further challenge when dealing with fine registration is the selection of the thresholds for search, weighting, and correspondence rejection. This can be partly avoided by using a statistical approach for the dynamic handling of thresholds [64]. Typical ICP-algorithms allow the fine registration of two point clouds only. Using velocity vector fields [44] or embedding the generalized procrustes analysis [59], also multiple point clouds can be registered simultaneously.

The results of ICP-algorithms heavily depend on the selection of correspondences. Gelfand et al. [18] propose to use only correspondences which can contribute to a geometrically stable solution. Another robust approach for handling noisy or only partly overlapping measurements uses trimmed least squares and is described in [10]. Hereby an approximated overlap parameter is calculated by minimizing an objective function with a Golden Section Search algorithm. For cases where an enlarged convergence region of the ICP-algorithm is needed [2] presents the Geometric Primitive ICP, which is based on geometric primitives (surface normal vectors, curvature, and change of curvature) and neighborhood search.

9.3.4 Imperfect and Subdivided Features

As mentioned above, laser scanners provide an efficient way to capture complex three-dimensional objects. The achievable quality, however, is still an open question and depends on several parameters. According to [30, p. 28] the instruments' accuracies can vary from instrument to instrument. Among the main error sources for laser scanning there are instrumental errors (laser beam propagation, mixed edge problem, range-uncertainty, angular uncertainty, axes error), object-related errors, environmental conditions (temperature, atmosphere, interfering radiation, distortion from motion), or methodological errors. As illustrated in [23] also surface properties such as the color contribute to the overall quality of the point cloud. More information is found in [20, p. 28].

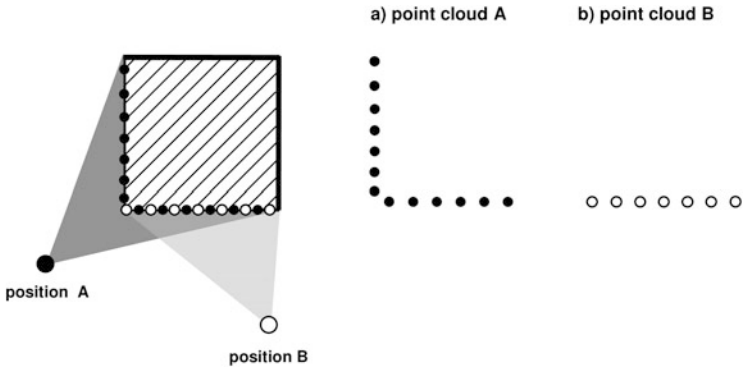

Fig. 9.4 Scanning a border from different positions

To avoid misunderstandings in this work the following conventions are made. The "measurement uncertainty" (sensor uncertainty) is the result of all influences which cause an ideal planar object looking jagged in the captured point cloud. By scanning surfaces (point-wise as discrete grid) with a lower point density, time and data volume can be saved. The so-created point clouds of rough surfaces, however, may seem to follow a random distribution of points ("surface noise"). Together with measurement uncertainty all these effects are generally referred to as "noise."

In order to save precious money and time when scanning objects, the amount of different laser scanner stations is generally held as low as possible. Due to this reduced number of stations, however, occlusions may occur in an increased quantity. When there is, e.g., more than one object in the scanning area of the laser scanner or the object covers itself in parts, such occlusions can arise (see Fig. 9.4). Moreover, with rough object surfaces important details may get lost. Furthermore, in many documentation tasks no perfect geometric figures are given. Especially in rural areas, planes, cylinders, spheres, and other figures are rare. On the contrary, objects are often composed of complex free-form surfaces. Further, they can be characterized by the abovementioned noise. This is labeled "imperfect geometry." "Incomplete geometry," however, refers to geometrical elements which are visible only in part.

The here discussed approach is especially about cases where only a low point density is given or the quality of borders, edges, and other local details may not be usable for the registration of point clouds.

9.3.4.1 Imperfect Features

When trying to automatically register point clouds characterized by significant noise level, imperfect and/or incomplete geometry as well as occlusions, the choice of a feasible registration strategy can be challenging. In such cases more complex features like planes instead of single points are generally preferable. This is shown

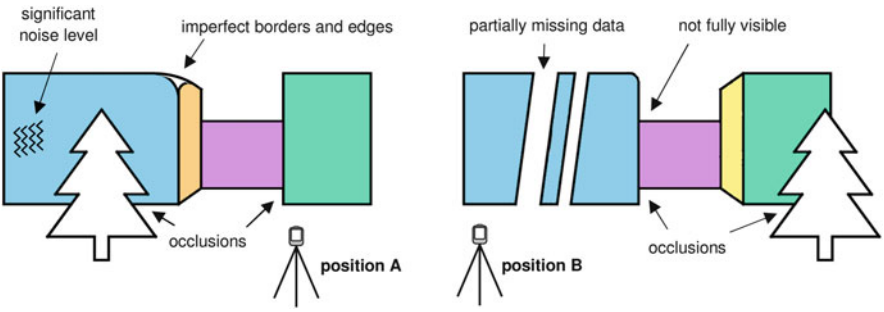

Fig. 9.5 Typical problems in data acquisition

in [6, 29], and [61]. Figure 9.5 illustrates typical problems with the generation, registration and post-processing of point clouds. Especially in urban areas human built objects consist of parallel planes and thus render registration more taxing. Moreover, a high noise level and incomplete or incorrectly captured surface areas can influence negatively on the overall registration process. Incomplete borders and/or edges may appear differently in point clouds of different stations.

The term "imperfect feature" refers to a characteristic area of a point cloud (e.g., plane patch) which can be used for the registration process; one has, however, to keep in mind that it may misrepresent the original object. Following this indication, most features in real-life point clouds can be detected only approximately. It thus makes sense to elaborate an overall registration concept that enables to handle imperfect features.

9.3.4.2 Subdivided Features

Most objects contain a variety of characteristic features which can be used for registration purposes. Through a significant noise level, disadvantageous occlusions, or incomplete datasets the number of such features can get rather low. Therefore, we propose to subdivide features into smaller parts. Those parts that aren't influenced anymore by occlusion or other perturbances are called "subdivided features" and can be used for the automatic registration. This way a higher robustness of the features and a greater flexibility can be achieved. The proposed approach differs from the "complete plane patches (CPP)" mentioned in [29] as we propose to subdivide features (in this case plane patches) into smaller sub-parts. The subdivision of point clouds was already mentioned in [61]. Hereby the point clouds are split into 3D raster cells based on the local coordinate system of each point cloud. For all cells a dominant plane (surface element) is calculated and neighboring elements are afterwards grouped to larger plane patches for faster processing. We suggest developing this approach further and use subdivided features directly for the automatic registration. At the same time we propose to maintain the concept

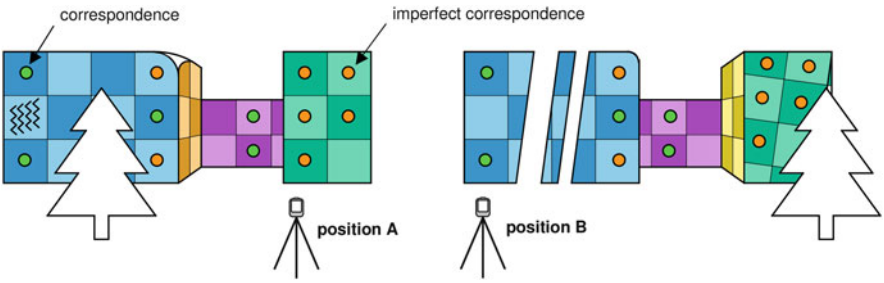

Fig. 9.6 Subdivided features

of imperfect features, so that also not ideal geometric figures can be used in the registration process.

As we want to increase the chance that subdivided features will be similar in different point clouds, we identify first all planar patches in a point cloud and then calculate their barycenters and their principal axes. Based on this information for each plane patch a local coordinate system for the subdivision is established. This way fully visible plane patches may have similar subdivided features with corresponding barycenters (see Fig. 9.6). In this work such cases are called perfect correspondences, whereas all other cases where barycenters don't match that well to a different subdivision are called imperfect correspondences.

Different authors showed that also other geometric elements can be used for the registration of point clouds. So [56] and [61] used linear elements whereas [47] used among others cylinders. Also in those cases the concept of imperfect and subdivided features would be applicable.

9.4 Registration Strategy

Various challenges can be encountered during the automatic registration of point clouds without known approximations or artificial targets. Not only the huge quantity of data, but also the direct connection between the aspired correspondences and the simultaneously looked for transformation parameters is not unproblematic. GAReg-ISF ("Genetic Algorithm Registration with Imperfect and Subdivided Features") represents one possible approach and is described in detail in [54]. It is especially designed for datasets with a significant noise level, imperfect and incomplete geometry as well as occlusion. The method consists of three main parts as seen in Fig. 9.7. Every puzzle piece represents schematically a single point cloud. Similar algorithm structures are also used in [35] and [61].

First each point cloud is analyzed and characteristic information is gathered. During the so-called pair-wise registration iteratively two point clouds are registered with each other. Finally, in the global registration the results of the pair-wise regis-

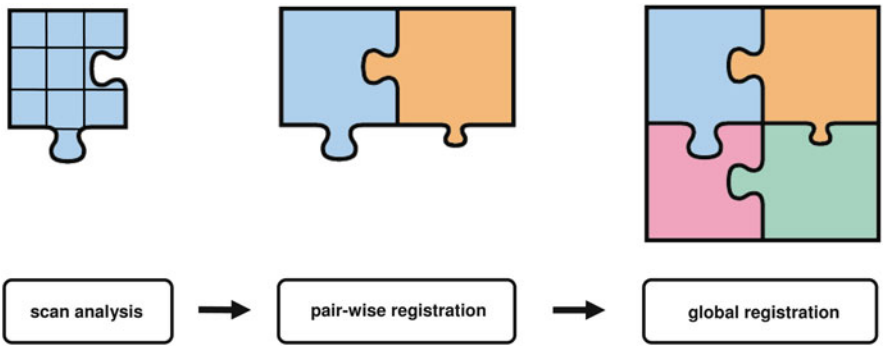

Fig. 9.7 Automatic registration strategy

tration are merged to a consistent digital representation of the original object. One of the big challenges in automatic point cloud registration is that the relationship between the single point clouds is initially unknown and has to be reconstructed as well during the registration. The basic objective of our approach is to enable also the registration of bigger objects with reasonable computational effort. To reach this target we combined the advantages of already established approaches while trying to minimize their drawbacks.

9.4.1 Scan Analysis

GAReg-ISF was designed to handle different file types of point clouds. Typically these files consist each of an unsorted list of single points [30, p. 67] which have to be pre-processed for later registration. The main import parameters as well as those for the other registration steps are supplied in GAReg-ISF by an external configuration file.

As published in [54, p. 73], the single steps of the scan analysis in GAReg-ISF contain the creation of kd-trees [3] for a faster nearest neighbor search, the calculation of normal vectors as well as the detection of imperfect and subdivided features. Also a smoothing of the point clouds is possible whenever the maintenance of edges and borders [39] is useful for registration purposes.

The huge amount of data is one of the challenges in automatic point cloud registration. Further, point clouds are typically characterized by varying point densities which may influence the automatic registration process. Haring [28, p. 49] e.g. shows a voxel-based approach to reduce the number of points and generate a homogeneous point density all over the point cloud. A too low point density, however, means to lose details, whereas a high number of points results in considerable computational effort.

By generating a data pyramid, each registration step can be executed with the most useful point density. As our examples showed [54, p. 135], GAReg-ISF works well with 20,000–100,000 points with regard to imperfect and subdivided feature detection and with up to 2,000 points (or features) with regard to the employed Genetic Algorithm. Similarly, other authors also use 100 [7] up to 10,000 points [55, p. 85].

For each level in the data pyramid a so-called kd-tree [3] is generated. According to [28, p. 42] a kd-tree is a data structure for saving points of k-dimensional space which is especially efficient for nearest neighbor searches. A popular implementation method can be found in [1].

Most automatic registration algorithms make use of normal vectors to improve the registration process. For calculating an approximation of the three normal vectors $\mathbf{n}_1$, $\mathbf{n}_2$, $\mathbf{n}_3$ in a considered point $\mathbf{p}$, first its neighboring points have to be found. This computationally intensive and time consuming step can be improved by using the already generated kd-tree or, in cases were during point cloud import a mesh was created, by implementing region growing. As [47, p. 36] and [2, pp. 62–63] mention, about 30 points are a good number for normal vector approximation.

Possible approaches for normal vector estimation can be found in [2, p. 47] or [28, p. 46]. For a given point $\mathbf{p}$ its n nearest neighboring points are stored in a list N_p. Next, the barycenter of all the points in N_p is calculated and is then used to form a covariance matrix $\mathbf{C}$ according to [43]. Carrying out an eigenvalue analysis, the eigenvector of the smallest eigenvalue $\boldsymbol{\lambda_1}$ is the normal vector $\mathbf{n}_1$ of the best-fit plane in $\mathbf{p}$. $\boldsymbol{\lambda_2}$ and $\boldsymbol{\lambda_3}$ lead to the normal vectors $\mathbf{n}_2$ and $\mathbf{n}_3$ which may be used to form a local coordinate system. GAReg-ISF defines normal vectors as positive if they look in the direction of the laser scanner position; this approach is mentioned in [28, p. 46]. It is quite obvious that normal vector estimation may be influenced by a high noise level.

As [43] show, normal vectors can also be used to get an approximation of the surface curvature, called "surface variation" σ_n. It can be calculated for the neighborhood of a point $\mathbf{p}$ by

$$\sigma_n(\mathbf{p}) = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} \tag{9.6}$$

whereby λ_i, $i = 1, 2, 3$ are the eigenvalues of $\mathbf{C}$ with $\lambda_1 \leq \lambda_2 \leq \lambda_3$ [43]. In GAReg-ISF the surface variation is used to find planar (flat) areas.

Generally, two- or three-dimensional geometric features contain more information than single points and are therefore preferable in many cases for automatic registration. In the following sections especially the use of imperfect and subdivided planar patches is described.

The detection of planes can happen in different forms. Von Hansen [61] introduces a 3D raster and calculates the dominant plane patch in each raster cell by using a RANSAC-algorithm before merging neighboring co-planar planes. The approach is very fast and works also for millions of points [61]. By just looking at the dominant plane in each raster cell, smaller planes are, however, omitted.

For objects showing a lot of small planes or complex geometry, region growing or voxel growing [14] are possible approaches. In GAReg-ISF a simplified method of the region growing steps described in [60] are used. Hereby all points are first sorted in a list according to their surface variation. Each point can be seen together with its normal vector as small plane patch approximating the local surface. A starting point (starting plane patch) is selected and neighboring points are added as long as they are compatible with the starting plane patch. If no point can be added anymore, the planar patch is saved and the next starting point is chosen from the sorted list.

To follow the principle of imperfect features, region growing is executed with a higher tolerance (typically between two and five times higher) than it would be necessary. This way also slightly "defective" surfaces can be approximated as planar patches. The region growing parameters are, however, selected empirically and may be adjusted by the user.

During region growing also the normal vectors $\mathbf{n}_1$, $\mathbf{n}_2$, $\mathbf{n}_3$ of each detected plane patch are stored. These can be used to create a local coordinate system with its origin in the barycenter of the plane patch. Those local coordinate systems are then taken to subdivide the corresponding planar patches into smaller sub-parts. The difference between this approach and the method described in [61] is that the subdivision is not performed on a global coordinate system, but for all planes individually.

By doing so, if a plane is fully visible in other point clouds, subdivision will take place in equal (or quite similar) manner. In other cases, GAReg-ISF is able to handle the possible matching uncertainties (see Sect. 9.4.2.2).

9.4.2 Pair-Wise Registration

The automatic registration of point clouds can be described as search process in six-dimensional space. To solve this complex task a separation of different registration steps is suggested. It can, for instance, be useful to align first only pairs of point clouds (pair-wise registration) and then use these pairs to create the digital representation of an object (global registration).

As discussed in Sect. 9.3, pair-wise registration is generally split into rough alignment and fine registration. Hereby, rough alignment is often done by using some kind of features. The results are then optimized during fine registration. By separating these processes and using specialized algorithms, the overall computational complexity can be reduced while at the same time the robustness of the solution is increased. But in case of significant noise level, incomplete and imperfect geometries, or occlusions, this separation may be problematic. As the features used during rough alignment can be distorted by the discussed influences, this can lead to a "wrong" rough solution.

Moreover, in this stage due to the introduced simplifications the "global" optimum may not be clearly detectable or may seem "worse" than an only local optimum. If the wrong rough solution is selected for refinement, the fine registration algorithm is hardly able to identify it and may return completely wrong results.

An alternative approach includes the use of Genetic Algorithms. They work on the point cloud itself so there is no restriction to specific geometric figures. Further, they are known to be quite robust in finding the global optimum in search space. There is, however, the drawback that Genetic Algorithms are computationally expensive resulting in not being able to work on the original point cloud but on a highly reduced one only.

Therefore, we suggest combining the advantages of rough alignment using geometric features such as plane patches and Genetic Algorithms. First, rough alignment with imperfect and subdivided features is executed. Afterwards, the results are used to mark promising areas in search space. By using the knowledge of rough alignment, the Genetic Algorithm can be applied in a very targeted and efficient way. Its main task is to improve the quality of the rough registration results while iteratively reducing their quantity. The aim is to identify the true "global" optimum on a step by step basis.

As mentioned in Sects. 9.3.4.1 and 9.3.4.2, adopting imperfect and subdivided features may result in "approximations." By using a Genetic Algorithm for optimization these are reduced before passing them on to a fine registration algorithm.

There are, however, cases where the rough alignment process is already able to identify the correct "global" optimum. The Genetic Algorithm will then terminate early as no improvement of the quality will occur. This case then presents the typical approach of rough alignment followed directly by fine registration.

9.4.2.1 Rough Alignment

In Sect. 9.4.1 the detection of imperfect features (plane patches) was discussed. For robust plane patch identification a minimum number of 90 points for each plane patch was considered feasible; moreover, it was agreed on the fact that each subdivided part may consist of more than 30 points. Both values can be adapted by the user to the specific properties of the point cloud.

Rough alignment in GAReg-ISF works on undivided features first and only afterwards on the subdivided features. This way computation can be speed up. In case of a high number of plane patches it is thinkable to use only the largest plane patches (i.e., 50) for rough alignment as proposed in [6]. In our test, however, we were able to successfully work with all of the detected plane patches.

As [6] show, it is necessary to detect three corresponding plane patch pairs in two point clouds to form a valid registration solution. When implementing also the barycenters of the plane patches, however, He et al. [29] state that two corresponding pairs are sufficient. Nevertheless, with each point cloud containing about 50 plane patches this results in over three million possible combinations. It is, thus, advisable to introduce hierarchical testing to reduce the number of possible solutions.

He et al. [29] propose to calculate and compare the angles between the normal vectors of plane patches in the two point clouds. As threshold a value between 2 and 3° is mentioned in [6]; in case of rough surfaces also values up to 15° can be feasible. Also the plane area [29] or the plane circumference [15] can be compared.

This will, however, fail, if a part of the (undivided) plane patch is occluded. For this reason it was omitted in the here presented approach. Further, the laser beam reflection value (intensity) and the surface variation can be used for eliminating completely contradictory correspondences.

In a next step the remaining combinations are refined by looking at their subdivided features (plane patches). As already mentioned, due to the subdivision of features the maximum distance error between the barycenters of the subdivided patches is limited, relating to the subdividing raster cell size. Brunnström and Stoddart [7] propose to compare four invariants between two point pairs and their normal vectors to eliminate wrong correspondences. The same approach can be applied also to subdivided plane pairs with their barycenters and their normal vectors. We set the angle threshold in our tests to 3°, whereas for the distance half the raster size was used.

By looking at the neighborhood of each subdivided plane patch a so-called neighborhood matrix can be created. Basically this is a 3×3 matrix containing information about the relative distribution of points (or areas) in eight different directions around the subdivided patch (see [54, p. 95] for details). This way features near borders or edges can be fast identified and wrong correspondences discovered.

For the remaining subdivided plane pair combinations the registration solution in form of the registration matrix $\mathbf{R}$ can be calculated as described in [33] and [6]. Software implementations herefore can be found, e.g., in the "Pointcloud Processing Toolbox" [46] as well as in commercial software packages like Innovmetric Polyworks™.

As [5] show, the solutions with the highest number of overlapping features don't necessarily also represent the global optimum of the registration. Due to approximations and uncertainties the correct solution may look worse than actually "wrong" solutions. By supplying all results to a Genetic Algorithm, this problem can effectively be counteracted.

9.4.2.2 Genetic Algorithm

The use of Genetic Algorithms is especially advisable when the search space is particularly large and contains lots of local optima. They are able to find approximated solutions where other algorithms may fail. We propose therefore to implement a Genetic Algorithm right in between rough alignment and fine registration. Rough alignment identifies promising areas in the search space and enables a more directed application of the following Genetic Algorithm, thus resulting in higher robustness and less computation time. The here presented Genetic Algorithm reduces and refines solutions simultaneously. This way the "correct" solution may become more apparent and pseudo solutions can be eliminated more efficiently.

The implemented Genetic Algorithm was originally designed for the optimization of tunnel designs [49]. Only a small part of the algorithm required is being

adapted to the registration of point clouds which proves once more the high flexibility of Genetic Algorithms. The algorithm is able to work with both imperfect and subdivided features as well as single points. In case of planar patches the former ones can be represented by their barycenters so that the handling is similar to the use of single points. Including also normal vectors into the registration process, a greater robustness can be achieved. By default GAReg-ISF uses the approach with imperfect and subdivided features as they generally occur in a lower number than single points which leads to faster processing. It has, however, to be mentioned that the purpose of the Genetic Algorithm in GAReg-ISF is not fine registration as such but to serve as intermediation between rough alignment and fine registration.

Most Genetic Algorithms show a typical structure. Our algorithm is designed similarly to the structure proposed in [55, p. 47]. One of the biggest challenges is, however, to define a chromosome encoding for the specific task. Similarly to [55, p. 49], a possible registration solution (individual) is defined by $\mathbf{X} = [\theta_x, \theta_y, \theta_z, t_x, t_y, t_z]$, where $\theta_x, \theta_y, \theta_z$ represent the three Euler angles and t_x, t_y, t_z the components of the translation vector. According to the suggestions in [40] regarding the drawbacks of Euler angles we substitute them with a rotation quaternion $\mathbf{Q}$ which leads to

$$\mathbf{X} = [\mathbf{Q}, t_x, t_y, t_z] \; . \tag{9.7}$$

Genetic Algorithms are typically initialized by creating a set of random individuals (start population). In our approach this step may be skipped in cases where rough alignment already provides a large set of rough results. If not already done, a kd-tree is now created to speed up the overall process. All solutions are then evaluated by the fitness-function (see Eq. (9.4)) and subsequently ordered according to their quality. Next, highly similar individuals are grouped by comparing their chromosomes; those showing the highest fitness in each group are used for the start population whereas the rest is eliminated. The maximum number of generations was set to 200 following [55, p. 55]. An earlier termination of the generation loop will, however, take place when the solutions converged (see Fig. 9.8) or the quality of the solutions can't be improved for five generations.

GAReg-ISF implements several genetic operators. The main purpose of each operator is to create new individuals (chromosomes) which may be better than their parent(s). A mutation operator is applied in 10 % of all cases and as for the other 90 % [55, p. 54] the cross-over operator is chosen. Further, all individuals are randomly mutated by a chance of 2 % [55, p. 58]. Two different types of mutation operators are available in GAReg-ISF. The first specializes in the mutation of the rotation quaternion $\mathbf{Q}$ of a parent chromosome and is further subdivided into two different variants: While one adds a random rotation (with arbitrary rotation axis) to the quaternion up to a maximum of 10°, the other restricts the rotation around the normal vector of a randomly selected feature (plane patch). The second type of operators mutates the translation elements t_x, t_y, t_z. Also in this case two variants

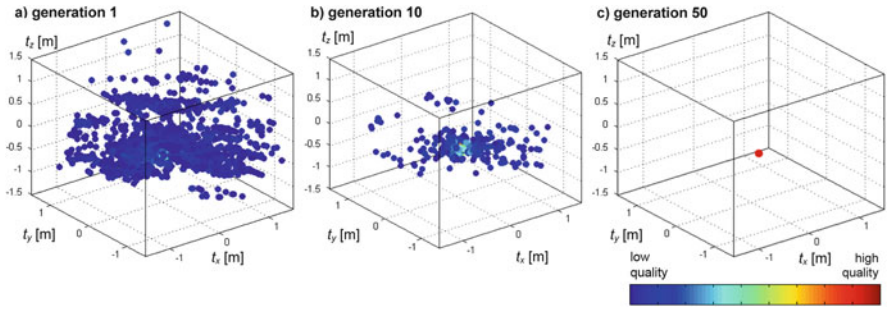

Fig. 9.8 Convergence of the solutions

are implemented: one for random translation and another for translation along a randomly selected plane patch. The maximum size of the translation can be set, i.e., to half the size of the subdivided features. To improve the convergence process, all thresholds are decreased linearly after each generation.

Contrary to the mutation operator, the cross-over operator works with two parent chromosomes and recombines them to two new ones. Hereby first an interpolation factor is determined randomly between 0 and 1. Such an interpolation can be done quite easily for the translation vectors; for the quaternions, e.g., a spherical lineal interpolation (SLERP) can be used. Most geometric libraries already contain such a function.

Selection generally serves two different objectives in GAReg-ISF. There is the selection of genetic operators which is handled by a classical roulette wheel. More important is the selection of those individuals forming new generations. This is done by a binary tournament [16, p. 75]. Hereby first two individuals are selected by a weighted roulette wheel as competitors; individuals with a higher fitness are more likely to be selected. Subsequently, the selected two individuals are compared against each other and the one showing the higher fitness is selected for the next generation. These steps are repeated until the specified size of the new population (i.e. 100) is reached.

Then again the creation of new individuals is started and the above described steps are repeated until a termination criterion is met. As already mentioned, this can be, e.g., the maximum number of generations, the convergence of the solutions or a combination of both.

9.4.2.3 Fine Registration

Optimized by the Genetic Algorithm, the solutions resulting from rough alignment are successively processed by a fine registration algorithm. Fine registration, however, only works with one solution at a time and has to be repeated for all others consecutively. Therefore the solutions are first sorted according to their quality (see

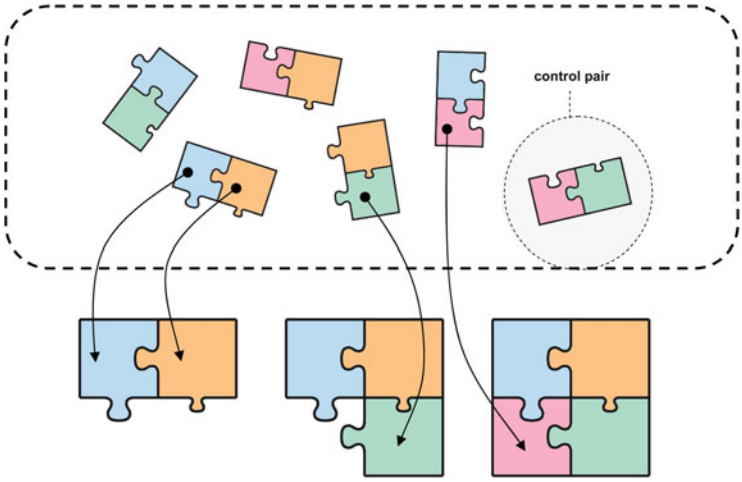

Fig. 9.9 Global registration

again Eq. (9.4)); hereby similar solutions are grouped together. Those showing the highest fitness in each group are passed on. Next, solutions presenting a quality lower than 10 % or more than that of the best solution are eliminated. This way in our examples only a couple of solutions remained for fine registration.

Several approaches of fine registration are already discussed in Sect. 9.3.3. GAReg-ISF implements the improved ICP-algorithm according to [18]. Hereby only the points contributing to a geometrically stable solution are used for registration.

There are, however, cases where more than one solution can seem feasible during pair-wise registration. These are individually refined and passed on to the following global registration process in which the "correct" solution may be identifiable.

9.4.3 Global Registration

As for the global matching of the already calculated pair-wise results, GAReg-ISF follows an approach described in [45]. First, the pair-wise results are sorted according to their quality. Next, the best pair-wise solution is fixed and iteratively the next pair is added (see Fig. 9.9). To identify globally inconsistent solutions a visibility check [42] is advisable. Whenever a point cloud is added, the already fixed point clouds are realigned to avoid the accumulation of errors. Remaining pairs may be used as control pairs. The result of the global registration step is a set of globally consistent arranged point clouds.

9.5 Software Implementation

C-Sharp, an object-oriented programming language developed by Microsoft within its .NET initiative, allowed the fast and flexible development of GAReg-ISF; it must, however, be mentioned that this might slightly slow down computationally expensive calculations. Generally, the automatic registration of point clouds requires a lot of different geometric calculations. These can be done by using, for example, SlimDX, a free open source framework based on .NET technologies.

A lot of different software packages are currently available for point cloud processing and point cloud manipulation. A widely used software for point cloud processing is Innovmetric Polyworks™. It provides specialized tools for importing, triangulating, fine registering, and analyzing point clouds, which can be used by external programs.

Based on [54], GAReg-ISF was redesigned as a client application which makes use of different Innovmetric Polyworks™ v10 modules. Several steps during the automatic point clouds registration are actually done within these modules. This regards, e.g., the import and the triangulation of the point clouds to create (reduced) meshes. GAReg-ISF is able to work with point clouds as well as with triangulated meshes. In the latter case the vertices of the mesh are used as (reduced) "point cloud" further on.

Also part of the pair-wise fine registration and the iterative realignment of all point clouds during global matching as well as final merging of the single point clouds (meshes) and the visualization and output are carried out with Innovmetric Polyworks™ modules. The software communication hereby is handled by Microsoft's Component Object Model (COM) architecture.

9.6 Experimental Results

To evaluate and test the GAReg-ISF approach in reality, several experimental datasets were used. Beneath some simulated datasets [54], measured data from architecture, civil, and mining engineering and other fields, especially archaeological data from excavations, proved to show very clearly the advantages and power of the approach presented.

9.6.1 Introduction

The special research program HiMAT (*Hi*story of *M*ining *A*ctivities in *T*yrol and Adjacent Areas) was established at the University of Innsbruck as an interdisciplinary project in 2007. The aim of this international research program was the analysis of the impact of mining activities on the environment and human society.

Fig. 9.10 Excavation site with wooden chest

Twelve international institutions of different scientific fields of Natural Sciences, the Humanities, and Engineering were taking part in this research consortium. During the HiMAT program, the objective and rigorous documentation of the archaeological excavation sites and findings was carried out by the Unit for Surveying and Geoinformation (see [25] and [26]). In the frame of the HiMAT-Special research program the involved archaeologists of the internationally well-known German Mining Museum in Bochum, Germany were able to continue their excavations in the industrial area alongside the famous main load mining area at the Mitterberg. This mine can be considered as one of the largest Bronze Age mining districts in Europe. Aside the mining depressions an extensive area of ore-beneficiation is known [57]. The excavations in 2008 and 2009 uncovered an area of wet beneficiation and an ore-washery (see Fig. 9.10). In the center of these installations a fully preserved wooden chest was discovered, in which ore was washed and perhaps heavier, fine grinded ore residua were concentrated. While the wooden chest is singular and outstandingly preserved, a complete documentation using different scanning techniques was desirable.

When the archaeologists found this wooden structure, the condition of the object was unrivaled. The geometry of the chest seemed undamaged, therefore it was possible to investigate the Bronze Age mounting techniques of the wooden boards. The three-dimensional documentation of this site was a top priority, because later investigation of these mounting methods would have been impossible after the excavation of the find.

The terrestrial laser scanning data acquisition was accomplished with a Trimble GX 3D Scanner in October 2009 (see Fig. 9.10). The earliest and the final stage

of the archaeological site (extension about 5 by 6 m) have been measured. The first documentation was managed, when the approximately 1.5 × 1.5 m wide and 0.5 m high wooden structure was found. The second data acquisition was before the beginning of the conservation process. The resolution of the two scans of excavation layers was 2 mm but the surrounding was scanned with resolution of only 20 mm. At an average scanner to object distance of 5 m, the single point accuracy proved to be around 3 mm. These parameters qualified the three-dimensional documentation of the excavations with the used Trimble GX 3D Scanner [41]. Each attitude of the instrument was carefully planned to ensure complete coverage of the object. The two excavation situations were scanned during 2 days from a total of 14 different positions. The volume of the raw dataset was about 14 million points (see [27] and [38]).

Accompanying the laser scanning surveys, the wooden chest was recorded also photogrammetrically with a Nikon D200 calibrated digital camera. The resolution of the eighty photos of 3,872 × 2,592 pixels provides accurate image data for later texturing of the achieved 3D object.

9.6.2 Example

The potential and the limits of GAReg-ISF were studied carrying out an experiment on a dataset captured the second day at the Mitterberg mining area. Archaeological excavations generally are well-suited for testing automatic registration strategies (without the use of artificial target spheres) as they usually don't contain ideal features such as planes or other perfect geometric elements. Such cases can be demanding for the automatic registration of point clouds.

The considered excavation extends about 5 by 6 m and is about 1 m deep. Figure 9.11 gives an overview of the site showing also the position of the six closest target spheres which were later used for accuracy studies (see Table 9.2). In overall that day five single laser scans/point clouds (see Fig. 9.12) were taken from different positions with about 1.1 up to 1.7 million points. The scan resolution was set to 20 mm at 7 m distance for the excavation and 2 mm for target spheres and the wooden chest itself. The following steps were fully and automatically managed by GAReg-ISF. First, the scans were imported and preprocessed by making use of the IMAlign module of Innovmetric Polyworks™. As import parameters the interpolation step (grid sampling step) was predefined with 20 mm, the maximum angle between the laser beam and the surface with 85° and the maximum edge length for the triangles with 300 mm.

Next, triangulation was automatically executed by the Innovmetric IMMerge module. The parameters were set to 100 mm regarding the maximum search distance, 20 mm regarding the surface sampling step as well as the maximum smoothing radius with an overall smoothing level set to medium. The whole import and the triangulation took about 1 min per scan. The resulting number of vertices and triangles can be seen in Table 9.1. All further calculations were based on vertices

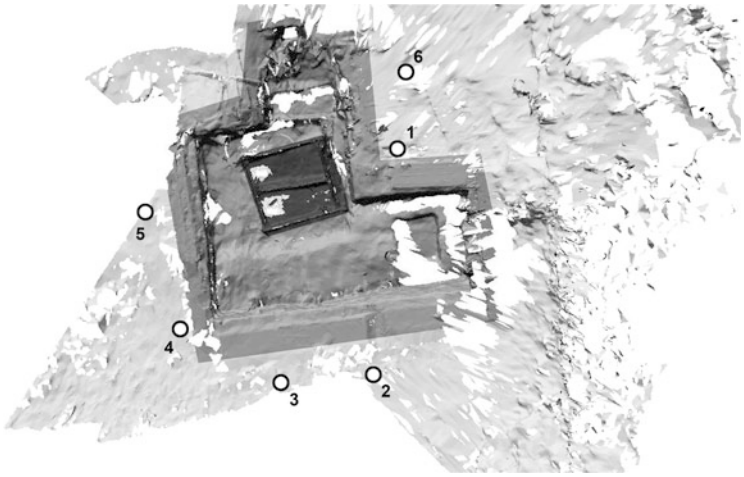

Fig. 9.11 Overview with target spheres

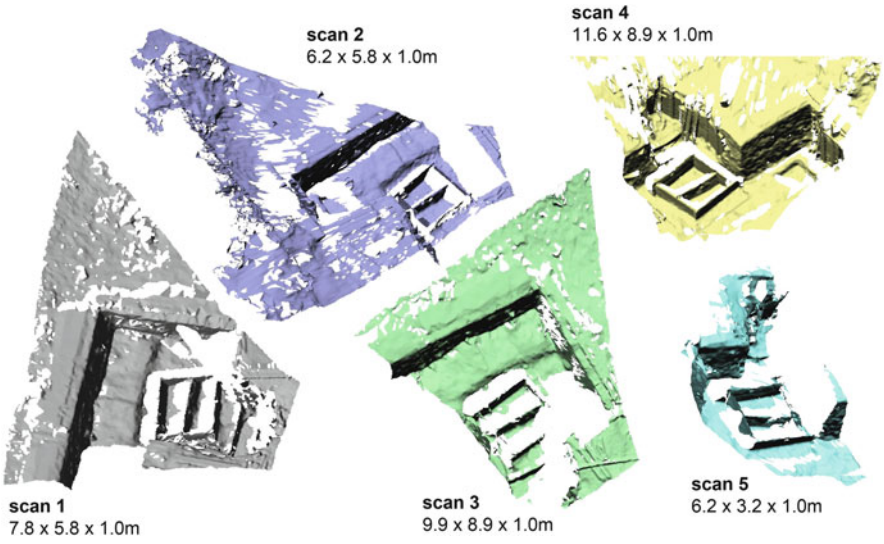

Fig. 9.12 Laser scans from different positions

which can be seen themselves as pruned copy of the original point cloud and, therefore, enabled a faster processing.

Next, the five scans were registered pair-wisely by the GAReg-ISF algorithm, executing as a first step a scan analysis and generating the imperfect and subdivided features with a size of 400 mm. Figure 9.13 shows the detected imperfect and subdivided features of scan 1 and scan 4, whereas Table 9.1 indicates their quantity. The high quantity of subdivided plane patches in scan 4, however, comes mostly

Table 9.1 Laser scan details

Scan	Points	Vertices[a]	Triangles[a]	Plane patches	Subdivided plane patches
1	1,156,482	31,198	56,867	41	208
2	1,176,283	26,894	49,495	35	170
3	1,698,914	59,873	105,227	100	295
4	1,747,802	116,373	196,562	197	533
5	1,142,105	16,787	30,010	38	87

[a]Triangulated mesh created with Innovmetric Polyworks™ v10

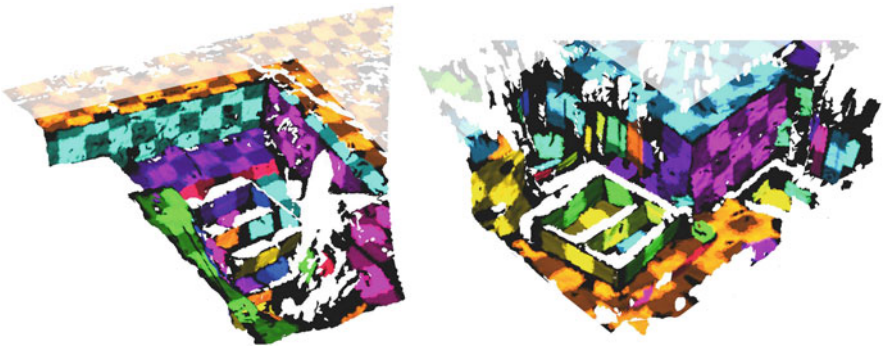

Fig. 9.13 Imperfect and subdivided features of scan 1 and 4

from the surrounding area of the excavation which does not directly contribute to the registration process. The detected features were then used for a rough pair-wise alignment and for further processing using Genetic Algorithms. Finally, the best solutions were fine-registered. In this example, the pair-wise registration of two scans took between 1 and 5 min on an Intel i5 M460 processor (using a single core only). When all scans were registered pair-wisely, global registration was executed as described in Sect. 9.4.3. It has, however, to be mentioned that the global visibility check [42] was indeed necessary to enable the identification of the correct registered pairs. The result of the global registration is shown in Fig. 9.14.

9.6.3 Accuracy and Comparison

To test the capabilities of GAReg-ISF algorithm in comparison to "classical" registration using targets spheres, a number of experiments were conducted on the dataset of the excavation which can be divided into two main phases.

First, "classical" registration was done using the globally known position of 13 artificial target spheres positioned around the excavation area to calculate the registration matrices of the five single scans. Hereby an additional overview laser scan, showing only the 13 spheres with high density, was used to capture the global

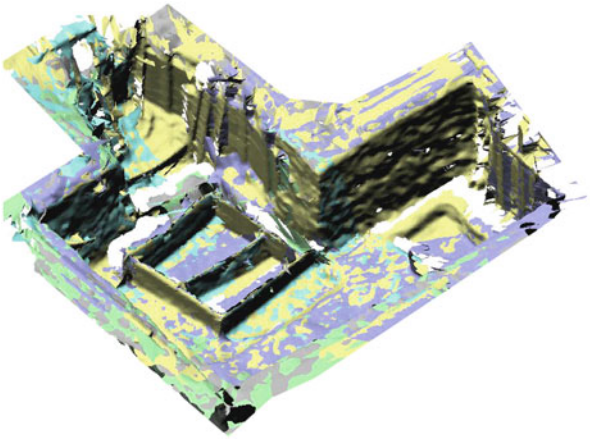

Fig. 9.14 Registered scans

position of the target sphere centers; these served as reference system. Scanning at least three spheres from each single laser scanner position enabled the registration of the five single scans (see Fig. 9.12) into the global reference system.

Next, the identical five scans were registered a second time; this time, however, without any target spheres, but by implementing the automatic registration method GAReg-ISF. For comparison reasons the hereby generated registration matrices were applied afterwards to the centers of the six closest target spheres around the excavation area (see Fig. 9.11) which are given for each scan. The means of the target sphere centers were used to register the results of GAReg-ISF further into the established global reference system. Table 9.2 shows the captured sphere centers, their means, and the standard deviation of both above described methods. It can be seen that the registration using target spheres leads to standard deviations between 0.04 and 2.09 mm, whereas GAReg-ISF gives values between 0.57 and 5.43 mm. While the "classical" approach is focused only on the optimization of the target spheres, GAReg-ISF is working without any spheres but with the surfaces as such; this also explains the higher standard deviations of GAReg-ISF. Furthermore, it has to be added that already minimal translations and rotations of the scans have a bigger influence on the spheres positioned outside around the excavation area. The spatial discrepancy of both solutions is between 0.20 and 6.33 mm, which clearly encourages further research.

In a next step, the registered triangulated surfaces of the scans themselves were analyzed and considered. Hereby the standard deviation and the root mean squared error (RMSE) were calculated between one scan, which was fixed as reference, and all others using Innovmetric Polyworks™. Using a upper threshold distance of 25 mm, Table 9.3 shows that the standard deviation of the classical registration ranges between 7.91 and 10.11 mm, and those of GAReg-ISF between 7.80 and 10.27 mm. Similarly, the RMSE of the target sphere registration is between 8.00 and 10.16 mm whereas those of GAReg-ISF are between 7.80 and 10.29 mm. In four of five scans GAReg-ISF produces lower values with a maximum difference

Table 9.2 Comparison of the target sphere centers

	Registration using 13 target spheres			Registration using GAReg-ISF[a]		
	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)
Sphere 1	1,445.37	8,977.35	349.00	1,441.11	8,970.72	345.62
	1,445.33	8,977.28	348.97	1,450.41	8,975.48	348.45
	1,445.29	8,977.62	348.01	1,448.37	8,980.04	350.28
Mean[b]	1,445.33	8,977.41	348.66	1,446.63	8,975.41	348.12
Std. dev.	0.04	0.18	0.57	4.89	4.66	2.35
Sphere 2	5,257.30	7,758.05	60.12	5,250.83	7,761.09	60.62
	5,257.63	7,757.24	60.18	5,251.07	7,760.64	60.03
	5,256.36	7,756.82	59.98	5,253.11	7,752.87	58.51
Mean[b]	5,257.10	7,757.37	60.09	5,251.67	7,758.20	59.72
Std. dev.	0.66	0.63	0.10	1.25	4.62	1.09
Sphere 3	4,993.79	6,128.50	−23.30	4,989.71	6,131.15	−22.69
	4,994.17	6,127.43	−23.22	4,992.06	6,130.12	−23.49
Mean[b]	4,993.98	6,127.96	−23.26	4,990.89	6,130.63	−23.09
Std. dev.	0.27	0.76	0.05	1.67	0.73	0.57
Sphere 4	3,665.75	4,610.29	−119.73	3,663.90	4,611.00	−119.16
	3,664.53	4,610.69	−119.81	3,666.57	4,609.76	−120.12
Mean[b]	3,665.14	4,610.49	−119.77	3,665.24	4,610.38	−119.64
Std. dev.	0.86	0.29	0.05	1.89	0.87	0.68
Sphere 5	1,535.99	4,464.69	−110.82	1,534.35	4,462.29	−110.50
	1,533.67	4,467.64	−111.15	1,536.12	4,460.88	−111.33
Mean[b]	1,534.83	4,466.17	−110.99	1,535.24	4,461.59	−110.92
Std. dev.	1.64	2.09	0.23	1.25	1.00	0.59
Sphere 6	211.78	9,472.94	−354.94	218.34	9,472.94	−356.76
	211.90	9,474.41	−354.20	216.34	9,480.63	−351.63
Mean[b]	211.84	9,473.68	−354.57	217.34	9,476.79	−354.20
Std. dev.	0.08	1.04	0.53	1.41	5.43	3.63

[a]At the end the resulting registration matrices were applied to the unregistered target spheres
[b]The spatial distance between the calculated means for sphere 1 is 5.50 mm, sphere 2: 2.45 mm, sphere 3: 4.60 mm, sphere 4: 0.20 mm, sphere 5: 4.09 mm, sphere 6: 6.33 mm

Table 9.3 Standard deviations of the surfaces

	Registration using 13 target spheres			Registration using GAReg-ISF		
Reference[a]	Points	Std. dev.[b]	RMSE[b]	Points	Std. dev.[b]	RMSE[b]
Scan 1	40,622	8.50 mm	8.52 mm	40,469	8.39 mm	8.39 mm
Scan 2	40,206	8.26 mm	8.43 mm	40,042	7.96 mm	7.96 mm
Scan 3	44,002	7.91 mm	8.00 mm	44,079	7.80 mm	7.80 mm
Scan 4	36,710	8.78 mm	8.96 mm	36,682	8.66 mm	8.76 mm
Scan 5	18,297	10.11 mm	10.16 mm	18,031	10.27 mm	10.29 mm

[a]The corresponding scan was fixed as reference and compared to all other scans
[b]The upper threshold distance was set to 25 mm

of 0.30 mm regarding the standard deviation and 0.47 mm of the RMSE. Looking at the compared quantity of points (resulting from the maximum comparing distance of 25 mm) one can see that, apart from scan 3, GAReg-ISF uses a slightly lower number of points. This may have contributed to the results.

In overall, comparing the surfaces gives quite similar results for both methods. We could prove that GAReg-ISF, due to the use of imperfect and subdivided features, is able to automatically register also laser scans which don't consist of typical geometric elements such as ideal planes, and can achieve results of comparable quality regarding the registration of surfaces.

9.7 Summary

In this contribution the authors propose a method for the automatic complementary registration of arbitrary point clouds originating from terrestrial laser scanners without using artificial targets. The presented approach can be seen as improvement to the state of the art as it combines aspects of different already well-studied methods such as feature matching and engineering applications of Genetic Algorithms. Combining and optimizing the positive facets of these techniques, their application to the automatic registration of partially occluded point clouds that are indicated by even significant noise level and imperfect geometry was enabled.

One of the great advantages of the presented concept is the acceptance of "a certain imperfectness" of the features in the individual datasets and the setup of a registration framework capable to handle them. The rigorous subdivision of features into smaller sub-features allows overcoming occlusions at the object and both robustness and computational performance were increased when registering point clouds of any size. Together with Genetic Algorithms this was implemented as clearly targeted step in between classical rough alignment and fine registration.

The provided example shows that the approach of subdivided features is not only applicable to objects consisting of classical geometric primitives such as planes (as in most architectural and engineering applications); due to the concession of imperfect features also scenes consisting of approximated features are manageable. Within a geometric documentation of an archaeological excavation we proved exemplarily that the presented approach is able to reach results of comparable accuracy as the classical registration using artificial spheres. Even in this complex surrounding it supplies evidence for the efficiency of the presented overall registration work-flow.

Acknowledgements The dataset discussed in this article was provided by SFB HiMAT, the Special Research Program on the HiMAT. The project was mainly funded by the Austrian Science Fund (FWF), TransIDEE (the Science and Technology Transfer Center of the University of Innsbruck), the regional authorities of Tyrol, Salzburg, and Vorarlberg as well as the Autonomous Province Bolzano/South-Tyrol (Italy).

References

1. Arya, S., Mount, D.M., Netanyahu, N.S., Silverman, R., Wu, A.Y.: An optimal algorithm for approximate nearest neighbor searching fixed dimensions. J. ACM **45**(6), 891–923 (1998)
2. Bae, K.-H.: Automated registration of unorganised point clouds from terrestrial laser scanners. Ph.D. thesis, Curtin University of Technology (2006)
3. Bentley, J.L.: Multidimensional binary search trees used for associative searching. Commun. ACM **18**(9), 509–517 (1975)
4. Besl, P., McKay, N.: A method for registration of 3D shapes. IEEE Trans. Pattern Anal. Mach. Intell. **14**(2), 239–256 (1992)
5. Brenner, C., Dold, C.: Automatic relative orientation of terrestrial laser scans using planar structures and angle constraints. In: Proceedings of the ISPRS Workshop 'Laser Scanning 2007 and SilviLaser 2007', Espoo, vol. XXXVI, Part 3 / W52 of IAPRS, The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 84–89 (2007)
6. Brenner, C., Dold, C., Ripperda, N.: Coarse orientation of terrestrial laser scans in urban environments. ISPRS J. Photogram. Rem. Sens. **63**(1), 4–18 (2008)
7. Brunnström, K., Stoddart, A.J.: Genetic algorithms for free-form surface matching. In: Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria, 1996
8. Chen, Y., Medioni, G.: Object modeling by registration of multiple range images. Image Vis. Comput. **10**(3), 145–155 (1992)
9. Chen, C.-S., Hung, Y.-P., Cheng, J.-B.: RANSAC-based DARCES: a new approach to fast automatic registration of partially overlapping range images. IEEE Trans. Pattern Anal. Mac. Intell. **21**(11), 1229–1234 (1999)
10. Chetverikov, D., Stepanov, D., Krsek, P.: Robust Euclidean alignment of 3D point sets: the trimmed iterative closest point algorithm. Image Vis. Comput. **23**(3), 299–309 (2005)
11. Chow, C., Tsui, H., Lee, T.: Surface registration using a dynamic genetic algorithm. Pattern Recognit. **37**(1), 105–117 (2004)
12. Chua, C.S., Jarvis, R.: Point signatures: a new representation for 3D object recognition. Int. J. Comput. Vis. **25**(1), 63–85 (1997)
13. Cordón, O., Damas, S., Santamaría, J.: Feature-based image registration by means of the CHC evolutionary algorithm. Image Vis. Comput. **24**(5), 525–533 (2006)
14. Deschaud, J.-E.: A fast and accurate plane detection algorithm for large noisy point clouds using filtered normals and voxel growing. In: 3DPVT'10: Proceedings of the 5th International Symposium on 3D Data Processing, Visualization and Transmission, Espace Saint-Martin, Paris (2010)
15. Dold, C., Brenner, C.: Registration of terrestrial laser scanning data using planar patches and image data. In: Proceedings of the ISPRS Commission V Symposium 'Image Engineering and Vision Metrology', Dresden, vol. XXXVI, Part 5 of IAPRS, The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences (2006)
16. Dumitrescu, D., Lazzerini, B., Jain, L.C., Dumitrescu, A.: Evolutionary Computation. CRC Press, Boca Raton (2000)
17. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
18. Gelfand, N., Ikemoto, L., Rusinkiewicz, S., Levoy, M.: Geometrically stable sampling for the ICP algorithm. In: 3DIM 2003: Proceedings of the 4th International Conference on 3-D Digital Imaging and Modeling, Banff, Alberta, pp. 260–267 (2003)
19. Gelfand, N., Mitra, N.J., Guibas, L.J., Pottmann, H.: Robust global registration. In: SGP 2005: Proceedings of the 3rd Eurographics Symposium on Geometry Processing 2005, pp. 197–206. Vienna (2005)
20. Gordon, B.: Zur Bestimmung von Messunsicherheiten terrestrischer Laserscanner. Ph.D. thesis, Technische Universität Darmstadt (2008)

21. Grussenmeyer, P., Hanke, K.: Einsatz eines TLS-Systems in der Ingenieurvermessung. In: Chesi, G., Weinold, T. (eds.) 14. Internationale Geodätische Woche Obergurgl 2007, pp. 21–30 (2007)
22. Grussenmeyer, P., Landes, T., Voegtle, T., Ringle, K.: Comparison methods of terrestrial laser scanning, photogrammetry and tacheometry data for recording of cultural heritage buildings. In: Proceedings of the ISPRS Commission V Congress, Beijing, vol. XXXVII, Part B5 of IAPRS, The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 213–218 (2008)
23. Hanke, K., Grussenmeyer, P., Grimm-Pitzinger, A., Weinold, T.: First experience with the trimble GX Laser Scanner. In: Proceedings of the ISPRS Commission V Symposium 'Image Engineering and Vision Metrology', Dresden, Germany, vol. XXXVI, Part 5 of IAPRS, The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences (2006)
24. Hanke, K., Moser, M., Grimm-Pitzinger, A., Goldenberg, G., Toechterle, U.: Enhanced potential for the analysis of archaeological finds based on 3D modeling. In: Proceedings of the ISPRS Commission V Congress, Beijing, vol. XXXVII, Part B5 of IAPRS, The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 187–192 (2008)
25. Hanke, K., Hiebel, G., Kovács, K., Moser, M.: Surveying and geoinformation - contributions to an interdisciplinary special research program on the history of mining activities. In: Proceedings of the 22nd CIPA Symposium, vol. XXII-2009 of The ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences and The CIPA International Archives for Documentation of Cultural Heritage (2009)
26. Hanke, K., Hiebel, G., Kovács, K., Moser, M.: Documentation challenges in an international and interdisciplinary research project. In: Proceedings of the ISPRS Commission V Midterm-Symposium in Newcastle upon Thyne, UK, vol. 39 (5) of The ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences and The CIPA International Archives for Documentation of Cultural Heritage (2010)
27. Hanke, K., Stöllner, T., Kovács, K., Moser, M.: Combination of different surveying methods for archaeological documentation: the case study of the bronze age wooden chest from Mitterberg. In: Contreras, F., Melero, F.J. (eds.) Proceedings of the 38th Annual Conference and Computer Applications and Quantitative Methods in Archaeology, BAR International Series 2494. Archaeopress, Oxford (2013)
28. Haring, A.: Die Orientierung von Laserscanner- und Bilddaten bei der fahrzeuggestützten Objekterfassung. Ph.D. thesis, Technische Universität Wien (2007)
29. He, W., Ma, W., Zha, H. Automatic registration of range images based on correspondence of complete plane patches.In: 3DIM 2005: Proceedings of the 5th International Conference on 3-D Digital Imaging and Modeling, Ottawa, pp. 470–475 (2005)
30. Heine, E., van Genechten, B., Lerma García, J.L., Santana Quintero, M.: Theorie und Praxis des terrestrischen Laserscannings: Deutsche Version des Theorieteils und des praktischen Trainingteils Kirche St. James aus der DVD-Publikation: Theory and practice on Terrestrial Laser Scanning (5 languages), vol. 152. Universidad Politécnica de Valencia, Valencia (2008)
31. Hofer, M., Pottmann, H.: O Laserscanner-Punktwolken. Vermessung Geoinformation **91**, 297–306 (2003)
32. Holland, J.: Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor and Michigan (1975)
33. Horn, B.K.P.: Closed-form solution of absolute orientation using unit quaternions. J. Opical Soc. Am. **4**(4), 629–642 (1987)
34. Huang, C.-W.: The application of genetic algorithms in rough registration of three-dimensional range images. Ph.D. thesis, Taiwan University (2002)
35. Huang, Q.-X., Flöry, S., Gelfand, N., Hofer, M., Pottmann, H.: Reassembling fractured objects by geometric matching. In: SIGGRAPH 2006: Proceedings of the 33rd International Conference and Exhibition on Computer Graphics and Interactive Techniques, Boston (2006)

36. Johnson, A.E.: Spin-images: a representation for 3-D surface matching. Ph.D. thesis, Carnegie Mellon University, Pittsburgh and Pennsylvania (1997)
37. Kitamura, K., D'Apuzzo, N., Kochi, N., Kaneko, S.: Automated extraction of break lines in TLS data of real environment. In: Proceedings of the ISPRS Commission V Mid-Term Symposium 'Close Range Image Measurement Techniques', Newcastle upon Tyne, UK, vol. XXXVIII, Part 5 of IAPRS, The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences (2010)
38. Kovács, K., Moser, M., Hanke, K.: Application of laser scanning for archaeological prospection and 3D documentation. In: Museen der Stadt Wien - Stadtarchäologie, Workshop 14 - Archäologie und Computer 2009. Phoibos, Wien (2009)
39. Lange, C., Polthier, K.: Anisotropic smoothing of point sets. Comput. Aided Geometric Des. **22**(7), 680–692 (2005)
40. Lomonosov, E., Chetverikov, D., Ekárt, A.: Pre-registration of arbitrarily oriented 3D surfaces using a genetic algorithm. Pattern Recognit. Lett. **27**(11), 1201–1208 (2006)
41. Moser, M., Hye, S., Goldenberg, G., Hanke, K., Kovács, K.: Digital documentation and visualization of archaeological excavations and finds using 3D scanning technology. In: Proceedings of ARQUEOLOGICA 2.0 (1st International Meeting on Graphic Archaeology and Informatics, Cultural Heritage and Innovation) (2009)
42. Neugebauer, P.J.: Reconstruction of real-world objects via simultaneous registration and robust combination of multiple range images. Int. J. Shape Model. **3**, 71–90 (1997)
43. Pauly, M., Keiser, R., Gross, M.: Multi-scale feature extraction on point-sampled surfaces. In: EG 2003: Proceedings of the Eurographics 2003, pp. 281–289. Granada (2003)
44. Pottmann, H., Leopoldseder, S., Hofer, M.: Simultaneous registration of multiple views of a 3D object. In: PCV02: Proceedings of the ISPRS Commission III Symposium 'Photogrammetric Computer Vision', Graz, Austria, vol. XXXIV, Part 3 of IAPRS, The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 265–270 (2002)
45. Pulli, K.: Multiview registration for large data sets. In: 3DIM'99: Proceedings of the 2nd International Conference on 3-D Digital Imaging and Modeling, Ottawa, pp. 160–168 (1999)
46. Rabbani, T.: PROTO: Pointcloud Processing Toolbox (Matlab). http://viztronics.com/tahir/data/PROTO/doc/menu.html (2003). Accessed 18 Apr 2014
47. Rabbani, T.: Automatic reconstruction of industrial installations using point clouds and images. Ph.D. thesis, TU Delft, The Netherlands (2006)
48. Rabbani, T., Dijkman, S., van Den Heuvel, F., Vosselman, G.: An integrated approach for modelling and global registration of point clouds. ISPRS J. Photogram. Rem. Sens. **61**(6), 355–370 (2007)
49. Reed, M.B., Schenk, S., Swoboda, G.: FTO: A genetic algorithm for tunnel design optimisation. In: GECCO 2005: Late breaking papers of the Genetic and Evolutionary Computation Conference, Washington, DC (2005)
50. Remondino, F.: Image-based modeling for object and human reconstruction. Ph.D. thesis, Swiss Federal Institute of Technology (ETH) Zurich (2006)
51. Rietdorf, A.: Automatisierte Auswertung und Kalibrierung von scannenden Messsystemen mit tachymetrischem Messprinzip. Ph.D. thesis, Technische Universität Berlin (2005)
52. Rusinkiewicz, S., Levoy, M.: Efficient variants of the ICP algorithm. In: 3DIM 2001: Proceedings of the 3rd International Conference on 3-D Digital Imaging and Modeling, Québec City, pp. 145–152 (2001)
53. Salvi, J., Matabosch, C., Fofi, D., Forest, J.: A review of recent range image registration methods with accuracy evaluation. Image Vis. Comput. **25**(5), 578–596 (2007)
54. Schenk, S.: Automatische Registrierung von Punktwolken: Genetische Algorithmen mit genäherten und unterteilten Merkmalen. Ph.D. thesis, Universität Innsbruck (2010)
55. Silva, L., Bellon, O.R.P., Boyer, K.L.: Robust range image registration using genetic algorithms and the surface interpenetration measure. Series in Machine Perception and Artificial Intelligence, vol. 60. World Scientific, Hackensack (2005)
56. Stamos, I., Leordeanu, M.: Automated feature-based range registration of urban scenes of large scale. In: CVPR 2003: Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison (2003)

57. Stöllner, T.: Bronzezeitliche Massenproduktion von Kupfer am Mitterberg. Archäologie Deutschland **4**, 32–33 (2008)
58. Tarsha-Kurdi, F., Landes, T., Grussenmeyer, P.: Hough-transform and extended RANSAC algorithms for automatic detection of 3D building roof planes from Lidar data. In: Proceedings of the ISPRS Workshop 'Laser Scanning 2007 and SilviLaser 2007', Espoo, vol. XXXVI, Part 3 / W52 of IAPRS, The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 407–412 (2007)
59. Toldo, R., Beinat, A., Crosilla, F.: Global registration of multiple point clouds embedding the Generalized Procrustes Analysis into an ICP framework. In: 3DPVT'10: Proceedings of the 5th International Symposium on 3D Data Processing, Visualization and Transmission, Espace Saint-Martin, Paris (2010)
60. Vieira, M., Shimada, K.: Surface extraction from point-sampled data through region growing. Int. J. CAD/CAM **5**(1), 19–27 (2005)
61. von Hansen, W.: Robust automatic marker-free registration of terrestrial scan data. In: PCV'06: Proceedings of the ISPRS Symposium of Commission III 'Photogrammetric Computer Vision and Image Analysis', Bonn, vol. XXXVI, Part 3 of IAPRS, The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 105–110 (2006)
62. von Hansen, W., Gross, H., Thoennessen, U.: Line-based registration of terrestrial and airborne LIDAR data. In: Proceedings of the ISPRS Commission III Congress, Beijing, vol. XXXVII, Part B3 of IAPRS, The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 161–166 (2008)
63. Vosselman, G., Maas, H.: Airborne and Terrestrial Laser Scanning. Whittles Publishing, Dunbeath (2010)
64. Zhang, Z.: Iterative point matching for registration of free-form curves and surfaces. Int. J. Comput. Vis. **13**(2), 119–152 (1994)